DOCUMENT RESUME

ED 372 376 CS 011 796

AUTHOR	Binkley, Marilyn, Ed.; Rust, Keith, Ed.
TITLE	Reading Literacy in the United States: Technical Report of the U.S. Component of the IEA Reading Literacy Study.
INSTITUTION	North Arlington School District, NJ.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-94-259
PUB DATE	Aug 94
NOTE	681p.; The study utilized contractual support from Westat, Data Recognition Inc., and the Council of Chief State School Officers (CCSSO).
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Statistical Data (110) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF04/PC28 Plus Postage.
DESCRIPTORS	Data Analysis; Grade 4; Grade 9; Intermediate Grades; Junior High Schools; Literacy; *Reading Ability; *Reading Achievement; Reading Research; Research Methodology
IDENTIFIERS	*International Evaluation Education Achievement; *United States

ABSTRACT

This technical report covers almost every aspect of the United States component of the International Association for the Evaluation of Educational Achievement's (IEA) International Reading Literacy Study, from the inception of the project to the production of the reports. The report notes that since much has already been learned from studies such as the National Assessment of Educational Progress about the reading achievement of fourth and ninth graders, a primary objective in analyzing the data was to go beyond traditional approaches to the analysis and reporting of results. The first part describes all aspects of the process by which data on students, teachers, and schools were collected. The second part describes various aspects of the properties of the reading achievement instruments used, and an initial overview of the reading literacy skills of a few major subgroups. The report culminates in the third part, which describes the methods used to analyze the data and the findings of these analyses. Contains 146 tables and 67 figures of data. Reading literacy tests and student, teacher, and school questionnaires are attached. (RS)

National Center for Education Statistics

Reading Literacy in the United States

Technical Report

U.S. Department of Education
Office of Educational Research and Improvement

NCES 94-259

25

Reading Literacy in the United States:

Technical Report of the U.S. Component of the IEA Reading Literacy Study

Editors

Marilyn Binkley
National Center for Education Statistics

Keith Rust
Westat, Inc.

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

Data Development Division
Jeanne E. Griffith, *Associate Commissioner*

International Activities Group
Eugene Owen, *Chief*

National Center for Education Statistics

"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e-1).

Suggested Citation

U.S. Department of Education, National Center for Education Statistics, *Reading Literacy in the United States. Technical Report.* Washington, D.C.: 1994.

August 1994

Technical Editor: Carol Litman, Westat, Inc.
Cover design: Ana Horton, Westat, Inc.

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328

PREFACE

In 1991, 32 countries participated in The International Reading Literacy Study to evaluate the reading literacy skills of their school students and to assess factors thought to be related to reading literacy. The study was conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). Two populations were assessed in the study: those in the grade with the most 9-year-old students (Population A, grade 4 in the United States) and those in the grade with the most 14-year-old students (Population B, grade 9 in the United States). Most of the participating countries, including the United States, conducted studies of both populations.

The IEA published an initial set of results for all countries in 1992 (*How in the World Do Students Read?* by W.B. Elley). While the IEA has been engaged in analyzing results from the study across nations, within the United States the National Center for Education Statistics has conducted an intensive and extensive analysis of the U.S. national data. The findings from this national research are being released in a series of three reports. This present volume constitutes a comprehensive technical report covering the conduct of the study within the U.S. and the methods of analysis employed on the U.S. data for students, teachers, and schools. A more general report, *Reading Literacy in the United States: Findings from the IEA Reading Literacy Study*, describes the findings from these analyses in a way that is accessible to educational policymakers and others interested in the reading literacy of U.S. students but without the full technical detail provided in this report. A third volume, *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study*, presents papers on a number of methodological issues that had to be confronted within the U.S. in the conduct of the study and the analysis of the data. A fourth volume, *Reading Literacy in an International Perspective*, will cover findings from a number of comparative analyses conducted in partnership with the study representatives from several European countries.

This technical report covers almost every aspect of the project that the U.S. investigators had to address, from the inception of the project to the production of the reports. The study, as part of an international effort, had to follow many major parameters relating to the study as set by IEA and the National Research Coordinators (NRCs) from the 32 participating nations. Consequently, the U.S. undertook the research in a fashion that conformed to these predetermined parameters, as did the other participating countries. As will be seen in the report, this at times required carrying out aspects of the study in a way that varied from principles of "best practice" that would be widely recognized within the U.S. The study design did, however, ensure a high degree of comparability from country to country, and resulted in a very sizable number of participating nations for a study such as this. In analyzing the data within the U.S. context, we were not hampered by constraints to fit an international model, but were constrained by the data that the international project provided.

Since a good deal has already been learned from studies such as the National Assessment of Educational Progress about reading achievement in this nation's schools, a primary objective in analyzing the Reading Literacy Study data was to attempt to go beyond traditional approaches to the analysis and reporting of results from national assessments of educational achievement. To this end, we were aided by recent methodological developments in the area of linear modeling of statistical data, namely, hierarchical linear models. Nevertheless, it must be kept in mind that the study was not designed specifically with U.S. analytical objectives in mind.

This report contains four sections, each having a number of chapters. The first chapter provides an introduction to the study, its components, and the processes by which it was developed. Part I of the body of the report comprises Chapters 2 through 6. This section describes all aspects of the process by which the data on students, teachers, and schools were collected in the U.S. This includes the design of the samples of schools and students, the process by which schools were recruited, and the way in which assessment sessions were organized. It also describes how the data were captured once the assessments had taken place. Part II includes Chapters 7 through 10, which describe various aspects of the properties of the reading achievement instruments used in the assessment to provide a context for the detailed analyses that follow. Part II also gives an initial overview of the reading literacy skills of a few major subgroups within the U.S. student population. The report culminates in Part III, which describes the methods used to analyze the data and the findings of these analyses. These analyses include simple one-way tabulations by a wide variety of relevant characteristics of schools, teachers, and students, as well as complex models that incorporate, simultaneously at the school and student levels, the interrelationships between factors associated with reading literacy.

Thus the various sections of this report attempt to build a comprehensive picture of the processes by which the findings drawn from the study, and presented both here and in the general audience report, were obtained. It details the opportunities and constraints presented to the researchers and describes the decisions and choices made by the researchers in the face of these. In this way, we hope to clarify the basis on which we claim to speak with any authority about the current state of reading literacy among America's school students.

ACKNOWLEDGMENTS

The project that gave rise to this report was part of a larger international effort conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). Our numerous colleagues who served as members of the International Steering Committee, the International Coordinating Center, and the National Research Coordinators in each of the participating countries (as listed in Chapter 1) played a large part in its development.

Within the United States, the research was sponsored by the International Activities Group of the Data Development Division of the National Center for Education Statistics (NCES). Marilyn R. Binkley was project director and the U.S. National Research Coordinator for the study. The data collection and much of the analysis was carried out by Westat, Inc., under the direction of Keith Rust, the corporate manager for this project. Overall project direction was provided by Jeanne E. Griffith, Associate Commissioner for Data Development, and Gary Phillips, Associate Commissioner for Education Assessment. Eugene Owen, Chief of the International Activities Group, provided support, guidance, direction, and a wealth of wisdom.

During the 5-year span of this project those of us at the center of the fray benefited from the cooperation and guidance of many individuals. The editors and authors wish to acknowledge these contributions. Without the cooperation and participation of almost 10,000 students, 600 teachers, and 350 principals who took the time to respond to the questionnaires and to take part in the assessment, the data reported here and in the forthcoming volumes would not be available. Without the generosity and endorsement from the chief state school officers, the state curriculum and assessment staffs, the district superintendents and their curriculum and assessment staffs, the Catholic Schools Association, and headmasters from the private schools, the study would not have been possible. To these people who often are forgotten at times like this, we say thank you.

The Council of Chief State School Officers provided valuable assistance in numerous ways: Gordon Ambach in his role as U.S. representative to the IEA helped steer both the international and national components of the study; Ramsay Selden, as he donned multiple hats, helped us gain the cooperation of states, districts, and schools, provided technical support, and generally kept us honest; the members of the Education Information Advisory Council graciously reviewed proposals, created subcommittees to assist in the development and review of test items, and ultimately endorsed and told others about this study. Zoe Leimgruebler, Tom Kerins, Wayne Martin, and Edward Roeber deserve special thanks for always responding on just a moment's notice.

Several consultants to the project, notably Albert Beaton, Barbara Kapinus, Stephen Norris, Linda Phillips, and Stephen Raudenbush, also got involved and helped refine our thinking on important substantive and methodological issues. The members of the National Steering Committee— Margaret Smith Burke, Marcia Farr, Jerome Harste, Barbara Kapinus, Thomas Kerins, Irwin Kirsch, Larry Mikulecky, Scott Paris, Maria Ramirez, Ed Roeber, Robert Tierney, and Sheila Valencia—provided intellectual leadership. Their efforts during the design and development phases of the study were Herculean and moved the project beyond the mediocre.

Within NCES and OERI, colleagues such as John Burkett, Cynthia Dorfman, Carol Sue Fromboluti, Edward Fuentes, Robert Leestma, Carlyle Maw, Dawn Nelson, Lois Peak, Deborah Sedlacek, Thomas Synder, and Tommy Tomlinson listened and helped us to structure the study, to develop the reporting strategies, and to write unbiased reviews of the literature. Their input often helped us move beyond a particular problem efficiently and effectively.

Susan Ahmed, Robert Burton, and Mary Frase unstintingly gave of their time to review, revise, and redirect our efforts so we might come closer to the high standards we set as our goal. Throughout this project we consistently called upon the same team to review Office of Management and Budget (OMB) packages, to vet analysis plans, and to adjudicate reports. To those people—Nabeel Alsalam, Sharon Bobbitt, Peggy Carr, James Guthrie, Edward Haertel, Sylvia Johnson, Irwin Kirsch, Paula Knepper, Andrew Kolstad, and Floraline Stevens—who stuck with us, thanks for helping us.

Many people within Westat often ran that extra mile for this project. We gratefully acknowledge their support. We particularly would like to acknowledge Steven Roey, who painstakingly managed the data and computing for all of us, Carol Litman, who carefully looked after the details of putting reports together and kept after us in the most cheerful manner, Sylvie Warren, who patiently created and recreated, and recreated the text of the reports, and Susan Hein, who created the numerous graphics. They deserve far more than the special thanks we extend to them at this time.

TABLE OF CONTENTS

PART I. SAMPLE DESIGN AND DATA COLLECTION

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

PART II. MEASURING READING LITERACY

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

PART III. THE VARIABLES THAT AFFECT READING LITERACY

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

Chapter **Page**

TABLE OF CONTENTS (continued)

LIST OF ATTACHMENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

Figure

TABLE OF CONTENTS (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES (continued)

TABLE OF CONTENTS (continued)

LIST OF EXHIBITS

Exhibit

1. INTRODUCTION

1.1. Background

The IEA Reading Literacy Study was an international study conducted to measure reading literacy across nations and to describe the factors associated with reading achievement. It was hoped that, as a result of this study, a measure of the comparative ability of educational systems to teach literacy skills could be devised and that a clear, unified definition of literacy would be developed. The study was a cooperative undertaking of the International Association for the Evaluation of Educational Achievement (IEA) and representatives from these 32 participating nations:

Belgium (French)
Botswana
Canada (British Columbia)
Cyprus
Denmark
Finland
France
Germany, East
Germany, West
Greece
Hong Kong
Hungary
Iceland
Indonesia
Ireland
Italy
Netherlands
New Zealand
Nigeria
Norway
Philippines
Portugal
Singapore
Slovenia
Spain
Sweden
Switzerland
Thailand
Trinidad and Tobago
United States
Venezuela
Zimbabwe

This report provides details on the U.S. portion of this study. Within the United States, the study was directed by the National Center for Education Statistics (NCES), U.S. Department of Education, with the assistance and support of a National Steering Committee and conducted through a contract with Westat, Data Recognition Corporation, and the Council of Chief State School Officers.

Specifically included in this report is information on sample design, data collection, and data analysis procedures used in the United States. Discussion of the instrument development at both the international and national levels are also provided. An elaboration and discussion of methodological issues relating to the procedures used in this study are provided in a separate volume of edited papers.

Consistent with the cooperative nature of IEA studies, each participating nation agreed to use a common set of procedures and to administer the same instruments in the interest of gathering comparative data that would provide both descriptive and explanatory information about reading literacy practices and achievement within each nation and across nations. In some cases these procedures were consistent with standard practice in the United States; in others there was marked divergence. Throughout the report, we discuss IEA requirements and how we met them while also meeting the necessary NCES standards and pragmatic U.S. procedures that would render the study feasible.

1.2. Organization of the Technical Report

The remainder of this chapter will provide the reader with an overview of the study. It provides only a brief description of the study design and organization that might serve as the scaffolding for the subsequent chapters where specific dimensions of the study are discussed in detail.

Part I of the report focuses on the field operations. It includes detailed description of the sample design, the enlistment procedures, response rates, data collection, and preparation of the data files. Throughout these chapters the reader will find reference to the requirements for the IEA and a description of the U.S. procedures used.

Part II of the report focuses on the instrumentation, while Part III focuses on the analysis done in the United States. For the most part, within the United States, we began with internationally developed instruments, even though these instruments were inconsistent in some respects with current theory and practice in the U.S. These chapters provide a discussion of the instruments, evaluating how well they function in the U.S. In addition, they describe how the U.S. team went about the analysis of the national data.

1.3. Overview of the Study

1.3.1. The International Nature of the Project

Within the United States there has been growing interest in cross-national comparisons of educational achievement. Although studies of this kind have been conducted by the IEA since the late 1960s, the publication of *A Nation at Risk* in 1983 brought this interest into national prominence. Since that time, and in light of the rapidly changing international political and economic climate, this interest has focused on a concern about the ability of our population to meet the growing challenges of an information society and a desire to maintain our competitive advantage in the world economy.

As the results of most of the previous cross-national comparative studies indicate, there are many reasons to believe that the U.S. education system is not producing the desired student outcomes. As noted in the Second International Mathematics Study (1981) and the Second International Science Study (1983), both IEA projects, and the International Assessment of Educational Progress (IAEP), conducted by the Education Testing Service (ETS) in 1988, the U.S. performance in mathematics and science was not particularly good. Researchers and policymakers were interested in uncovering those aspects of our education system that could be changed to improve student performance. Many hypotheses about instruction and learning that would be considered sacrosanct within the U.S. could be challenged internationally based on existing differences in current practices. Therefore, cross-national studies could provide insight into ways to improve our educational infrastructure.

Prior efforts to compare educational opportunities across countries focused primarily on differences in educational inputs, measured by the number of school days, the rate of expenditure per student, the number of books per student, and the like. More recently the emphasis has shifted to a desire to understand finer nuances of the inputs, such as differences in instructional methodology and curriculum organization and coverage, as they relate to a desired level of performance. The IEA studies, as well as the IAEP studies, provided a vehicle for gathering the desired data both on achievement and on background factors that would influence differences in achievement outcomes.

1.3.2. The IEA

The IEA was originally organized to act as the focus and coordinating agency for international comparative studies of educational achievement. Formed by a group of distinguished scholars, it operated as an independent network of researchers all over the world.

Over the past 30 years, IEA has conducted assessments in most of the main scholastic subject areas in a variety of nations. Mathematics, science, reading comprehension, second languages, civics education, classroom environments, written composition, literature, pre-primary education, and computers in education were the focus of most of their individual projects.

While carrying out these projects, this loosely organized collaborative body of research institutes developed a well-defined set of procedures for carrying out comparative studies. Their procedures called for the formation of an International Steering Committee (ISC) to oversee each of the studies, a National Research Center (NRC) in each participating country to conduct the study within each nation, and an International Coordinating Center (ICC) to process the data.

The IEA procedures include four levels of participation:

- **The obligatory international core.** All nations collect the agreed upon data in the specified manner using internationally developed instruments. These data are processed internationally.
- **The international option.** The instruments and data collection procedures are specified internationally, but nations may choose whether they wish to collect the data. The data are processed internationally.
- **The special national option.** The instruments and data collection procedures are specified internationally, but the analysis is left to individual nations. Nations may choose whether they wish to collect these data.
- **The national option.** The instruments, data collection procedures, and analysis are left to individual nations.

The IEA procedures also have some built-in quality safeguards. These include subjecting the study design to an IEA international review committee, the inclusion in each study of a sampling referee who is expected to monitor the execution of the sampling designs within each participating country and to maintain equitable standards for exclusion rates and cooperation rates, and data processing procedures that ensure that each country's data conform to an international format.

1.4. The Organizational Structure for the IEA Reading Literacy Study

Consistent with IEA procedures, the Reading Literacy Study began with the formation of an International Steering Committee (ISC) and the designation of an International Coordinating Center (ICC).

The members of the ISC included the following persons:

Chair:
Dr. Warwick B. Elley
University of Canterbury
Christchurch 1
New Zealand

International Coordinator:
Dr. T. Neville Postlethwaite
Institute for Comparative Education
University of Hamburg
Sedanstrasse 19
D-2000 Hamburg 13
Germany

Sampling Referee:
Dr. Kenneth Ross
School of Education
Deakin University
Geelong, Victoria
Australia

Dr. Ingvar Lundberg
Department of Psychology
University of Umea
Radhuseplanaden 2
S-90247 Umea, Sweden

Dr. John T. Guthrie
College of Education
University of Maryland
College Park, MD 20742
USA

Dr. Francis Mangubhai
Darling Downs
Institute of Advanced Education
Toowoomba, Queensland 4350
Australia

Dr. Alan C. Purves
School of Education
State University of New York
1400 Washington Avenue
Albany, NY 12222
USA

The International Coordinating Center, under the leadership of Dr. T. Neville Postlethwaite, was established at the Institute for Comparative Education at the University of Hamburg.

National Research Centers were established and the following National Research Coordinators were designated:

Dr. Dominique La Fontaine
Service de Pedagogie Experimentale
Universite de Liege au Sart-Tilman B32
B-4000 Liege 1
Belgium

Mrs. Serara Moahi
Ministry of Education
Research Testing Centre
P.O. Box 189
Gaborone
Botswana

Dr. Victor Froese
Department of Language Education
University of British Columbia
2125 Main Mall
Vancouver, B.C., V6T IZ5
Canada

Dr. Constantinos Papanastasiou
Department of Research and Evaluation
Pedagogical Institute
P.O. Box 512
Nicosia
Cyprus

Mr. Jan Mejding
Danish Institute for Educational Research
28, Hermodsgade
DK-2200N Copenhagen
Denmark

Professor Dr. Rainer Lehmann
Institute 1
Department of Education
University of Hamburg
Von-Melle-Park
D-2000 Hamburg 13
Federal Republic of Germany

Dr. Pirjo Linnakyla
Institute for Educational Research
University of Jyvaskyla
Yliopistonkatu 9-11
SF-40100 Jyvaskyla 10
Finland

Mme. Emilie Barrier
Centre International d'Etudes Pedagogiques
Division d'Etudes sur les Systemes
d'Education
1, Rue Leon Journault
F-92311 Sevres
France

Dr. Georgia Polydorides
Department of Mathematics
University of Patras
Patras 26110
Greece

Dr. Cheung Yat-shing
Department of Chinese Translation
and Interpretation
Hong Kong Polytechnic
Hung Hom, Kowloon
Hong Kong

Dr. Judit Kadar-Fulop
Ministry of Culture and Education
P.O. Box 1
Szalay u 1014
H-1884 Budapest
Hungary

Professor Dr. Sigridur Valgeirsdottir
Rannsoknastofnun Uppekdusnaka
Institute of Educational Research
v/Laufasveg (Gamia Kennaraskolahusid)
IS-101 Reykjavik
Iceland

Dr. Jiyono
Balitbang Dikbud
Jalan Jend. Sudirman
Senayan
Jakarta
Indonesia

Dr. Michael O. Martin
Educational Research Centre
St. Patrick's College
Dublin 9
Ireland

Professor Dr. Pietro Lucisano
Dip. oh Ricerche Stonco-filosfiche
e Pedegogiche
Univ. di Roma "La Sapicuza"
1-00161 Roma
Italy

Dr. Kees de Glopper
S.C.O.
Grote Bickerstraat 72
NL-1013 KS Amsterdam
Netherlands

Dr. Hans Wagemaker
Research and Statistics Division
New Zealand Department of Education
P.O. Box 1379
Private Bag
Wellington
New Zealand

Professor Finn Egil Tonnessen
Stavanger Laererhogskole
Senter for Leseforsking
Box 2521
Ullandhaug
N-4004 Stavanger
Norway

Dr. Mona D. Valisno
Executive Director
Department of Education, Culture, and Sports
National Education Testing and Research
Center
Maralco Avenue (DECS complex)
Pasig, Metro Manila
Manila
Philippines

Mrs. Maria Jose Rau
Gabinete de Estudos e Planeamento
Ministerio de Educacao
Av. Miguel Bombarda 20
P-1093 Lisbon Codex
Portugal

Miss Beatrice Tay
Ministry of Education
P.O. Box 746
Kay Siang Road
Singapore 1024
Republic of Singapore

Prof. Dr. Marjan Setine
Educational Research Institute
University of Ljubljana (ERI)
Center for International Comparative Education
Gerviceva 62
P.O. Box 76
61111 Ljubljana
Slovenia

Mr. Guillermo Gil
Centro de Investigacio y Documentacion
Educativa (C.I.D.E.)
Ministerio de Educacion y Diencia
Ciudad Universitaria s/n
E-208040 Madrid
Spain

Dr. Karin Taube
Department of Psychology
University of Umea
Radhuseplanaden 2
S-90247 Umea
Sweden

Dr. Francois Stoll
Psychologisches Institut
Abt. Angewandte Psychologie
Schonberggasse 2
CH-8001 Zurich
Switzerland

Dr. Malee Nitsaisook
Susan Sunandha Teacher's College
Utong Nok Road
Dusit
Bangkok 10300
Thailand

Dr. Hyacinth E. McDowall
Curriculum Development Division
Ministry of Education
Hayes Street
St. Clair
Port-of-Spain
Trinidad and Tobago

Dr. Marilyn Binkley
U.S. Department of Education
National Center for Education Statistics
555 New Jersey Avenue, N.W.
Washington, D.C. 20208
United States of America

Dr. Armando Morles
Universidad Pedagogica Experimental
Libertador
Vicerectorado de Investigacion y Posgrado
Apartado Postal 2939
Caracas 1010
Venezuela

Mrs. Rosemary Moyana
Department of Curriculum Studies
Faculty of Education
University of Zimbabwe
P.O. Box MP 167
Mount Pleasant
Harare
Zimbabwe

Within the United States, the National Research Center was established within the National Center for Education Statistics. Dr. Marilyn Binkley was named National Research Coordinator, and a National Steering Committee was established. It included the following members:

Albert Beaton
Center for Study of Testing
Boston College
500 McGuinn Hall
Chestnut Hill, MA 02167

Marilyn Binkley
National Research Coordinator
National Center for Education Statistics
555 New Jersey Avenue, N.W.
Room 506A
Washington. D.C. 20208-5653

Margaret Smith Burke
New York University
70 Washington Square South
New York, NY 10012

Marcia Farr
University of Illinois at Chicago
P.O. Box 4348
Chicago, IL 60680

Jerome Harste
Indiana University
Bloomington, IN 47405

Barbara Kapinus
Maryland State Department of Education
200 West Baltimore Street
Baltimore, MD 21201

Thomas Kerins
Illinois State Board of Education
Student Assessment Department
100 North 1st Street
Springfield, IL 62777

Irwin Kirsch
Educational Testing Service
Room 51D
Rosedale Road
Princeton, NJ 08540

Larry Mikulecky
Learning Skills Center
Indiana University
Bloomington, IN 47405

Scott Paris
3112 School of Education
University of Michigan
Ann Arbor, MI 48109

Gary Phillips
Associate Commissioner,
Education Assessment Division
National Center for Education Statistics
555 New Jersey Avenue, NW
Room 513C
Washington, D.C. 20208-5653

Maria Ramirez
Executive Director
Center for Multinational and
Comparative Education
New York State Education Department
Room 225 EB
Washington Avenue
Albany, NY 12234

Ed Roeber
Council of Chief State School Officers
1 Massachusetts Avenue, NW
Washington, D.C. 20001

Keith Rust
Westat, Inc.
1650 Research Boulevard
Rockville, MD 20850

Ramsay Selden
Council of Chief State School Officers
1 Massachusetts Avenue, NW
Washington, D.C. 20001

Robert Tierney
Ohio State University
1945 North High Street
Columbus, OH 43210

Sheila Valencia
University of Washington
122 Miller Hall, DQ 12
Seattle, WA 98195

1.5. The Substantive Nature of the IEA Reading Literacy Study

The intent of the IEA Reading Literacy Study was

- To develop internationally valid instruments for measuring reading literacy suitable for establishing internationally comparable literacy levels in each of the participating countries;
- To describe on one international scale the literacy profiles of 9- and 14-year-olds in school in each of the participating countries;
- To describe the reading habits of the 9- and 14-year-olds in each participating country; and
- To identify the home, school, and societal factors associated with the literacy levels and reading habits of the 9-year-olds in school.

To accomplish these goals, a reading assessment instrument for students at both age levels and four sets of questionnaires (for students, their teachers, their principals, and the nation) were developed by committees working under the ISC. They were designed so that the same content would be applied in all participating countries in the appropriate languages for those countries.

The Reading Literacy Assessment

For the purposes of the IEA Reading Literacy Study, *reading literacy* was defined as

> *. . . the ability to understand and use those written language forms required by society and/or valued by the individual. Literacy occurs in a variety of language contexts (e.g., school, home, work, and religious or civic institutions) and involves both a range of competencies and a set of habits and/or practices, arrayed along various dimensions.*

Consistent with this definition, two reading assessments were developed to measure the reading proficiency of 9- and 14-year-olds. The assessments were designed to provide scaled scores that reflect students' understanding of three types of text -- narrative prose (continuous text materials in which the writer's aim was to tell a story, whether fact or fiction), expository prose (continuous text materials designed to describe or explain things), and documents (structured tabular texts, such as forms, charts, labels, graphs, lists, and sets of instructions). The assessments include questions that tapped six types of reading processes -- verbatim, paraphrase, inference, main theme, locating information, and following directions. For a full description and discussion of these instruments, see Chapter 7 of this report.

The Explanatory Variables

The four sets of questionnaires -- student, teacher, principal, and national -- were designed to collect data about those factors that were known to influence reading achievement and that might vary across nations. (The national questionnaire, which was completed by the national research team and returned to the IEA, is not discussed in this report.) Underlying these sets of questions was an implicit model of reading achievement.

After review of the available survey items, the U.S. team determined that the data could best be described in terms of two dimensions: to whom and to what they referred. In the case of the who dimension, the data describe students, their families, their teachers, and their schools. On the what dimension, the data describe their attributes, the kinds of environments provided, the forms of instruction used, and the reading behaviors they exhibited. Figure 1-1, which was developed by the U.S. team for its analysis plan, serves as a general framework for the items included in the questionnaires. For a discussion of the analyses done to relate these factors to reading achievement, see Chapter 8.

The Sample of Students

Within each of the participating countries, nationally representative samples were to be drawn based on two internationally defined target populations:

- **Population A:** All students attending school on a full-time basis at the grade level in which most students aged 9:00 - 9:11 years (during the first week of the eighth month of the school year) are enrolled.
- **Population B:** All students attending school on a full-time basis at the grade level in which most students aged 14:00 - 14:11 years (during the first week of the eighth month of the school year) are enrolled.

Within the United States, these definitions were implemented and modified in the following ways:

- **Population A:** All students attending school on a full-time basis at the grade 4 level in the 50 states and the District of Columbia, during the 1990/91 school year, who, in the opinion of school personnel, are capable of taking the test.
- **Population B:** All students attending school on a full-time basis at the grade 9 level in the 50 states and the District of Columbia, during the 1990/91 school year, who, in the opinion of school personnel, are capable of taking the test.

Using a three-stage sampling plan, approximately 200 schools were identified at each grade level with one class per school for the ninth grade, and at least one class (two, if available) for the fourth grade. The sample involved approximately 11,000 students, 470 teachers, and 332 principals. For a more detailed discussion of the sampling procedures, see Chapter 2. Procedures for determining which students were deemed not capable of taking the test are described in Section 4.4.3.1.

Figure 1-1. General framework and item map

ATTRIBUTES

STUDENT

1.
- a. Age
- b. Gender
- c. Language Of Test
- d. Ethnicity
- e. Bilingualism

FAMILY

2.
- a. Father's Education
- b. Mother's Education
- c. Family Wealth
- d. Ethnicity
- e. Bilingualism
- f. Family Size

SCHOOL

3.
- a. Principal's Experience
- b. Age
- c. Gender
- d. Ethnicity
- e. Preservice Educ. Administration
- f. Inservice Educ. Administration
- g. Years of Teaching Experience
- h. Professional Training
- i. Education in TeachingReading
- j. Inservice Reading Education
- k. General Reading Interest
- l. School Enrollment
- m. # Students/Grade/Sex
- n. Public/Private
- o. Urbanicity
- p. Resources/Activities
- q. # FTE Teaching Teachers
- r. Total Instructional Time
- s. Time School Open; Weeks/Year
- t. Days Instruction Lost
- u. % Students Absent
- v. Standardized Tests
- w. Rating Student Progress

TEACHER

4.
- a. Gender
- b. Language Of Test
- c. General Education
- d. Preservice Teacher Education
- e. Postsecondary Education
- f. Further Education In Reading
- g. Inservice Reading Education
- h. General Reading Interests
- i. Years Of Teaching Experience
- j. Aims Of Reading Instruction
- k. Attitudes To Reading Instruction
- l. Teacher Rating S Literacy
- m. Ethnicity
- n. Age
- o. Teacher Certification

ENVIRONMENT

QUALITY OF SCHOOL LIFE

5. Quality Of School Life

FAMILY

6.
- a. Meals Eaten Each Day
- b. Responsibilities/Job
- c. Newspaper In Home
- d. Hours Of TV Watching
- e. # Books In Home
- f. Wealth (Possessions)
- g. Parent Interest
- h. Parents Read (Test Language)
- i. Others Read (Test Language)
- j. Parents Read (Other Language)
- k. Others Read (Other Language)
- l. Reads To Parent (Test Language)
- m. Help With Reading Homework
- n. Read Aloud At Home

SCHOOL

7.
- a. T's Work Evaluated By P?
- b. P Discusses With T
- c. Staff Meetings; Frequency
- d. Items at Staff Meetings
- e. Community Resources
- f. Parent Cooperation
- g. Books In Library
- h. Books Added
- i. Students Borrow Books
- j. Special Reading Programs
- k. Informal Reading Programs
- l. Programs Improve Reading Instruction
- m. Reading Problems At School
- n. P's Activities
- o. P Evaluates Teachers; Frequency
- p. P Evaluation Procedures

CLASSROOM

8.
- a. T's Years Teaching Class
- b. Multigrade Class?
- c. Class Size
- d. # Non-Test-Language-Students
- e. Meet with Parents
- f. # Students Need Remedial
- g. # Students Receive Remedial
- h. Instructional Time; Total
- i. Instructional Time; Test Language
- j. Instructional Time; Reading
- k. Availability of Reading Texts
- l. Classroom Library?
- m. Classroom Library; # Books
- n. Classroom Library; Mags. Newsp.
- o. Students Can Borrow?
- p. School Library
- q. Class Visits To School Library
- r. Students Can Borrow?
- s. Students' Reading Ability

TEACHER BEHAVIORS

TEACHER INSTRUCTIONAL STRATEGIES

9.
- a. General Homework
- b. Teaching; T Directed/S Directed

TEACHER INSTRUCTION ; CONTENT

10. Time On Narr./Expos./Doc.

TEACHER READING STRATEGIES

11.
- a. Reading Homework
- b. Time Silent Reading
- c. Assigned Written Work On Reading
- d. Instruction Strategies
- e. Instructional Grouping
- f. Type Of Grouping
- g. # Of Groups
- h. Assign Reading Homework?
- i. Assign Reading Homework; Freq.
- j. Assign Reading Homework; Time
- k. Use Of 28 Reading Activities
- l. Encourage Reading Out Of School
- m. Increase S Comprehension
- n. Use Of Comprehension Strategies

TEACHER REMEDIAL STRATEGIES

12. No Questions Asked In This Area

TEACHER ASSESSMENT STRATEGIES

13.
- a. 9 Methods Assess Student Needs
- b. Assess Reading, Frequency
- c. 6 Methods Of Assessment

STUDENT BEHAVIORS AND ATTITUDES

STUDENT READING STRATEGIES

14. Students' Reading Strategies

STUDENT READING ACTIVITIES

15.
- a. Books for School/Homework
- b. Leisure Reading
- c. Magazine Reading
- d. Newspaper Reading
- e. Document Reading

STUDENT READING ATTITUDES/BEHAVIORS

16.
- a. Borrow Library Books; Frequency
- b. Self Concept Ability; Reading
- c. 3 Ways To Become Good Reader

STUDENT ATTITUDES; EDUCATION GENERALLY

17. Education Expectations

READING ACHIEVEMENTS

READING SCALES

- Narrative Comprehension
- Expository Comprehension
- Document Comprehension

1.6. The Study Timeline

Begun in 1989 in the United States, this study has had two major field operations -- a pilot study conducted in February - March of 1990, and a main data collection effort in February - March of 1991. Between the pilot test and the main study, analyses of the pilot test results were conducted in the U.S. and internationally. As a result of these analyses, instruments and procedures were revised. A complete description of these analyses is contained in Chapter 6 of the Pilot Test Report.

Reporting for this study is divided into a number of major thrusts. At the international level, the IEA has produced a series of international analyses focusing on specific topics of interest. In addition, the ICC produced an extensive academic report of the findings. These were released by the IEA during the period July 1992 through January 1994.

Within the United States, the National Center for Education Statistics also sponsored a series of publications. This report, *Reading Literacy in the United States: A Technical Report,* is designed to provide the interested reader with descriptions and discussions of how the study was carried out in the United States. It also includes descriptions of how the national data analyses reported in the more general publication, *Reading Literacy in the United States*, were conducted.

Reading Literacy in the United States: Findings from the IEA Reading Literacy Study is the first of two more general reports and is a distillation of the findings from detailed analyses of the data regarding U.S. students alone. These analyses were undertaken with a view to finding the attributes of students, families, schools, classrooms, and teachers associated with differences in student performance in this country. The second of the more general reports, *Reading Literacy in an International Perspective*, moves one step further, by looking at specific aspects of reading literacy performance in the United States as compared to selected countries that also participated in the IEA Reading Literacy Study. This report is organized as a series of very pointed papers targeting particular issues of importance.

Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study deals with methodological issues that were faced in this study and are likely to occur in other large-scale national or international comparative studies. It covers methodological issues such as sampling and sampling errors, imputation for missing data, and multilevel modeling. In addition it looks at psychometric issues regarding order effects, dimensionality, and test construction. Its purpose is to make available discussions of methodological problems and their potential solutions as they arose during the course of this study.

PART I. SAMPLE DESIGN AND DATA COLLECTION

This section of the report describes the procedures used to collect the study data and to ensure the confidentiality of the data. The procedures used were designed to be in accord with the standards for sampling and administration set forth for the study by the International Coordinating Center. At the same time, the study made use of best practices within the U.S. for sample design, administrator training, assessment administration, and data confidentiality. The next five chapters of the report describe the procedures used to collect representative samples of student, teacher, and principal data, and document the results of the sampling and data confidentiality procedures. The chapters are as follows:

- **Chapter 2: Sample Design**. *Edward Bryant, Marilyn Binkley*. A description of the procedures used to obtain a stratified probability sample of schools for each population.
- **Chapter 3: Enlistment of Schools and Class Selection Within Schools.** *Kenneth Burgdorf, David Bayless*. A discussion of the approach used to obtain the participation in the study of the sampled schools. The success of this recruitment is documented, and an analysis of the possible consequences of school nonparticipation is also provided.
- **Chapter 4: Field Data Collection.** *Kenneth Burgdorf*. A description of the procedures used to select sample classrooms from within participating schools and to administer the assessments to students, and background questionnaires to students, teachers, and principals.
- **Chapter 5: Receipt Control, Response Rates, and Processing of Raw Data.** *Kenneth Burgdorf, Edward Bryant, Nadir Atash, Stephen Roey, Valerjia Smith*. A description of procedures for returning the data to central location, coding it, and creating files ready for data analysis. This chapter includes a description of the procedures used to derive sampling weights and generate sample error estimates and documents the levels of student participation.
- **Chapter 6: Confidentiality.** *Stanley Legum, Nadir Atash*. An outline of the procedures used to ensure that the data files delivered to the International Coordinating Center protected the confidentiality of the participating schools and students.

2. SAMPLE DESIGN

2.1. Background

The IEA international studies are designed to provide comparable information about nationally representative samples of students in participating countries. All IEA studies comprise two elements: a cognitive achievement test in an agreed upon area and a set of background variables collected by questionnaire. The background variables are designed to provide an explanation of the differences within and between countries in the cognitive measure.

To make such comparisons possible, IEA attempts to design comparable samples of students across the participating countries, cognitive tests that are more or less equivalent, and survey instruments that capture important aspects related to instruction and practice within the subject area in question. This is accomplished through the development of a proposal by the organization's International Steering Committee (ISC), which is then approved by the IEA standing committee and the General Assembly. Subsequently, countries are invited to participate. Once underway, the study design is refined through consensus building processes.

Within the structure of the study design, certain aspects are obligatory in order to maintain comparability across countries. Adherence to the general principles of the sampling manual and approval of the sampling design by the sampling referee figure prominently among these. The international sampling manual leads participating countries through a step-by-step selection of schools using a single-level sampling frame based on a comprehensive national list of schools. Each school on the list is associated with a measure of school size, and schools accordingly are selected with a probability proportional to the size of their target population.

In the United States a number of practical sampling issues necessitated some departures from the procedures proposed in the IEA sampling manual (Ross 1991). First, the geographic dispersion of schools made it fiscally impossible to consider collecting data from a stratified random sample of schools. Second, because we lack a single set of national policies that would control such factors as entrance age, retention in grades, and placement in mainstream classes, the study designers in the U.S. could not identify a single grade with a clean majority of the target population.

With the view to accommodating the intent of the sampling manual within this context, project staff of the U.S. National Study tailored the sampling plan in ways that best met the specifications for the desired international target population. For example, we used a three-stage sampling frame designed to facilitate data collection, but which introduced additional clustering effects that had to be offset by increasing the sampling size, and we defined our national target population so that the modal grade for each desired age group was chosen. These modal grades contained more than 50 percent (i.e., a majority) of students of the relevant age in each case.

In the remainder of this chapter, we discuss in detail the characteristics of the IEA sampling design so that the reader can understand the constraints and goals of the sampling plan. We follow with a discussion of the sampling methods used in developing the U.S. national sample, the differences between the sampling designs for the two populations, and the way in which we have compensated for these differences through various refinements in the sample. The calculation of survey weights, sampling variances, and estimates of degrees of freedom are discussed later in this report.

2.2. Objectives of the IEA Sampling Design

2.2.1. Description of the Target Populations

The **desired international target population** definitions were prepared by the ISC to reflect the objectives of the IEA Reading Literacy Study. As is consistent with the educational research studies conducted by IEA, these desired international target populations have mainstream national education systems as their focus. Two desired international populations were defined as follows:

- **Population A:** All students attending school on a full-time basis at the grade level in which most students aged 9:00-9:11 years (during the first week of the eighth month of the school year) are enrolled.
- **Population B:** All students attending school on a full-time basis at the grade level in which most students aged 14:00-14:11 years (during the first week of the eighth month of the school year) are enrolled.

Each country developed its own defined national target population and was at liberty to tailor the definition to reflect its own needs. Although it was recognized that the defined national target population would vary from country to country and might represent only a subset of the students described by the desired international target population, it could not depart markedly from the desired international target population needs.

Factors that might have influenced the definition could have included the ability to deliver specified subsets of the eligible cohort, the desire to gather additional information for national use, or national policy constraints that might further limit access to the entire eligible cohort. The difference between the **defined national target population** and the **desired international target population** was then considered to be the **excluded population** for that country. However, decisions about the degree of divergence allowed across countries rested with the ISC and, in particular, with the sampling referee. In cases where the excluded population was deemed to be too large, the sampling referee could recommend that a country's data be excluded from the international reports.

2.2.2. Obligatory Specifications for the Sample

Based on the guidelines put forth in the sampling manual, six sampling specifications emerged.

1. **Reference point.** The age range was referenced to a particular point in the school year as opposed to a particular point in time. The ISC wished the testing date to occur when the school and class instruction was most likely to follow a regular pattern of activities. This designation of first week of the eighth month of the school year also could be seen as an attempt to limit between-nation variations in the amount of instruction received. However, given differences in policies regarding commencement age and retention, this objective was not always achieved.

2. **A single grade sample determined by age distribution.** It was the intent of the ISC that each country identify one grade level where the majority of students were of the specified age, thus simplifying administrative procedures. However, due to varying national policies concerning school commencement age and/or grade promotion

policies, there was reason to believe that there would be major differences between countries with respect to the within-grade age distributions of students. While this was seen as a possible problem for the interpretation of results, the ISC believed the benefits for study administration would outweigh the difficulties introduced in this way.

A related issue was whether the ISC intended the targeted age of students to be majority or modal. For some countries, such as Canada (as represented by British Columbia) and the United States, no single grade contained the majority of students of the specified age. Given that decisions as to grade selection were made at the country level, diversity did result. For the younger students, British Columbia chose grade 3, while the United States settled on grade 4. As an end result of this process, the age distribution across countries varied significantly.

3. **Use of intact classes.** The ISC decided to select intact classes, rather than designating individual students within schools as the sampling unit. While the ISC believed it more likely that there would be sufficient numbers of students to have reliable within-class estimates, this form of sampling increased the clustering effect, making it necessary to increase the sample size to retain the same degree of representativeness.

4. **Sampling accuracy.** IEA defined sampling accuracy as follows: "The IEA standard for sampling accuracy requires that all samples should aim to have an **effective sample size** for the main criterion variables of at least 400 students. That is, all samples should have sampling errors which are equal to, or smaller than a simple random sample of 400 students" (Ross 1991).

   This specification was designed to ensure that each country drew a large enough sample so that, with 95 percent probability, the sample estimates of population values for means, percentages, and correlation coefficients would be within +/- 0.1 of the standard deviation, +/- 5.0%, and +/- 0.1, respectively.

5. **Response rates.** On the matter of unit response, the IEA position was that "national centers should aim to obtain achieved samples that represent **at least 85 percent and preferably over 95 percent** of the designed sample of schools" (Ross 1991). IEA also allowed the use of replacement schools. The manual specified that three parallel samples were to be selected. When the first school was unavailable, it was permissible to replace it with the parallel school from the second sample, and if necessary from the third sample. In computing response rates, there is some difference of opinion as to how to account for replacements and the unknown bias that may have been introduced. This issue will be discussed in detail later in the chapter.

6. **Exclusion versus response rates.** According to the manual, "it is **not acceptable** to permit school system staff, school principals, or classroom teachers to have any influence over the selection of either a) school grades within selected schools, or b) students within selected classes." Obviously, it is possible that vested interests in selecting particular kinds of classes and/or students may lead to major distortions in sample estimates (Brickell 1974).

   Differences in policies with regard to who was in mainstream programs in various countries led to some inadvertent inconsistencies. In certain countries, for example, subgroups such as special education students were not in the mainstream program, and these students were part of the defined excluded population. Alternatively, in countries where these same subgroups were in mainstream programs, the IEA definition would lead us to include them in the sample. Clearly, these samples may not be comparable. In the United States, standard practice permits exclusion at the site level of students whose Individual Education Program (IEP) specification specifically prohibits standardized testing, and those whose native tongue is not English and who have not been in mainstream classes for at least 2 years. In order to maintain cooperation and to comply with standard policies, we followed this practice.

2.3. Objectives of the U.S. Sampling Design

2.3.1. Description of the U.S. Target Populations

Consistent with the international guidelines, staff on the U.S. study sought to define our national target population so that it matched the international target definition as much as possible. The two **defined national target populations** were as follows:

- **Population A**: All students attending school on a full-time basis at the grade 4 level in the 50 states and the District of Columbia during the 1990-91 school year who, in the opinion of school personnel, were capable of taking the test.

- **Population B**: All students attending school on a full-time basis at the grade 9 level in the 50 states and the District of Columbia during the 1990-91 school year, who, in the opinion of school personnel, were capable of taking the test.

  The term "school" in these definitions referred to regular public and private schools; that is, those schools offering solely special education or employing other nontraditional teaching methods involving no recognizable grade structure or those teaching by correspondence were regarded as being ineligible under the desired international target population definitions.

The **excluded populations** were defined in the following ways:

- **Population A:** The **excluded population** consisted of fourth grade students in the five U.S. territories of American Samoa, Guam, Northern Marianas, Puerto Rico, and the Virgin Islands. This group included 62,900 students in 1,060 schools, and represented 1.6 percent of the desired national target population. In addition, an estimated 188,600 students in eligible schools were known to be incapable of taking the test (4.9 percent of the desired national target population).

- **Population B:** The **excluded population** consisted of ninth grade students in the five U.S. territories of American Samoa, Guam, Northern Marianas, Puerto Rico, and the Virgin Islands. These 61,500 students in 341 schools, represented 1.8 percent of the desired national target population. In addition, an estimated 165,900 students in eligible schools were known to be incapable of taking the test (4.9 percent of the desired national target population).

The excluded populations contained two distinct portions. First, students in the five U.S. territories were excluded because of operational difficulties associated with administering the assessment in these far-flung territories and because students in Puerto Rico, who composed 90 percent of the total fourth grade population of these territories, speak and are educated in Spanish. However, the distant states of Alaska and Hawaii, often excluded from the target population of U. S. national surveys, were included in this case. The second portion of the excluded population consisted of those students incapable of taking a pencil and paper assessment in English. These students either were handicapped, or, because their native language was other than English, had limited proficiency in English. The determination as to whether a particular student should be excluded was made by school personnel under the proviso that when in doubt, the student was to be assessed.

In the U.S., as in other countries participating in the IEA Reading Literacy Study, the "desired population" comprised the defined and the excluded populations for the specific age group (Table 2-1). Data to compute the population sizes were obtained from the U.S. Bureau of the Census, the National Center for Education Statistics, and the Quality Education Data (QED) file.

Table 2-1. United States population: Desired, defined, and excluded populations

Population	Desired		Defined		Excluded	
	Schools	Students	Schools	Students	Schools	Students
A	64,660	3,835,500	63,800	3,584,000	1,060	251,500
B	21,306	3,390,300	20,965	3,162,900	341	227,400

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

2.3.2. Considerations in Defining the U.S. National Target Population

Reference Point. The importance of establishing a date to indicate the beginning of the school year for study purposes became obvious as we started to define the U.S. national target population. The IEA definition for each of the populations of interest was based on a particular time period in the school year. While in some countries determination of the beginning of the school year may not be problematic, in the U.S. each state and/or school district is responsible for its own calendar. Thus, there is a great deal of variation among schools in terms of their calendar; in fact, some schools operate a year-round calendar.

The variation in school calendar caused problems in terms of determining a date that could be considered as the beginning of the school year for all U.S. schools. Although reliable data are not available, it is known that some schools in the U.S. each year begin classes around the second week of August, whereas many other schools start around the first week of September. It is hard, therefore, to reach a consensus among educators as to the beginning of the school year for U.S. schools. For the purposes of defining the appropriate age distribution, we simply decided to consider the period from the third week of March to the first week of April as roughly corresponding to the desired first week of the eighth month of the school year.

A Single Grade Sample Determined by Age Distribution. Given the IEA target definition of age, we looked at all available information about the age distribution for 9- and 14-year-olds across grades. Direct information about the number of students of a given age who were in a particular grade were available only from limited sources. The 1980 Census gave the numbers for age as of April 1, 1980

(Table 2-2), while the 1988 NAEP provided data about the numbers for students age 9 and 13 in grades 4 and 8, respectively, at different points during 1988 (Table 2-3). NAEP data also gave an estimate of the grade distributions for 9- and 13-year-olds as of January 1, 1988. The annual October supplement of the *Current Population Survey*, produced by the Bureau of the Census, gives the information for age in years defined in early October. Thus, we had no way of establishing directly what were the modal grades for students aged 9 and 14 at the beginning of the eighth month of the school year.

As one notes by reviewing the data from the 1980 Census of Population as presented in Table 2-2, 9- and 14-year-olds were enrolled in a wide band of grades. Clearly no single grade contained the majority. However, there is clear indication that the modal grades for 9- and 14-year-olds were fourth and ninth, respectively, although, it is equally evident that large numbers were also present in the third and eighth grades.

Table 2-2. School enrollment for persons 9 and 14 years of age: 1980

Age of student	Grade in which student is enrolled									
	2		3		4		5		6	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
9 years (Pop. A)	197,371	5	1,616,763	43	1,806,517	49	73,571	2	7,322	*
Age of student	6		7		8		9		10	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
14 years (Pop. B)	34,259	1	267,001	7	1,560,064	41	1,797,251	48	83,681	2

*Less than 1 percent.

SOURCE: Excerpted from Detailed Population Characteristics, 1980 Census of Population.

Table 2-3. Percentage of students aged 9 and 13 years who are in grades 4 and 8, respectively, at the start of the each month: 1988

Students	January	February	March	April
Age 9; Grade 4	60.9	54.8	48.9	42.3
Age 13; Grade 8	59.3	53.2	47.5	41.0

SOURCE: NAEP estimates for 1988.

A slightly different way of determining which grade was most appropriate can be shown using the NAEP 1988 grade samples. These data show that a large percentage of students who were 9 and 13 years of age were in the fourth and eighth grades. While the percentage declined across subsequent months, the figure was still significant as late as April.

While it is clear that there is no obvious grade where a majority of 9- and 14-year-olds are best represented, it is clear based on all available data, and the IEA guidelines, that grades 4 and 9 were a reasonable choice for the study.

Use of Intact Classes. Consistent with the IEA guidelines, the U.S. sampled intact classes rather than selecting individual students within schools. In fact, this decision made it much easier to obtain cooperation from the schools. Two reasons were offered. First, this method was less disruptive to the school schedule. Second, it minimized the impact on teachers and students. In the fourth grade sample, the decision was made to sample two classes in schools where two or more classes were present. Since we did not know the number of classrooms per grade initially, the decision rule was "if there are fewer than an estimated 50 fourth grade students in the school, take all. If 50 or more, sample two classrooms at random." For the ninth grade, a single classroom was selected per school.

2.4. Sample Design and Sampling Procedures

2.4.1. Introduction

The survey was designed to collect test scores and information on student, teacher, school, and family characteristics, family environments, school environments, classroom environments, instructional strategies, and student reading activities and behaviors on a sample of fourth grade students (Population A) and ninth grade students (Population B). The first stage sample was drawn from the primary sampling units (PSUs) constructed for the NAEP surveys, after the changes in stratification described below. The sample was allocated to the strata in proportion to 1980 population, which was the basis for construction of the NAEP PSUs. The schools in the sample of PSUs were further stratified by student enrollment in the fourth or ninth grade (the two populations were handled independently) and by public and private control.

The structure of the sampling design differed somewhat from the models suggested by the international referee (Ross 1991). The U.S. adopted the approach, approved by the referee, of arranging for personnel from outside the school system to administer the assessments. This approach was taken to maximize school participation by minimizing the burden on schools and to assist in maintaining uniformly high standards of assessment administration throughout the sample by using field workers who were trained as a group by study staff. In most other countries, school personnel administered the assessments in the interest of minimizing costs.

The basic U.S. sample plan called for sampling intact classrooms and/or classes. For grade 4, if a sample school had fewer than an estimated 50 fourth grade students, all were included. In schools with 50 or more fourth graders, two classrooms were taken at random. For grade 9, in schools with fewer than an estimated 25 ninth grade students, all were included. Otherwise, the plan called for taking one classroom (typically, the language arts class). The number of students in the grade was estimated by dividing the total enrollment, as reported on the Quality Education Data (QED) file, by the grade span of the school.

2.4.2. Stage I Stratification of Schools

The NAEP PSUs are counties (or independent cities) and groups of counties with a minimum population of 60,000 as of the 1980 Census. The counties composing metropolitan areas are kept together; other aggregations avoid mixing urban and rural counties.

The IEA specifications did not require certain estimates by subgroups (such as minorities) that were mandated by NAEP. Hence, the NAEP PSUs were restratified for use in the IEA study. The first level stratification was by NAEP region (four geographic strata) and two degrees of urbanization strata (Metropolitan Statistical Area -- MSA -- and non-MSA). In addition, the Southeast and West regions were stratified by percent minority, those with less than 20 percent minorities in one class and those with 20 percent or more in another. Fourteen PSUs were of sufficiently large size that it was appropriate to include them in the sample with certainty. Minorities (outside of the large cities, included with certainty) are relatively less prevalent in the Northeast and the Central regions, so the minority stratification was not used in those regions (Table 2-4).

Table 2-4. Allocation of sample PSUs to strata

Region	Urbanicity	Certainty	Minority	Number of PSUs
Northeast	MSA	Certainty	All	7
		Noncertainty	All	4
	Non-MSA	Noncertainty	All	2
Southeast	MSA	Certainty	High	2
		Noncertainty	High	2
		Noncertainty	Low	2
	Non-MSA	Noncertainty	High	2
		Noncertainty	Low	2
Central	MSA	Certainty	All	3
		Noncertainty	All	6
	Non-MSA	Noncertainty	All	4
West	MSA	Certainty	High	2
		Noncertainty	High	4
		Noncertainty	Low	4
	Non-MSA	Noncertainty	All	4
Total PSUs				50

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The high minority, non-MSA stratum in the West contained so few PSUs that it was combined with the low minority, non-MSA stratum. It was possible to subdivide them by percent minority in the second stage of stratification. A sample of 50 PSUs in total was drawn according to the above allocation. Sampling weights equal to the inverse of the probabilities of selection were attached to them.

2.4.3. Stage II Stratification of Schools

The schools in the sampled PSUs were extracted from the Quality Education Data (QED) file and were substratified by stage II strata, which, in some cases, cross-cut the first level of stratification (Tables 2-5 and 2-6). The two stage II stratifying variables were type of control (public schools in one class; Catholic and other private in the other class) and enrollment in the fourth grade for Population A or the ninth grade for Population B. It was presumed that the distinction between private and public schools was so important that the design should adequately represent the relatively thin population of private schools and the large number of small schools with small enrollments, an objective that could not be reached without some cross-cutting of the major strata. The schools were put into three classes at Population A and two classes at Population B on the basis of their estimated grade enrollment. The amount of collapsing of first stage stratifying factors necessary to effect the second stage of stratification is evident from the tables. Note that the last stratum in each table consisted of the large number of

Table 2-5. Substrata for Population A

Sub-stratum number	NAEP stratum characteristics				Substratum		Number of schools	
	Region[1]	Urbanicity	Minority level[2]	Certainty status	Ownership	School size[3]	Sample	Population[4]
1	Northeast	MSA	NA	Certainty	Public	15-49	2	1,029
2	Northeast	MSA	NA	Certainty	Public	50+	12	2,627
3	Northeast	MSA	NA	Certainty	Private	15+	4	1,685
4	Northeast	All	NA	Noncertainty	Public	15-49	6	2,268
5	Northeast	All	NA	Noncertainty	Public	50+	13	3,221
6	Northeast	All	NA	Noncertainty	Private	15+	3	1,408
7	Southeast	MSA	High	Certainty	All	15+	4	915
8	Southeast	MSA	High	Noncertainty	Public	15+	9	2,282
9	Southeast	MSA	Low	Noncertainty	Public	15+	11	2,323
10	Southeast	All	All	Noncertainty	Private	15+	4	1,579
11	Southeast	Non-MSA	High	Noncertainty	Public	15+	8	1,920
12	Southeast	Non-MSA	Low	Noncertainty	Public	15+	9	2,393
13	Central	MSA	NA	Certainty	Public	15+	7	2,067
14	Central	MSA	NA	Certainty	Private	15+	2	782
15	Central	All	NA	Noncertainty	Private	15+	5	2,304
16	Central	MSA	NA	Noncertainty	Public	15-49	4	1,880
17	Central	MSA	NA	Noncertainty	Public	50+	14	3,718
18	Central	Non-MSA	NA	Noncertainty	Public	15-49	8	3,106
19	Central	Non-MSA	NA	Noncertainty	Public	50+	6	1,728
20	West	MSA	High	Certainty	All	15+	9	2,081
21	West	All	All	Noncertainty	Private	15+	4	1,696
22	West	MSA	High	Noncertainty	Public	15+	17	3,543
23	West	MSA	Low	Noncertainty	Public	15+	19	4,383
24	West	Non-MSA	All	Noncertainty	Public	15-49	5	1,538
25	West	Non-MSA	All	Noncertainty	Public	50+	7	1,630
26	All	All	All	Noncertainty	All	<15	8	10,408

[1] Region definitions (note that these region definitions are those used by NAEP and hence were used for forming strata for the Reading Literacy Study).

Northeast
Connecticut
Delaware
District of Columbia
Maine
Maryland
Massachusetts
New Hampshire
New Jersey
New York
Pennsylvania
Rhode Island
Vermont
Virginia (the part in Washington, DC MSA)

Southeast
Alabama
Arkansas
Florida
Georgia
Kentucky
Louisiana
Mississippi
North Carolina
South Carolina
Tennessee
Virginia (outside Washington, DC MSA)
West Virginia

Central
Illinois
Indiana
Iowa
Kansas
Michigan
Minnesota
Missouri
Nebraska
North Dakota
Ohio
South Dakota
Wisconsin

West
Alaska
Arizona
California
Colorado
Hawaii
Idaho
Montana
Nevada
New Mexico
Oklahoma
Oregon
Texas
Utah
Washington
Wyoming

[2] Minority level of PSU only used in Southeast and West regions: Low = less than 20%, High = 20% or more.

[3] Enrollment in the given grade estimated by dividing the school enrollment for the school as listed in the Quality of Education Data (QED) file by the number of grades in the grade span of the school.

[4] Tabulated from the QED file.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 2-6. Substrata for Population B

Sub-stratum number	NAEP stratum characteristics				Substratum		Number of schools	
	Region[1]	Urbanicity	Minority level[2]	Certainty status	Ownership	School size[3]	Sample	Population[4]
1	Northeast	MSA	NA	Certainty	Public	15+	18	961
2	Northeast	MSA	NA	Certainty	Private	15+	3	599
3	Northeast	MSA	NA	Noncertainty	Public	15+	14	1,265
4	Northeast	MSA	NA	Noncertainty	Private	15+	3	453
5	Northeast	Non-MSA	NA	Noncertainty	All	15+	5	726
6	Southeast	MSA	High	Certainty	All	15+	4	278
7	Southeast	All	All	Noncertainty	Private	15+	3	882
8	Southeast	MSA	High	Noncertainty	Public	15+	9	750
9	Southeast	MSA	Low	Noncertainty	Public	15+	12	680
10	Southeast	Non-MSA	High	Noncertainty	Public	15+	10	1,003
11	Southeast	Non-MSA	Low	Noncertainty	Public	15+	9	1,078
12	Central	MSA	NA	Certainty	All	15+	10	619
13	Central	All	NA	Noncertainty	Private	15+	3	602
14	Central	MSA	NA	Noncertainty	Public	15+	22	1,695
15	Central	Non-MSA	NA	Noncertainty	Public	15+	14	2,826
16	West	MSA	High	Certainty	Public	15+	9	471
17	West	All	All	Noncertainty	All	15+	2	588
18	West	MSA	High	Noncertainty	Private	15+	18	857
19	West	MSA	Low	Noncertainty	Public	15+	19	1,103
20	West	Non-MSA	All	Noncertainty	Public	15+	11	1,863
21	All	All	All	Noncertainty	All	<15	2	4,088

[1] Region definitions (note that these region definitions are those used by NAEP and hence were used for forming strata for the Reading Literacy Study).

Northeast
Connecticut
Delaware
District of Columbia
Maine
Maryland
Massachusetts
New Hampshire
New Jersey
New York
Pennsylvania
Rhode Island
Vermont
Virginia (the part in Washington, DC MSA)

Southeast
Alabama
Arkansas
Florida
Georgia
Kentucky
Louisiana
Mississippi
North Carolina
South Carolina
Tennessee
Virginia (outside Washington, DC MSA)
West Virginia

Central
Illinois
Indiana
Iowa
Kansas
Michigan
Minnesota
Missouri
Nebraska
North Dakota
Ohio
South Dakota
Wisconsin

West
Alaska
Arizona
California
Colorado
Hawaii
Idaho
Montana
Nevada
New Mexico
Oklahoma
Oregon
Texas
Utah
Washington
Wyoming

[2] Minority level of PSU only used in Southeast and West regions: Low = less than 20%, High = 20% or more.

[3] Enrollment in the given grade estimated by dividing the school enrollment for the school as listed in the Quality of Education Data (QED) file by the number of grades in the grade span of the school.

[4] Tabulated from the QED file.

NOTE: NA - Not applicable.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

schools with small enrollments. A relatively thin sample of such schools was drawn to increase the efficiency of the design, since the per-student assessment costs for such schools were high. This had the effect of increasing the weights of the small schools so that their effect on national projections was proportionate to the total enrollment of the stratum. The sample of 200 schools from each population was allocated to the deeply stratified universe in proportion to the number of students in the given grade projected from the sampled PSUs, since, at the time the sample was drawn, total counts for the universe were not available in time to meet the deadline for the design work. This required a later adjustment in the sampling weights, as is discussed in Section 5.5.

2.4.4. Sample Selection

The schools chosen for the sample were coded by substratum number and given a measure of size that reflected the way in which the within-school sample was to be drawn (see Tables 2-5 and 2-6). The measures of size were determined by multiplying the PSU weight by the average per-school enrollment for the schools in the defined class as follows.

Population A

Enrollment under 15
    Measure of size is 7.6 times the PSU weight.

Enrollment at least 15, but under 50
    School is in an MSA and is Private
        Measure of size is 26 times PSU weight.
    School is not in an MSA and is Private
        Measure of size is 38 times PSU weight.
    School is not in an MSA and is Private
        Measure of size is 21 times PSU weight.
    School is not in an MSA and is Public
        Measure of size is 29 times PSU weight.

Enrollment is 50 or greater
    Measure of size is enrollment size times PSU weight.

Population B

Enrollment is under 15
    Measure of size is 7.9 times PSU weight.

Enrollment is 15 or greater
    Measure of size is enrollment size times PSU weight.

After assignment of the measures of size, the samples were drawn with probability proportional to size within the substrata after selecting with certainty any school with measure of size equal to or greater than three-fourths of the sampling interval. Schools were given a probability of inclusion of one divided by the revised sampling interval, after exclusion of the certainty selections. The product of the within-substratum probability of selection and the PSU probability of selection is the overall probability of selection of the schools. The sampling was done using WESSAMP, Westat's proprietary package for sample selection. This software also provided the overall probability of selection of each of the schools. The base weight of each selected school is the inverse of the probability of selection. These base weights were adjusted for school nonresponse (see Chapter 5).

As required by the sampling referee, checks were made on the selected sample of schools and their base weights to ensure that the samples had been drawn without error. By stratum, the weighted measures of size of the selected schools were summed and then compared with the total of the measures of size for the stratum. They agreed exactly in each case, as was appropriate.

2.4.5. Correspondence of Sample Design and Selection Procedures to International Requirements

The sampling manual (Ross 1991) imposed certain minimum requirements on the sampling procedures used to draw the samples of schools and students. The manual specified that stratification of schools be implemented to the maximum extent practicable, but with the limitation that there be at least two schools selected per stratum, in order to permit unbiased estimation of sampling errors. The variables suggested as possible stratifiers were region, urbanization, socioeconomic status, school type, school size, and school program. As Tables 2-4, 2-5, and 2-6 show, the U.S. design was heavily based on a stratified sample using four of the six strategies suggested in the sampling manual. The classification of school program (academic/vocational) was, for the most part, not applicable at the school level in the U.S. Because the frame of schools did not have complete, reliable, and valid information on socioeconomic status, all four stratifiers that were available and appropriate were incorporated into the U.S. design. In addition the stratification of the geographic PSUs included a high-low minority population classification, which has been previously found to be an important stratifier for U.S. samples.

The sampling manual also included instructions as to how to draw the sample of schools from the frame. The procedure used to draw the U.S. samples were exactly in accord with the directions for drawing a three-stage probability proportional to size (PPS) sample of geographic PSUs, schools, and classrooms included in the manual. The U.S. sample did vary from the specified procedures in that no back-up sample of schools was drawn. In the U.S. the preferred procedure for handling initial nonparticipation by the original school sample involved a two-pronged approach. The first was to use extensive recruitment followup procedures of the initial refusals in an effort to include them in the study. These efforts were successful in increasing the participation rate from below 70 percent to 87 percent (see Table 3-2). The second procedure involved the use of weighting adjustments to reduce the potential bias from both school and student nonparticipation (see Section 5.5). This approach is a well-recognized procedure for compensating for nonresponse in sample surveys (Little and Rubin 1987, Chapter 4).

2.4.6. Correspondence of Sample Design and Selection Procedures to Those Used in the National Assessment of Educational Progress

This study was similar to the ongoing National Assessment fo Educational Progress national studies in that both are assessments of educational achievement of school students. The methods and materials for designing and selecting the samples of schools and students, and for defining eligible schools and students, were very similar across the two studies, with one exception. NAEP does not select samples of classrooms, but rather samples individual students directly from within participating schools. The procedures for data collection and sample weighting used in this study (see Chapters 3 through 5) are also very similar to those adopted by NAEP.

2.4.7. Sampling for Order Effect Study

To test the impact of item ordering, a small experiment was incorporated within the overall study design for students in grade 4. Within each class where the test was administered, one randomly selected student was given a test booklet that used an alternate ordering of the questions. Data from this small study has been analyzed to assess the impact of item ordering on classical and IRT item statistics (e.g., p-values, discrimination in index, Rasch difficulty values). A report on this analysis is available from Westat, Inc.

References

Brickell, J.L. (1974). Nominated samples from public schools and statistical bias. *American Educational Research Journal*, 11 (4), 333-341.

Little, R.J.A.,.and Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Ross, K.N. (1991). *Sampling manual for the IEA international study of reading literacy.* University of Hamburg, Hamburg, Germany: International Coordinating Center, IEA International Study of Reading Literacy.

3. ENLISTMENT OF SCHOOLS AND CLASS SELECTION WITHIN SCHOOLS

3.1. School Enlistment

The National Center for Education Statistics began its efforts to gain support for the IEA Reading Literacy Study through presentations to the Council of Chief State School Officers' (CCSSO) Education Information Advisory Council (EIAC). The EIAC has the responsibility within CCSSO of reviewing proposed data collection activities for proper justification, reasonable data burden, appropriate methodology, consistency with state and local record keeping practices, and value of proposed uses of the data collected. At four EIAC meetings over a 2-year period beginning in September 1989, NCES representatives described the preliminary plans, the pilot test, and implementation plans for the main test and sought input from state representatives. As a result of the presentation of information and discussion of the activity with the state representatives at those meetings, EIAC endorsed the study and encouraged its members to participate fully in all activities.

According to the specifications of the IEA, those who would conduct the Reading Literacy Study should first obtain permission to test in the schools. In the U.S., because the school system is decentralized and locally autonomous, this requirement necessitated adhearance to a protocol of contacting several levels of government officials. First, responsibility for education constitutionally rests with the individual states. Secondly, most states have legislatively authorized local school districts to operate the schools in their area. Finally, the school organization has a principal as the chief administrative officer of the building complex responsible for activities within it. Consequently, in the U.S., it was necessary to secure permission from the chief state school officer, local district superintendent, building principal, and the classroom teacher prior to conducting the IEA Reading Literacy Study at each location. The following sections describe the processes used to accomplish this task.

3.1.1. Contacting States

Following EIAC approval, Westat sent a letter on October 13, 1990, to the chief state school officer in each of the 31 states with schools selected to participate in the study sample and the District of Columbia. On October 23, 1990, the CCSSO directed a letter to the EIAC representative and testing/evaluation representative of each of the states involved requesting their assistance in identifying and anticipating any problems in working with the local districts selected as part of the study sample and seeking their cooperation in addressing those problems. Although the ICC provided prototypical letters for such contacts, the governmental and administrative hierarchy unique to the U.S. school system made it more appropriate to use materials written specifically for this country.

In early December 1990, each state education agency was contacted by telephone, first to ensure that the state had granted permission to pursue the study, and then to determine the method or protocol that should be followed in contacting local school districts. Most state agencies advised Westat to contact the districts directly. Some state education agencies wanted to make the first contact with the sampled districts themselves, followed by a more complete explanation of the study methodology and district requirements from Westat. In Louisiana and Wyoming, the state agency made all the arrangements with the district.

3.1.2. Contacting Districts

The initial district contact was made in the form of a letter from Westat to the districts. This letter, generated on a flow basis as soon as permission was received from the individual states beginning in December 1990, provided a general description of the study, identified the agencies involved in conducting the U.S. portion of the study, and referenced the authorization of the chief state school officer to conduct the study in that state. Additionally, a packet of materials was included that described the study and the participants worldwide (the informational brochure produced by the National Center for Education Statistics), listed the requirements imposed on participating schools, specified the details of the study coordination between the schools and Westat, listed the schools in the district selected as part of the sample, and requested the cooperation and participation of the district. The letter indicated that Westat would contact the local school district in one week to make arrangements with schools.

It was not unusual to place a minimum of three to five calls before having the opportunity to talk with the district superintendent. In a large district, the process was further complicated because materials were often referred to other central office staff, and it was necessary to determine who had the information before anyone could make a decision concerning the district's participation. Although almost all districts that agreed to participate premised their decision on the subsequent concurrence of the selected schools, some districts refused to grant district permission until they had the opportunity to discuss the activity and its implications with the principals of these schools.

The reason given most often by districts that declined participation was that the cumulative loss of instructional time to all tests being given at the school was so significant that the district was forced to reject any additional intrusion not currently planned or required by regulation. Figure 3-1 is a cause and effect diagram that reflects some of the major reasons given for nonparticipation. The diagram reflects the results of an initial brainstorming session of the Westat enlistment staff to identify the major reasons for nonparticipation, followed by an analysis of the call records maintained for each district and school to quantify the content and incidence of those reasons (causes) for nonparticipation.

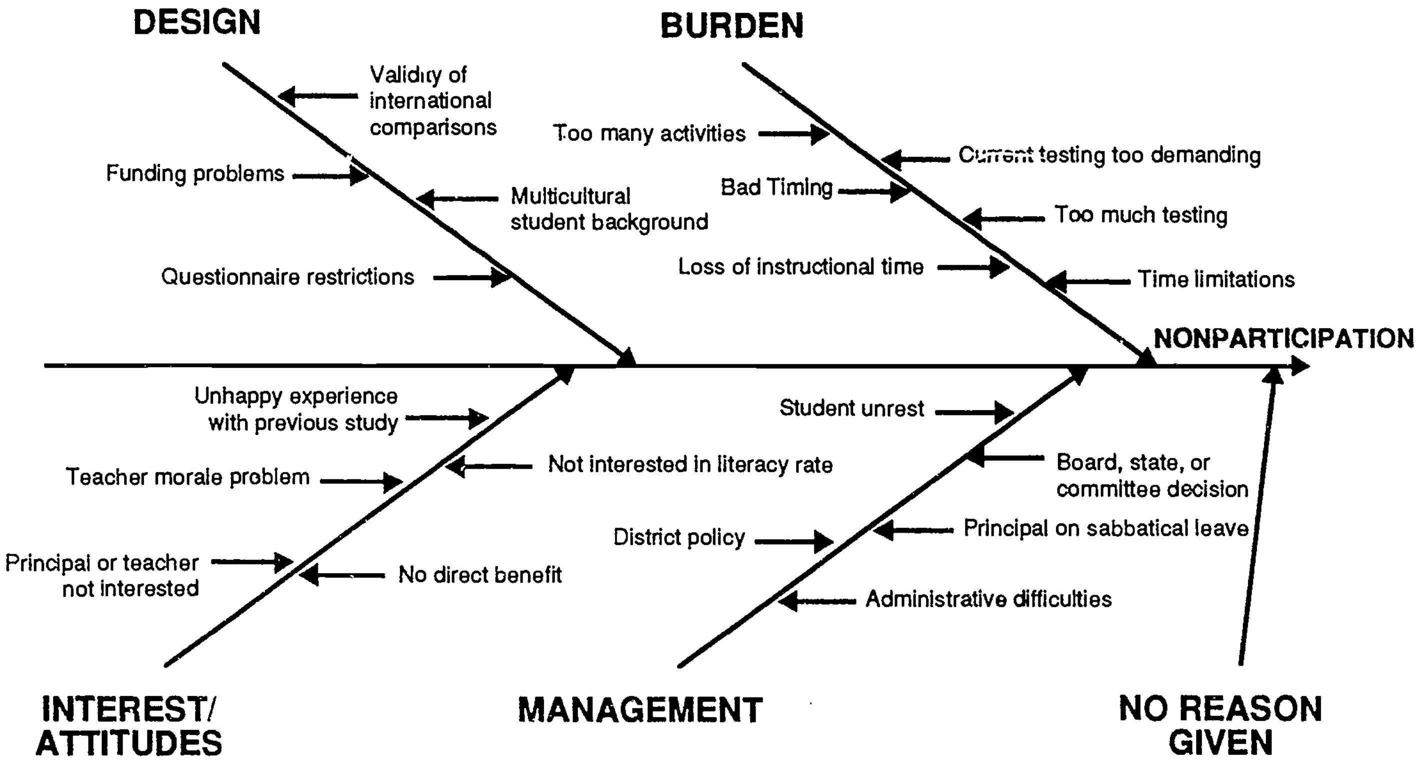
Invariably, the individuals contacted acknowledged the importance of the study and the opportunity to generate useful new knowledge for practitioners but felt that local circumstances forced them to "permit other districts the privilege of generating this new information." The most widespread disappointment among both participating and nonparticipating districts was the inability of the study design to produce school- or district-level data that could be useful in assessing and comparing their individual programs to national and international results.

3.1.3. Contacting Schools

Officials of the districts that agreed to participate were asked to verify, or provide, names of principals and addresses of the schools within their districts selected to be part of the sample. The district representatives were also asked if they wished to coordinate the study activities centrally, assume responsibility for contacting the schools, or give Westat the permission to contact the schools directly. At the request of several districts, letters were prepared that were addressed to the principals but sent to the district administrator, who in turn hand delivered them to the principals. Additionally, some school districts requested that liaison between the survey staff and school personnel in their district be handled centrally by a coordinator designated by the central office.

As suggested by the IEA specifications, the letter to the principal described the general objective and design of the study and specified the participation requirements for that particular school.

Figure 3-1. Cause and effect diagram of district nonparticipation in the IEA Reading Literacy Study



The letter indicated that after Westat received a list of classes in the appropriate category, a Westat staff member would be in contact with the school coordinator to begin the process of scheduling the school visit and to inform the school of the class(es) selected. The school principal was told that after the schedule was arranged, a packet of materials containing Teacher and School Questionnaires would be mailed. The principal was asked to complete the School Questionnaire, to distribute the Teacher Questionnaire to the teachers of the selected classes, and to have both completed and available for the assessment administrator to collect during the visit. A packet of materials containing the following items accompanied this letter:

- A School-Westat Coordination Procedures description - An instruction sheet asking the principal to name a school coordinator to function as the primary contact between the school and the IEA Reading Literacy Study staff and to compile a list of appropriate classes in the school using forms provided.
- A Fourth/Ninth Grade Class List Form - A form asking for names and identifying information for all eligible classes within that school. The class name, class identifier, teacher name, and number of students in the class were requested. This Class List Form was used to select the sample of the class(es) participating in the study.
- A Federal Express package and preaddressed label for use in returning the Class List Form.

Followup telephone calls were made to each school principal (or district coordinator) receiving these materials requesting his/her approval to proceed and seeking to respond to any questions that may have arisen. During the telephone call, the plan to select a sample of one or two classes from the school was reiterated and emphasis was placed on the importance of receiving the list of eligible classes to draw the sample. It is noteworthy that none of the district or school personnel contacted indicated that a selection of more than one classroom in his/her school would create a problem or influence the decision to allow the school to participate in the study. This position was true of the nonparticipating schools as well as those that agreed to take part in the study.

Recognizing the short time frame and the intrusion into normal school operations, Westat encouraged the schools to provide information that already existed (even in a different format), to use FAX technology when available, and even to provide necessary sampling information over the telephone.

As soon as Westat received the Class List Form from the school, the sample was selected and the Westat field supervisor selected tentative dates for a school visit that would expedite the scheduling of area assessment administrators. Using information and instructions compiled from the earlier district contacts, a Westat scheduler contacted the school to arrange for the visit dates, negotiate changes, and seek any information that would facilitate the assessment administrator's visit and enhance his/her contact and communication with the school coordinator and classroom teachers.

3.1.4. Refusal Conversion

Many districts and schools that had initially refused participation were persuaded to reconsider this decision and to participate in the study. Due to schedule constraints, conversion efforts were almost exclusively implemented by telephone. Westat contacts in state education agencies assisted in the process of persuading districts and schools to participate and were influential in converting original refusals in 16 states. The NCES project officer also assumed an active and successful role in aiding the reconsideration of refusals by sampled districts and schools. Additionally, a concerted effort at refusal conversion by Westat staff, as the enlistment period drew to a close, resulted in a combined participation rate of 86.5 percent (unweighted), increasing the original response rate by over 20 percent (see Table 5-1).

3.1.5. School Participation Rate

Of the 32 jurisdictions for the sample, 15 states had no schools refusing participation in the study (Table 3-1). An additional 6 states had only one refusing school and 5 states had only 2 refusing schools--consequently 26 of the 32 jurisdictions had refusals from 2 or fewer schools in their jurisdiction. Of the 17 schools that were out of scope, 8 had reconfigured their organizational pattern and no longer had the grade of interest (e.g., changing from a junior high school with grades 7-9 to a middle school with grades 6-8), 5 schools had closed, 2 schools were small ungraded schools without fourth grade equivalent students, and the remaining 2 schools were excluded because of unique school circumstances that caused them to be nonrepresentative.

School participation by grade and type of school is reflected in Table 3-2. The percentage of variation was less than 1/2 of 1 percent between fourth and ninth grade public school participation. A greater differentiation occurred in private schools, but over a much smaller number of schools. As expected, participation rates at the elementary level were slightly higher than at the ninth grade level, but the differential was actually less than expected. We assumed that elementary schools would be more likely to participate as the three-period requirement for administering the instruments created greater scheduling difficulties and intrusion on schedules for departmentalized secondary schools than for self-contained elementary programs.

3.2. Class Selection and Assessment Scheduling

Following enlistment of the school, participating schools were asked to use the Class List Form to indicate all fourth grade classes or English/language arts classes that contain ninth grade students in their school. For each class, they were asked to provide an identifier (room number, period number, section number, etc.), teacher name, and the number of students. Using these forms, operations staff began making phone calls to arrange assessment sites and dates.

The sampling design called for selecting intact classrooms and/or classes as follows:

- Fourth Grade - If fewer than 50 students in the school, take all. If 50 or more, sample two classrooms at random.
- Ninth Grade - If fewer than 25 students, take all. If 25 or more, take one ninth grade language arts class.

Using these general criteria, the staff used a random number table to select the appropriate number of lines from the form to represent the sample from that school. In a subsequent phone call, the school coordinator was told which class(es) had been selected, and asked to schedule a session date that would be consistent with the school calendar and the assessment administrator's schedule.

Table 3-1. School participation by state

States	Approvals				Refusals				Out of scope				Total schools in sample					
	Public		Private		Public		Private		Public		Private		Public		Private		Overall	
	4th	9th	4th	9th	4th	9th	4th	9th	4th	9th	4th	9th	4th	9th	4th	9th	4th	9th
ALABAMA	4	4	1	0	0	0	0	0	0	0	0	0	4	4	1	0	5	4
ARKANSAS	5	5	0	0	0	1	0	0	0	0	0	0	5	6	0	0	5	6
CALIFORNIA	15	14	3	1	3	4	0	0	0	2	1	0	18	20	4	1	22	21
COLORADO	3	2	0	0	1	2	0	1	0	0	0	0	4	4	0	1	4	5
CONNECTICUT	6	6	0	2	2	0	0	0	0	0	1	0	8	6	1	2	9	8
DISTRICT OF COLUMBIA	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
FLORIDA	8	7	1	2	0	0	0	0	0	0	0	0	8	7	1	2	9	9
GEORGIA	2	1	0	0	0	0	0	0	0	1	0	0	2	2	0	0	2	2
IDAHO	4	4	0	0	0	0	0	0	0	0	0	0	4	4	0	0	4	4
ILLINOIS	6	6	3	1	0	1	0	0	0	0	0	0	6	7	3	1	9	8
INDIANA	1	2	0	0	2	1	0	1	0	0	0	0	3	3	0	1	3	4
IOWA	4	4	1	1	0	0	0	0	0	0	0	0	4	4	1	1	5	5
KANSAS	4	3	0	0	0	0	0	0	0	0	0	0	4	3	0	0	4	3
LOUISIANA	4	5	2	1	0	0	0	0	1	0	0	0	5	5	2	1	7	6
MARYLAND	1	2	2	0	1	1	0	0	0	0	0	0	2	3	2	0	4	3
MASSACHUSETTS	2	2	0	0	0	1	0	0	0	0	0	0	2	3	0	0	2	3
MICHIGAN	4	5	0	2	0	0	0	0	0	0	0	1	4	5	0	3	4	8
MINNESOTA	2	2	0	0	0	0	0	0	0	0	0	0	2	2	0	0	2	2
MISSOURI	2	3	2	1	1	1	0	0	1	0	0	0	4	4	2	1	6	5
MONTANA	3	3	0	0	0	0	0	0	0	0	0	0	3	3	0	0	3	3
NORTH CAROLINA	8	7	0	0	0	0	0	0	0	1	0	0	8	8	0	0	8	8
NEW JERSEY	4	2	1	0	1	3	1	1	0	1	0	0	5	6	2	1	7	7
NEW YORK	8	9	1	1	3	6	1	0	2	0	1	0	13	15	3	1	16	16
OHIO	6	8	0	0	1	0	0	0	0	0	0	0	7	8	0	0	7	8
PENNSYLVANIA	0	1	1	1	1	1	0	0	0	1	0	0	1	3	1	1	2	4
SOUTH CAROLINA	3	4	0	1	1	1	0	0	0	0	0	0	4	5	0	1	4	6
TEXAS	14	17	2	0	3	1	0	0	0	0	0	0	17	18	2	0	19	18
UTAH	7	6	0	0	0	0	0	0	0	0	0	0	7	6	0	0	7	6
VIRGINIA	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	1
WEST VIRGINIA	5	6	0	0	0	0	0	0	1	0	0	0	6	6	0	0	6	6
WISCONSIN	5	6	3	0	1	0	0	0	0	1	0	0	6	7	3	0	9	7
WYOMING	3	3	0	0	0	0	0	0	1	0	0	0	4	3	0	0	4	3
Total	144	151	23	14	22	24	2	3	6	7	3	1	172	182	28	18	200	200

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 3-2. School participation by grade and type

Control	Approvals	Refusals	Out of scope	Total	Percent of eligible schools participating
Schools	332	51	17	400	87
Public					
Fourth grade	144	22	6	172	87
Ninth grade	151	24	7	182	86
Total public schools	295	46	13	354	87
Private					
Fourth grade	23	2	3	28	92
Ninth grade	14	3	1	18	82
Total private schools	37	5	4	46	88
Total fourth grade	167	24	8	200	87
Total ninth grade	165	27	8	200	86

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

3.3. Analysis of the Willingness to Participate Index (WTPI)

To study possible distortions or biases due to schools chosen for the sample deciding not to participate in the IEA Reading Literacy Study sample, Westat performed a profile analysis based on a Willingness to Participate Index (WTPI) among sampled schools that was developed for the study. The analysis was based on assigning the schools that participated in the sample into three categories.

Each individual enlistment case was rated by staff involved in the enlistment process for that school and its parent district, that is, by experienced Westat data collection staff using the field collection telephone calling records as a primary data source. The rating was based on the level of effort required to secure participation. The ratings were subjectively assigned using the following general categories:

Easy - Both the district and the school agreed to participate, and few logistical or administrative problems were encountered, with only a moderate effort required to convince one or the other of the importance of justifying an intrusion on their program.

Moderate - Effort was required by caller or enlistment contact to convince either the district or the school that their participation was worth their efforts and time. More than one call or contact was required to talk to all concerned personnel, explain the study, answer questions, or overcome administrative difficulties before receiving authorization to conduct the study.

Difficult - Substantial efforts were required to overcome initial refusals, including alternative strategies such as enlisting support from states and outside agencies and seeking to make any accommodations that would make the study acceptable.

The school characteristics analyzed and studied were school type (public/private), type of community (rural, small town, or large city[1] -- as measured on the Principal Questionnaire) and percentage of nonwhite student enrollment (0 to 10 percent, 11 to 60 percent, or 61 to 100 percent--which also was collected on the Principal Questionnaire).

The first issue analyzed was whether distribution of WTPI was relatively similar for schools within the analysis variables. Figures 3-2 and 3-3 provide profiles of the schools using the Willingness to Participate Index for grades 4 and 9, respectively, by school type, community type, and percentage of nonwhite student enrollment.

The major conclusion drawn from the profiles is that there was no substantial difference in profiles of schools (in terms of WTPI distribution) across categories of variables under consideration. However, the following minor differences were observed:

1. The private sample schools in grade 4 were very willing to participate (only 4 percent of the sample schools were categorized as difficult in terms of gaining participation);
2. For the private schools in grade 9, however, 33 percent of the sample schools were difficult to enroll in the study;
3. The grade 9 schools with a large enrollment of nonwhite students were difficult to enlist in the study (31 percent were categorized as difficult to gain participation versus 21 percent for the U.S.).
4. A slightly larger percentage of the schools in grade 4 with a enrollment of 61 percent or larger of nonwhite students was categorized as easy to gain participation (64 percent versus 48 percent for the U.S.).
5. Among the large city schools in grade 4, 59 percent (versus 48 percent for the U.S. overall) were easy to enlist.

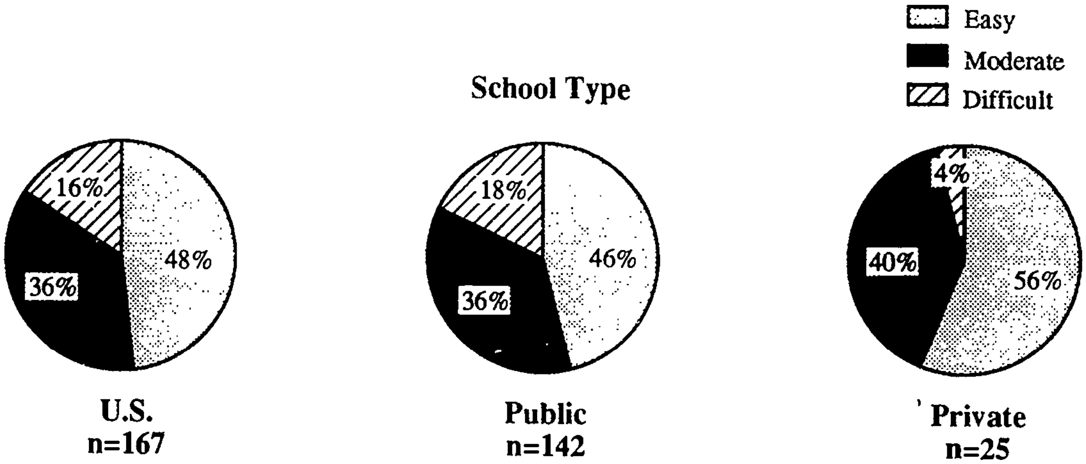
Figure 3-4 provides the reason for refusal by the major categories of the cause and effect diagram (refer to Figure 3-1). Burden seemed to be the most prevailing reason for schools refusing to participate.

---

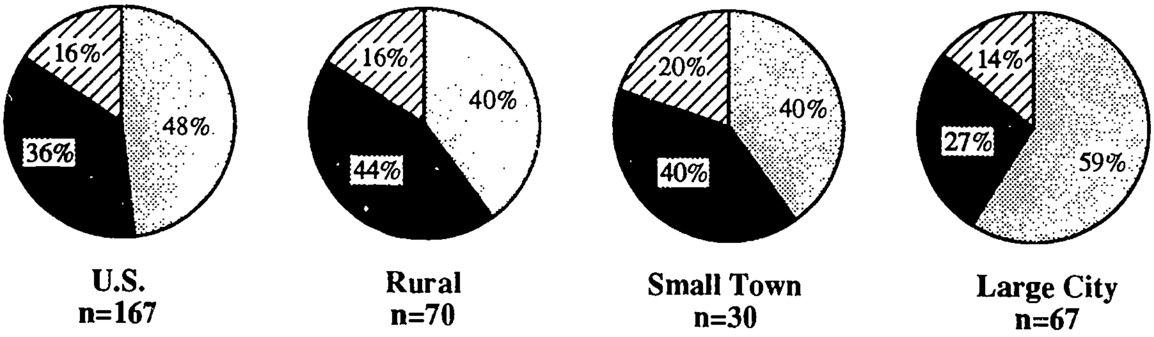
[1] For a description of these categories, see Chapter 9, Section 9.3.4.

Figure 3-2. Profiles of Willingness to Participate Index by school type, community type, and percentage of nonwhite student enrollment: Grade 4

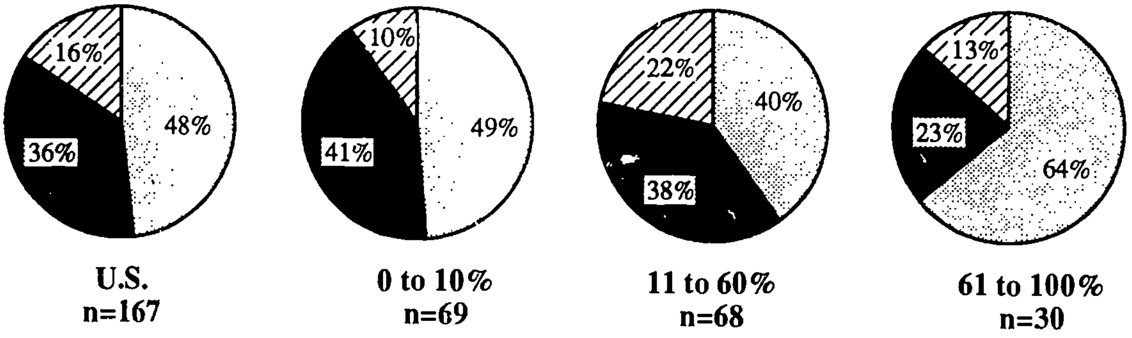
School Type



Community type

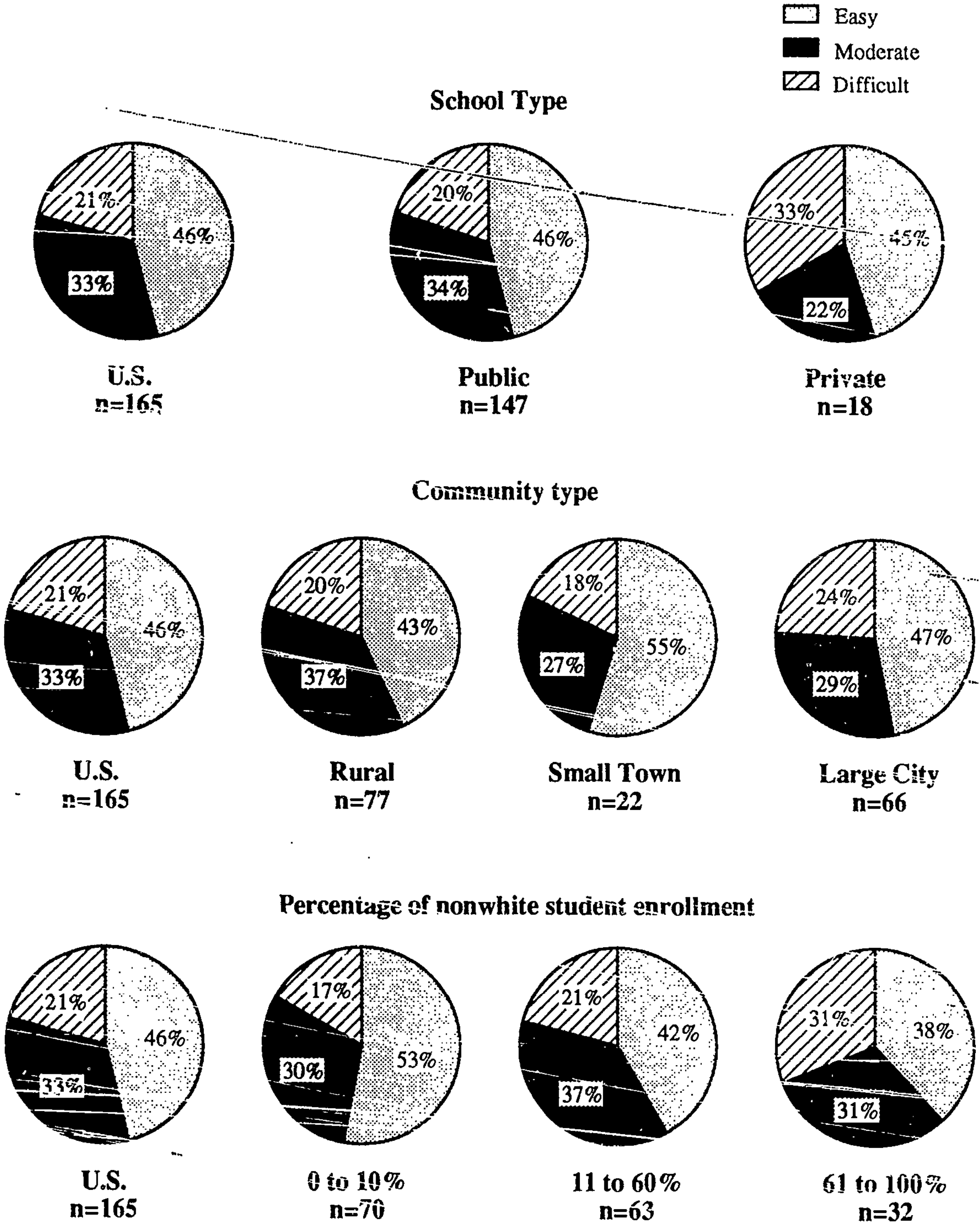


Percentage of nonwhite student enrollment



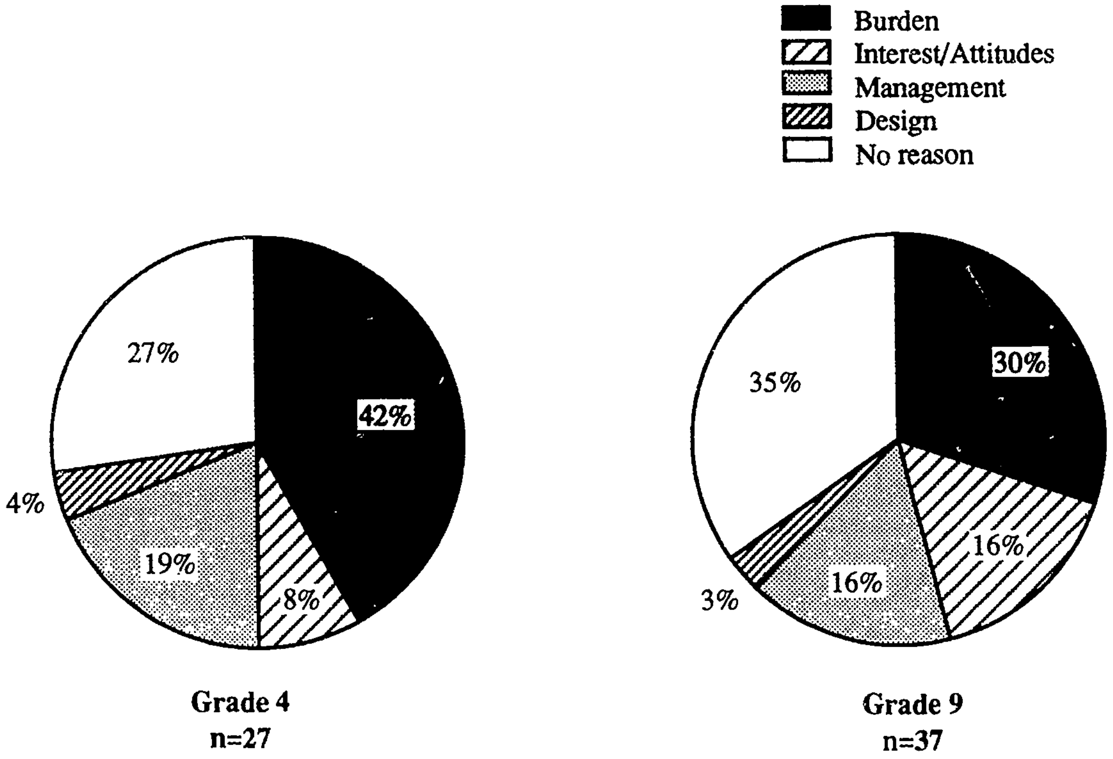
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 3-3. Profiles of Willingness to Participate Index by school type, community type, and percentage of nonwhite student enrollment: Grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 3-4. Profile of reasons for nonparticipation

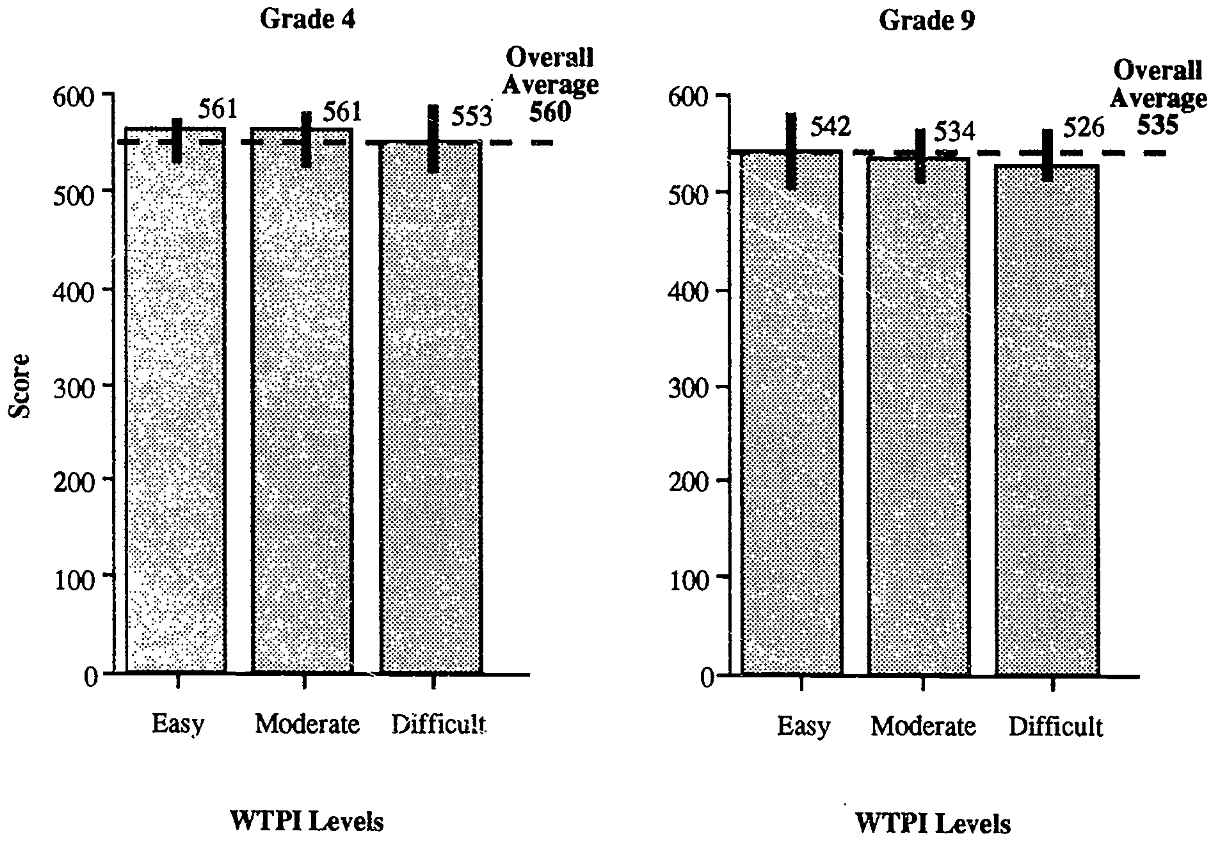


SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The next question we analyzed was whether student performance varied by the subgroups defined by the three categories of the WTPI. The measure used for student performance in this analysis was the narrative scale score. It will be especially interesting to compare student performance in schools categorized at the difficult level of the WTPI (which can be considered to be near refusal schools) to the student performance in schools at the easy level of the WTPI. If no difference exists among these levels, it provides some suggestive evidence that little distortion can be attributed to school nonparticipation.

Figure 3-5 provides a graphic display of the average student performance for each of the three levels of the WTPI for both grades 4 and 9. The confidence interval bands associated with each average clearly demonstrate that variability of the averages across the three levels of the WTPI may be due to chance or random causes. This finding, based on WTPI analysis, suggests that there may be little distortion in the IEA Reading Literacy Study data due to willingness of the schools to participate in the study. There are no data from the schools that did not participate in the study (the refusals) to compare with the data of those that did participate in order to conduct a more comprehensive analysis of the bias due to nonresponse. If the assumption that refusal schools were similar to difficult-level schools is tenable, it can be concluded that nonresponse bias, if any, would be quite small.

Figure 3-5. Average narrative achievement by Willingness to Participate Index (WTPI) levels



Sample sizes n= 167 schools (overall)
n= 80 Easy
n= 61 Moderate
n= 26 Difficult

Sample sizes n= 165 schools (overall)
n= 76 Easy
n= 54 Moderate
n= 35 Difficult

NOTE: The estimated standard error of the mean narrative scores (a measure of the variation due to sampling) has been used to exhibit the precision of average narrative scores across the three school categories. If all possible samples were surveyed under similar conditions, intervals of 1.965 standard errors below to 1.965 standard errors above the mean would include the average result of these samples in approximately 95 percent of the cases. For example, for the estimated average narrative scores for Population A schools categorized as "Easy," the 95 percent confidence interval ranges from 542 to 574. If the above procedure were followed for every possible sample, about 95 percent of the intervals would include the average number from all possible samples. This confidence interval is shown by the black band at the top of the columns.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

4. FIELD DATA COLLECTION

In conducting the U.S. component of the IEA Reading Literacy Study, Westat collected data on approximately 7,200 students in the fourth grade and 3,800 students in the ninth grade at 332 public and private schools. The sample included schools distributed in 227 districts across 31 states and the District of Columbia.

4.1. Field Plan

The ICC specifications permitted participating countries to choose field administrators from a range of categories, including classroom teachers, school administrators, and nonschool personnel. In considering these options, the U.S. study team felt that the study would be better served by the creation of a field staff that was no way associated with the schools themselves. The primary benefit would be that the assessment administrators could be trained together and would subsequently administer the test to all students in a standardized manner. It was felt that data collected in this way would be far more comparable than that collected under an infinite number of differing conditions. In addition, using study staff rather than school personnel would reduce the burden of response and might thereby increase the rate of participation.

To complete the data collection in a cost-effective manner and within the time parameters established by the International Steering Committee, study personnel decided that a field staff of approximately 45 assessment administrators and 2 supervisors was needed. The staffing was based on several assumptions:

- A 4-week field period with an additional week allocated to accommodate schools that required rescheduling or that were slow in agreeing to participate in the study;
- A 95 percent institution participation rate with 70 percent of the participating schools agreeing to single day visits and 30 percent requesting that testing be done over 2 days; and
- An average of three completions per week per assessment administrator.

The field plan included training Westat staff, supervisors, and assessment administrators involved in the data collection effort.

4.2. Recruiting Field Staff

Two experienced supervisors were recruited and each was assigned a geographical area. The supervisors in turn selected assessment administrators. Westat home office supplied the supervisors with suggested lists by state, PSU, and area. Westat used field staff drawn from a large pool of individuals with extensive field experience, including staff who worked for Westat on previous studies of this type (e.g., National Assessment of Educational Progress). Additional personnel were obtained by Westat's established recruiting networks and procedures that are used to provide qualified candidates for its other field studies and from the national field staff who have worked on other education studies. Supervisors began the recruitment of field staff, called assessment administrators, in December 1990, with all hiring completed by the end of January 1991.

4.3. Supervisor and Assessment Administrator Training

4.3.1. Training Plan

In developing the training plan, the following issues were considered:

- The size of the field staff warranted conducting two training sessions because the smaller the field staff to trainer ratio, the more likely the desired objectives would be achieved. Westat believed that cost-effective, optimum training could best be achieved with training sessions involving no more than 25 trainees. Smaller training sessions would enable the assessment administrators to participate more fully and provide an environment that encouraged all trainees to ask questions and clear up problems. The smaller size of the training also would help the trainer monitor each trainee's performance more effectively.
- Because of the distribution of assignments across the country, training should occur at an eastern and western site in cities that were easily accessible (Washington, D.C., and Los Angeles were chosen) and should be scheduled over weekends in order to qualify for the substantial discounts offered by airlines and hotels.
- All training materials should be scripted to ensure that the material presented and the manner of presentation was consistent across two sessions.
- Active participation of the NCES project officer in the development of the training package would be critical in order to ensure that a training package incorporating her insights and ideas be finalized by the scheduled training dates.
- Westat staff would present a full walk through of training approximately 3 weeks prior to the training sessions in order to provide NCES the opportunity for review of the entire training session and related materials so that any revision could be incorporated into the package well in advance oi training. In fact, Westat staff presented these materials to the NCES project officer on January 23, 1991, at the Westat home office, and revised them to incorporate the participants' comments.

4.3.2. Field Manual

The ICC allowed each country to devise its own training program using field manuals and other training materials as the study staff saw fit. In the U.S. an assessment administrator's manual was developed to provide each administrator with general information about the study and detailed instructions for contacting schools, preparing for the assessment, conducting the assessment, and completing necessary post-assessment sessions. All procedures were to be used as a guide by the assessment administrator to ensure that all field staff handled cases as consistently as possible.

The manual included the following general materials:

- An introduction to the purposes, goals, history, general study design, study schedule, management arrangements, and overview of administrator's tasks.

- A general description of behaviors required of anyone representing Westat and the National Center for Education Statistics. Included in this section were attitude and conduct considerations, required handling of materials and supplies, confidentiality requirements for instruments and data, and individual accountability.
- A review of the number and nature of previous contacts made with the state, district, and school.
- Step-by-step instructions for preparing the assessment sessions, including confirming the scheduled visit, school contacts, schedule and locations of sessions, policies and procedures for handling disruptive students, excluded students, etc.
- Step-by-step instructions for conducting the assessment sessions, including completing the administration schedule, preparing the test site, instructing students, monitoring the session, answering student questions, and handling problem situations.
- Step-by-step instructions for completing post-assessment activities, including completing the Administrative Schedule and the School Field Log and editing and shipping materials.
- Exhibits in the manual included checklists and scripts and copies of administration forms required by the various activities.

4.3.3. Other Training Materials

In support of the training sessions and the field manual, additional materials were developed and distributed in advance of the training sessions. Home Study Exercises to be completed prior to the training session provided administrators an opportunity to assess their overall knowledge of the activity after reviewing the manual. These exercises, which were reviewed at the training session, included true-false statements about administrator tasks, short answer questions concerning questions that were likely to arise, a request for the administrator's description of the study instruments, a list of actions to be completed, and finally a piece asking how the administrator would respond to the items on a list of potential student comments.

In addition, role-play activities were prepared that would cover the following scenarios:

- Meeting school study coordinators;
- Completing the administration schedule;
- Distributing and labeling of student-specific test materials; and
- Conducting the assessment.

4.3.4. Training

A training session for the two supervisors, conducted on February 6, 1991, at the Westat home office, covered the administrative aspects of the supervisors' responsibilities, their role in training, and an abbreviated version of the training their field staff would receive.

The two training sessions for the field staff were held February 8-10 and February 15-17. The sessions were built around the use of three basic training techniques--interactive lecture exercises, reviewing material and procedures presented in the lecture sessions, and role-play activities that emphasized test administration.

Training topics included an overview of project background, contacts with school coordinators, conducting assessment sessions, field procedures, quality control, and administrative responsibilities. The project director, assistant director, field manager, and federal project officer participated in the training sessions.

4.4. Site Visits

4.4.1. Producing and Distributing Field Assignments

Shortly after training, the assessment administrators met with their supervisors to review the assignments and make final arrangements for receipt of all necessary materials. Materials and supplies for the first several scheduled schools were packaged for assessment administrators to take with them; additional supplies were shipped to their homes or some other designated destination. The following materials were included in the packets:

> **Field Assignment Summary Sheet** (Exhibit 4-1). This summary identified all schools in the field group, both those scheduled and those not yet scheduled. It provided summary information such as school name and address, grade level, and confirmed visit dates.
>
> **Field Schedule** (Exhibit 4-2). The schedule, in a calendar format, contained all of the schools in the assignment with confirmed appointments. The schedule was updated as additional appointments were confirmed.
>
> **School Contact Sheet** (Exhibit 4-3). This computer-generated sheet was school specific. It provided information obtained during previous calls to the school, including the names and telephone numbers of contacts at the school, any special arrangements that had been made, on-site scheduled information, and space for summary information of any problems or unusual circumstances encountered while completing the assignment.
>
> **Administration Schedule** (Exhibit 4-4). This form was used to record the names of the students enrolled in a sampled class regardless of whether or not they attended the scheduled session. This form provided space to record the student's session status (present, absent, or excluded) and, for students who were identified as excluded, to record the reason for exclusion. The form also included space for recording session appointment places and times, teacher ratings of student reading literacy, and race/ethnicity.

Exhibit 4-1. Field assignment sheet

IEA READING LITERACY STUDY

FIELD ASSIGNMENT SUMMARY SHEET
FIELD GROUP

WESID	SCHNAME	SCHADDRESS	CITY	CONF. DATE(S)	GRADE	# OF ASSESS.	SCHOOL COORDINATOR	TITLE
A001	Bond Elementary School	1411 Hope Road	Camegie	2/25-26/91	4	2	Ms. Jan Brown	Principal
A002	Jones Bridge Elementary School	1205 Cherry Drive	Bridgeville	3/1/91	4	2	Mr. Tom Calhoun	Lang. Arts. Coord.
B002	Alvin Wake Jr. High	12 School Drive	Stone Rocks		9	1	Mr. McCraken	Guidance Counselor
A004	Carson Elementary	600 Parkway West	Greentree	3/5-6/91	4	2	Ms. Linda Goshen	Teacher
A005	Pine Hills Elementary	3200 Carson Street	Pittsburgh	2/18/91	4	1	Ms. Thelma Dill	Lead Teacher
B003	Raven Gap High School	1007 Long Lane	Pittsburgh	2/20/91	9	1	Mr. Sam Dosier	Principal

NOTE: All entries in this exhibit are fictitious.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Exhibit 4-2. Field schedule

FEBRUARY - MARCH 1991

SUN	MON	TUE	WED	THU	FRI	SAT
FEBRUARY	11	12	13	14	15	16
17	18 A005 PINE HILLS ELEMENTARY	19	20 B003 RAVEN GAP HIGH SCHOOL	21	22	23
24	25 A001 BOND ELEM. SCHOOL	26	27	28	MARCH 1 B002 JONES BRIDGE ELEM. SCHOOL	2
3	4	5 A004 CARSON ELEMENTARY	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

NOTE: All entries in this exhibit are fictitious.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Exhibit 4-3. School contact form

IEA READING LITERACY STUDY

SCHOOL CONTACT FORM

---

* CONTACT INFORMATION *

FIELD GROUP:

WESID: GRADE:

INSTITUTION NAME:

ADDRESS:

CITY: STATE: ZIP:

PRINCIPAL NAME:

COORDINATOR NAME:

COORDINATOR TITLE:

LOCATION OF COORDINATOR: PHONE #:

VISIT DATE(S): VISIT TIME:

---

* ASSESSMENT ARRANGEMENTS *

CLASS 1	CLASS 2
SCHEDULED DATE:	SCHEDULED DATE:
SCHEDULED TIME:	SCHEDULED TIME:
Session 1:	Session 1:
Session 2:	Session 2:
Session 3:	Session 3:

SPECIAL PROCEDURES:

---

* SCHOOL SUMMARY INFORMATION (A.A. COMPLETED) *
(e.g., pep rally, early dismissal, etc.)

NOTE: All entries in this exhibit are fictitious.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Exhibit 4-4. Administration schedule

ADMINISTRATION SCHEDULE

4th Grade, A001
Pottowa Elementary
Class 2
Ms. Mildred Marples, 2

Assessment Administrator: ______ ID#: ______

Session # 1______
Day/Date:______
Time:______
Location:______

Session # 2______
Day/Date:______
Time:______
Location:______

Session # 3______
Day/Date:______
Time:______
Location:______

Student Name		RACE ETHN	LIT LEVEL	WESTAT ID	SESSION STATUS P-PRESENT A-ABSENT E-EXCLUDED			REASON EXCLUDED (6 Spec. Ed.) (7 Non-Eng)
LAST	FIRST				1	2	3	
01				A00122-001-3				
02				A00122-002-2				
03				A00122-003-7				
04				A00122-004-8				
05				A00122-005-2				
06				A00122-006-9				
07				A00122-007-7				
08				A00122-008-0				
09				A00122-009-1				
10				A00122-010-0				
11				A00122-011-1				
12				A00122-012-9				
13				A00122-013-0				
14				A00122-014-9				
15				A00122-015-3				
16				A00122-016-7				
17				A00122-017-2				
18				A00122-018-9				
19				A00122-019-4				
20				A00122-020-5				

NOTE: All entries in this exhibit are fictitious.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Reading Literacy Tests (Attachment A-1). The reading test was the primary data collection instrument for this study. The test had two forms, one for fourth grade and the other for ninth grade.

Student Questionnaires (Attachment A-2). Questionnaires included items on student/parent background information such as parent's educational level, language spoken at home, student reading activities, etc. There was a separate questionnaire for fourth and ninth graders.

Teacher Questionnaires (Attachment A-3). Questionnaires were used to collect information on school and classroom policy, instructional approaches used by the teacher, and the teacher's educational background and experience.

School Questionnaires (Attachment A-4). Questionnaires were completed by the school principal or person designated by the school principal asking for information on school demographics, school policies and resources, and evaluation of instruction. One questionnaire was to be obtained from each participating school.

School Field Log (Exhibit 4-5). This log was the control form for the School Questionnaire and the Teacher Questionnaire(s). The name of the designated respondents to the questionnaires, as well as the final status of each questionnaire, was recorded in the log. Additionally, the log reported the expected class size and provided space for recording actual class size, number of excluded students, and number of absent students.

Assessment administrators were asked to plan their assignments by reviewing their field schedule calendar and the Field Assignment Summary Sheet. If an administrator could foresee any potential problems, he/she contacted the supervisor well in advance of the scheduled date.

4.4.2. Confirming Appointments

At least 3 days prior to the scheduled arrival at the school, the assessment administrator was supposed to contact the school coordinator by phone to confirm the arrangements. The assessment administrator was to accomplish four purposes with the call: to introduce him/herself, to verify the school visit dates, to inquire about the session times that had been arranged by the coordinator, and to arrange to meet with the coordinator prior to the session and answer any questions about the study and/or the scheduled visit. If there was a problem with the date scheduled for the school visit, the assessment administrator was to notify the supervisor immediately.

4.4.3. Site Visit

4.4.3.1. Initial Meeting With Coordinator

At the start of a site visit, the assessment administrator checked into the school office and identified him/herself as a representative of the IEA Reading Literacy Study. The assessment administrator then met with the school coordinator to review the schedule and arrangements for the assessments. If the coordinator was not available, the assessment administrator asked to see the principal. If neither school official was available, the session was rescheduled. The problem was documented on the School Contact Form and the supervisors were notified immediately.

Exhibit 4-5. School field log

4th GRADE, A

BOND ELEMENTARY SCHOOL, A001

Assessment Administrator ______________________ ID # ____________

# of Students (Class 1)		** # of Students (Class 2)	
Expected........	____	Expected........	____
Actual..........	____	Actual..........	____
Excluded........	____	Excluded........	____
Absent..........	____	Absent..........	____

RESPONDENT NAME		WESTAT SCHOOL ID	SCHOOL QUESTIONNAIRE STATUS CM-Complete RF-Refused OT-Other
LAST	FIRST		
		A001	

1 Class, 1 Teacher

TEACHER NAME		WESTAT CLASS/TEACHER ID	TEACHER QUESTIONNAIRE STATUS CM-Complete RF-Refused OT-Other
LAST	FIRST		
		A00111	

OR 2 Classes, 2 Teachers

TEACHER NAME		WESTAT CLASS/TEACHER ID	TEACHER QUESTIONNAIRE STATUS CM-Complete RF-Refused OT-Other
LAST	FIRST		
		A00111	
		A00122	

**Class 2 data on number of students will only appear for schools with 2 classes.

NOTE: All entries in this exhibit are fictitious.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Assessment administrators used the School Coordinator Initial Meeting Checklist (Exhibit 4-6) to ensure that they asked the appropriate questions and to confirm all assessment procedures. They obtained student lists for the classes chosen for assessment and completed the Administration Schedule, a five-part form providing multiple copies to be used as documentation. The form was designed so that student names appeared only on the first copy, which was left at the school in order to protect both privacy and confidentiality.

4.4.3.2. Excluding Students

During the initial meeting, the coordinator and assessment administrator determined which students appearing on the class roster should be identified as "excluded" on the Administration Schedule. For this study, a student was excluded from the assessment only for the following two reasons:

- A student was enrolled in a special education program and had an Individual Educational Plan (IEP) that specifically prohibited pencil-and-paper assessment; or
- A student was non-English speaking and had been enrolled in a mainstream English class for less than 2 years.

Table 4-1 shows the results of the process permitting the exclusion of certain students. In total, 183 students were excluded from the grade 4 sample and 18 students from the grade 9 sample. As can be seen from the table, the excluded students constituted very small proportions of the respective target populations. Note that the weighted estimates of population percentages were obtained by applying the weighting procedures, described in Chapter 5, to both the nonexcluded and excluded students in similar fashion. Note that the selection probability for a student who is excluded is the same as the selection probability of the assessed students in the same classroom of students.

Table 4-1. Number and weighted percentages of students excluded

Reason for exclusion	Grade 4		Grade 9	
	Number excluded	Weighted percent excluded	Number excluded	Weighted percent excluded
IEP with learning or physical disability	72	1.1%	6	0.2%
LEP with insufficient English language skills	111	1.1	12	0.4
Total excluded	183	2.2	18	0.6

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

4.4.3.3. Conducting Assessments

After meeting with the coordinator, the assessment administrator evaluated the designated testing area to make sure that it had all the required facilities. After determining that the testing area was properly arranged, supplies were set out and the assessment was ready to begin.

In conducting the student assessments, assessment administrators were urged to project a professional, friendly manner and, to the extent possible, to minimize the amount of disruption to the school day. As demonstrated by the timetables for the administration of the questionnaires and test (Exhibit 4-7), maintaining the schedule was a formidable task.

Exhibit 4-6. School coordinator checklist

IEA Reading Literacy Study
School Coordinator
Initial Meeting Checklist

______ 1. Confirm specific scheduled times for assessment sessions and record this information on the Administration Schedule. Ideal schedule is:

	4th Grade	9th Grade
Session 1	35+ min	45+ min
Break	≥15 min	≥10 min
Session 2	35+ min	45+ min
Break	≥15 min	≥10 min
Session 3	25+ min	25+ min

______ 2. Has coordinator notified the sampled classes' teachers and students of the scheduled sessions and enlisted the assistance of the classroom teacher during the assessments? If no, ask him to do so.

______ 3. Have the details of the 3rd session, administration of the Student Questionnaire, been worked out? (Refer to section 4.2 of the Field Manual if necessary.)

______ 4. What are the procedures for handling students who refuse to participate either before or during a session and students who show up after a session has begun?

______ 5. What options are available for dealing with disruptive students, and what is the school's preferred method for dealing with such situations?

______ 6. Will the classroom teacher be available to monitor the students during the test sessions and planned breaks? If not, what are the exact procedures you should follow when providing breaks between sessions?

______ 7. What procedures should you follow if it is necessary to excuse a student to the rest room during a session? (Does the school have some sort of monitoring system, (e.g., hall pass, sign out sheet, etc.), which you should follow?)

______ 8. Will you be given completed School and Teacher Questionnaires before you leave the school? If not, provide Westat labels and ask for a date on which the forms are expected to be completed.

______ 9. Obtain and review class rosters asking the coordinator to identify those students who should be excluded from the assessments. Be sure to get a reason for exclusion. (Refer to section 4.2 in the manual for definitions of exclusion.)

______ 10. On the class roster, highlight the names of those students identified as excluded or nonparticipating because of parental refusal; make sure the coordinator understands that these students are not to be at the assessment and that they are the school's responsibility.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Exhibit 4-7. Assessment timetables

Fourth Grade Timetable

Session 1 (45 minutes)

Part I	Word Recognition	1 1/2 minutes
Part II	Main Test (a):	35 minutes

Session 2 (45 minutes)

Part III	Main Test (b):	35 minutes

Session 3 (30 minutes)

Part IV	Student Questionnaire:	25 minutes

Ninth Grade Timetable

Session 1 (50 minutes)

Part I		45 minutes

Session 2 (50 minutes)

Part II		45 minutes

Session 3 (25 minutes)

Part III	Student Questionnaire:	25 minutes

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The procedures that follow were designed to provide a step-by-step approach to meeting this challenge:

- Using the Administration Schedule, the assessment administrator took the roll.
- He/she then used the script/checklist as a guide to introduce the students to the IEA Reading Literacy Study, distribute the forms and other materials, determine when the instructions were to be read verbatim to the students, administer the Reading Test and the Student Questionnaire, and collect all materials and conclude the session.
- As he/she had been instructed prior to the testing session, during the testing sessions the teacher recorded rating of student reading literacy levels and race/ethnicity on the student roster that had been used to construct the Administration Schedule. The task was accomplished according to the instructions in the Coding Guidelines (Exhibit 4-8). The assessment administrator was responsible for ensuring that the teachers fully understood the definitions of reading literacy proficiency levels before they rated the students.
- The assessment administrator wrote any concluding summaries or observations on the School Contact Form. All forms and materials were accounted for at this time and prepared for shipping to the home office. The assessment administrator then determined whether a makeup session was necessary by computing a response rate. If the number of absent students was greater than 20 percent of the total number of students who should have taken the test (see Section 4.4.3.4), the supervisor was contacted to determine if a makeup session was necessary.

4.4.3.4. Administering the Student Questionnaire

Each set of classroom sessions involved approximately 25 students, each of whom completed the Reading Literacy Test and the Student Questionnaire. Before the instruments were distributed, the assessment administrator affixed an ID label on each that exactly matched the ID number next to the student's name on the Administration Schedule. Once the assessment administrator left the school, this ID number was the only means of associating a student with the questionnaire and test he/she completed. The four copies of the Administration Schedule that left the school had a preprinted Westat ID number corresponding to each student name space so that when a student name was recorded on the schedule, an ID number was automatically assigned.

At the start of each session, the assessment administrator told students that they would be given no explicit assistance in responding to the material once the sessions started. They were told that time limits were specified for each test session and that it was important that they attempt to answer all questions. The assessment administrator was to create a nonthreatening test environment and to be encouraging and positive without actually reading test material or providing help in selecting an answer.

During the Reading Literacy Test, the assessment administrator was permitted to show students how to record answers correctly. Providing either specific instructions; information, or answers about individual questions; or assistance in reading or spelling was not permitted. During administration of the Student Questionnaire, the assessment administrator could answer questions about any items on the questionnaire.

Exhibit 4-8. Coding guidelines

CODING GUIDELINES

Please use the following sets of codes when providing student race/ethnicity and reading literacy proficiency levels. Write the appropriate numeric code on the class list from the sampled class. Write the codes to the left of the students' names: first the race/ethnicity, then the reading literacy level. Please note that if a student has been officially excluded from participating in the assessment, a reading literacy level is unnecessary.

RACE/ETHNICITY CODES

W = White (not of Hispanic heritage)
B = Black (not of Hispanic heritage)
H = Hispanic (regardless of race)
A = Asian or Pacific Islander
I = American Indian or Alaskan Native
O = Other (race/ethnicity unknown)

READING LITERACY PROFICIENCY LEVELS

The description of reading literacy levels described below are consistent with those being used by all participating nations.

1 = Very poor reader	Consistently demonstrates little understanding of what has been read. Interpretation is very literal. Often cites wrong information in response to a specific question.
2 = Poor reader	Generalizes based on only one dimension. Often overlooks relevant information that may be in surrounding text.
3 = Average reader	Tends to take a number of dimensions into consideration. Can develop some generalizations based on combining information from source materials but often does not account for all inconsistencies or alternative interpretations.
4 = Good reader	Uses all relevant information from texts discriminating between relevant and irrelevant information. Forms generalizations which account for a variety of possibilities. Draws from personal experiences to elaborate conclusions.
5 = Very good reader	Forms generalizations based on information from the text and his experience, accounting for alternative interpretations. Tests his generalizations in new situations and applies his knowledge in new contexts.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

It was important that as many students in the sample as possible were given the opportunity to participate in the session. However, once the session began, no student was admitted to the room. Any student not present (including latecomers) was regarded as absent from the session. Students who arrived after the start of the session were instructed to report to the school office (or follow whatever procedure was devised by the assessment administrator and the coordinator).

To begin the sessions, the assessment administrator went through the same routine for both fourth and ninth graders. The steps were as follows.

Session I:

1. Introduce him/herself.
2. Distribute tests and envelopes.
3. Read directions aloud.
4. Time the test in a nonobtrusive way.
5. Request that each student take a seal from the red strip stapled in the booklets and place it across the edges of finished pages. This prevented students from working on parts of the test that have already been concluded. (Steps 4 and 5 were repeated for part II of the Reading Literacy Test for fourth graders only.)
6. Remind students of next session, conclude this session, and collect materials.
7. Provide a break according to the procedures specified by the school coordinator.

Session II:

1. Distribute the now-labeled Reading Literacy Tests and Student Questionnaires to the proper students (if more than two students who were present at the first session were absent, the school coordinator was consulted about increasing attendance).
2. Review general session directions for the Reading Literacy Test.
3. Read the practice directions for the next part (III for fourth graders and II for ninth graders) and allot time for questions and answers.
4. Time the test in a nonobtrusive way.
5. Instruct students to place the test booklets inside the labeled envelopes when they have completed the final part of the Reading Literacy Test. If the Student Questionnaire was to be completed at this time, the students then removed this document from the envelope and continued in accordance with the procedures discussed in 6 through 9 below.
6. Read the directions in the front of the Student Questionnaire aloud.

7. Monitor the completion of the questionnaire.

8. Conclude the session and collect materials.

9. Ask students to return to their normal routine according to procedures specified by the school coordinator.

4.4.3.5. Field Coding

Student Check-In

As close as possible to the start time for the first session, the assessment administrator counted the number of students in the room and compared it to the number that was calculated from the School Field Log. If only a few students were missing or the count included a few more than expected, roll was called and attendance was recorded. The Administration Schedule was used as a roster and filled in with a "P" for present under "Session Status." At this time an "A" for absent was not filled in for those students not yet in attendance. For those students already identified as excluded or nonparticipating, the column was already filled in.

If many of the nonexcluded students on the Administration Schedule were not present, the assessment administrator waited a few minutes, bearing in mind that school officials would expect the sessions to end on schedule. If more than 20 percent of the students were absent, a makeup session was scheduled. Once the session began, an "A" was marked in the "Session Status" column for those students who were absent from the assessment and not already identified as excluded. After the session, the number of students identified as absent was written on the School Field Log.

Upon completion of the session, the assessment administrator had to account for all the forms used during the session and make sure that the Administration Schedule was properly filled out.

Administrators were responsible for coding administration information on the covers of each Reading Literacy Test and Student Questionnaire used. A one-digit code, which described the outcome of each session for the student who used the booklet/form, was placed in the top right hand-corner of the test cover and the bottom right corner of the questionnaire cover. The administration code was entered in the numbered box corresponding to the session when the student left the testing site. Sufficient time was scheduled between sessions for the administrator to complete the coding during the interim period. The following codes were used:

1 In session full time: Completed entire form or section, or completed, or tried to complete, part of the section or form.

2 In session full time: No response to entire form or section.

3 In session part time: A student leaves the session, regardless of whether or not he/she returns.

4 Session incomplete: The session was interrupted and no student was able to complete the exercise booklet (e.g., fire drill).

5 Absent: Student has been identified as absent on the Administration Schedule or student is absent in a subsequent session after being present in the first session.

6 Long-term absent: As identified by the teacher, this is a student who has been absent from school for more than 2 weeks or who is identified as chronically truant (thus, rarely present).

7 Parental refusal: Student has been identified as absent on the Administration Schedule because his/her parents have officially notified the school that they refuse to allow the student to participate in the assessment.

8 Excluded: Student is found to be eligible for exclusion subsequent to beginning the assessment, because he/she is enrolled in a special education program and has an IEP that prohibits pencil-and-paper assessment.

9 Excused: Student is found to be eligible for exclusion subsequent to beginning the assessment because he/she is identified as a non-English speaking student who has been enrolled in a mainstream English class for less than 1 year.

When leaving the school, the assessment administrator took all forms, used as well as unused. The first copy of the Administration Schedule and School Field Log was left with the school coordinator, as were any other forms and any envelopes marked with student names and ID labels.

4.4.3.6. Field Reporting

Materials were shipped to Westat via standard UPS as soon as an assignment was completed, preferably shipment on the same day as the last session at a school. Under no circumstances was the shipment to be delayed beyond the third day following completion of an assignment. Included in the shipment were the following forms:

- Teacher Questionnaire;
- School Field Log;
- School Contact Form;
- Administration Schedule;
- Student Questionnaire; and
- Shipping Transmittal Form.

5. RECEIPT CONTROL, RESPONSE RATES, AND PROCESSING OF RAW DATA

5.1. Receipt Control

"Receipt Control" is the term given to the procedures and programs for tracking the flow of materials to and from the survey sites. Those materials that were returned directly to Westat were the School Questionnaire, the Teacher Questionnaire, the Student Questionnaires, and one copy of the Administrative Schedule, a five-copy noncarbon form specific to each tested class. In addition, the School Field Log, a computer-generated, school-specific form used to record the final status of the School and Teacher Questionnaires and to indicate the number of students for each tested class, was also sent to Westat. The total number of students on the Administrative Schedule and the School Field Log had to agree with the number of Student Questionnaires received at Westat and later with the Reading Literacy Tests received at Westat from DRC.

The assessment administrators sent the Reading Literacy Tests to Data Recognition Corporation (DRC) for coding, keying, verifying, and basic editing. One copy of the Administration Schedule was sent to DRC along with the Reading Literacy Tests. This form was used to record the date and location of each of the three test sessions, one for the Student Questionnaire and two for the Reading Literacy Test. Further, it associated the name of each student in the class with a preprinted Westat ID, as well as ethnicity, literacy level, statuses for each of the sessions, and a coded "reason for exclusion," if appropriate, and also indicated the number of students at each session. For confidentiality reasons, student names were printed only on the first copy of the Administrative Schedule, which was retained at the school. The other four copies of the Administrative Schedule, which were sent to Westat and DRC, did not contain student names. When DRC personnel finished their tasks, the data were sent to Westat on tape, along with the tests themselves, which were returned to Westat for storage.

A copy of the Administrative Schedule and the School Field Log are included as Exhibits 4-4 and 4-5, respectively.

5.1.1. Receipt of the Tests

When the Reading Literacy Tests were received at DRC, each test booklet was "scan-edited," a process by which a staff member looked through the booklet to ensure that everything was in order (e.g., that there were responses to the questions if the student was marked present at the session). The Administration Schedule, which had been sent under separate cover, was then matched with the reading test from each student, to see if the statuses marked were consistent. All problems identified by this process were referred to the field supervisor at Westat for resolution. In some cases, a check of the School Field Log received at Westat might clarify a count inconsistency. Typically, the assessment administrator would be asked to resolve the discrepancies by referring to his/her own materials or making further contact with the teacher or school.

5.1.2. Receipt of the Questionnaires

The School Questionnaire, Teacher Questionnaire(s), Student Questionnaires, a copy of the Administration Schedule, and the School Field Log were boxed by the assessment administrator for each school and sent to Westat. When a shipment was received at Westat, the contents of each box were checked to ensure that the numbers and statuses recorded on both forms corresponded to the number of questionnaires received and the statuses indicated on them. The questionnaires were then scan-edited to screen for those that were obviously damaged, improperly filled out, or completed in some fashion that was contrary to the status indicated on the forms. These problem cases were set aside.

The number and status information, and notes about problem or missing cases, were entered into a daily manual log. The cases with missing pieces were referred to the field supervisor a week later if the missing pieces had not materialized; other problem cases were referred to the field supervisor immediately. Short of actually coding the responses, great effort was made at this early stage to ensure that the data were as consistent, accurate, and complete as possible. The staff member who was responsible for receipt control then entered the student counts and the questionnaire statuses from the School Field Log, and race/ethnicity, literacy level, the three statuses, and the administrator ID for each student from the Administration Schedule into the computerized receipt control system.

5.1.3. Resolution of Discrepancies

Both at Westat and at DRC, every effort was made to resolve discrepancies between the Administration Schedule and the tests or questionnaires when they were received. Even so, a final set of checks were made when all the materials were brought together at Westat.

By far, the majority of the problems had to do with status discrepancies. In a number of cases, students who were to have been excluded for administrative or other reasons did take the test and were marked as completes. DRC keyed the data in cases where there was doubt about the status of the student, and Westat revised the status and reset the variables to appropriate missing values while preparing the files. Of the approximately 20 cases with student ID problems, those with missing or extra digits turned out to match neatly to students with missing tests. The two pairs of duplicate IDs required handwriting comparison between the Reading Literacy Tests and the questionnaires. The most troublesome class of problems were missing tests. In the end there were none, but achieving that conclusion required a great deal of communication among the field personnel, school personnel, Federal Express, UPS, DRC, and Westat.

5.2. Response Rates

5.2.1. Overall Response Rates

Unweighted and weighted response rate data are summarized in Table 5-1. This table shows that despite the substantial response burden associated with the study, the response rates were high for all respondent categories for both study populations. The unweighted response rates were consistently similar or identical to the weighted rates. Using the weighted figures, the following results were noted:

- Ninety percent of the sampled grade 4 schools participated in the study, as did 88 percent of the sampled grade 9 schools. School Questionnaires were received from all participating schools. The response rate reflects the impact of district nonresponse as well as individual school nonresponse.

- Teacher Questionnaires were received from 100 percent of the designated teachers within participating schools, for both student populations.

Table 5-1. Unweighted and weighted response rates, by grade and type of respondent

Grade and type of respondent	Index					
	Unweighted			Weighted		
	Sample	Response	Rate	Grade	Response	Rate
Schools:[1]						
Grade 4	192	167	87%	61,225	55,104	90%
Grade 9	192	165	86	21,818	19,108	88
Teachers in participating schools:						
Grade 4	304	304	100	-	-	100
Grade 9	165	165	100	-	-	100
Students in participating schools:[2]						
Grade 4	7,041	6,544	93	3,223,966	2,992,863	93
Grade 9	3,738	3,223	86	3,075,604	2,667,004	87
Students, total:[3]						
Grade 4	-	-	-	-	-	84
Grade 9	-	-	-	-	-	76

- Not applicable.

[1] Responding schools are ones that agreed to participate in the study. Completed School Questionnaires were obtained from all participating schools.

[2] Responding students are ones who attended both administration sessions and from whom a usable Student Questionnaire and cognitive Reading Literacy Test were obtained. Excluded students are not included in response rate calculation.

[3] Total student response rate is the product of the weighted school response rate and the weighted response rate of sampled students within participating schools.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

- Within participating schools, the student response rate was 93 percent for grade 4 and 87 percent for grade 9. These figures reflect a conservative counting rule, whereby sampled students were classified as respondents only if they attended both administration sessions and provided usable data for both the Student Questionnaire and the cognitive Reading Literacy Test.

- The overall student response rates, obtained by multiplying the school participation rate and the student response rate within participating schools, were 84 percent for grade 4 and 76 percent for grade 9.

The levels of school response achieved met the ICC requirements for school response and also are in accord with NCES standards for school and student response. That is, the weighted school response rates at both grades exceeded 85 percent, and the weighted student response rate at both grades exceeded 85 percent.

5.2.2. Response Rates by Stratification Variables

As noted above, participating schools provided 100 percent of the School and Teacher Questionnaires required in the study design. There was no variation among strata in these respects. School and student response rates did vary somewhat from stratum to stratum, however. Tables 5-2 and 5-3 present (unweighted) school and student response rate data, by sampling stratum, for grades 4 and 9, respectively.

The data in these tables are summarized in Tables 5-4 and 5-5, which present school and student response rate data for grades 4 and 9, respectively, by four main school stratification variables: region, urbanicity, minority level, and type of control (public/private).

As shown, the Northeast region had the lowest school response rate for both grade 4 (70 percent) and grade 9 (68 percent), while the Southeast region had the highest school response rate (96 percent for both grades. The other two regions had intermediate school response rates for both grades. School response rates did not vary substantially, or consistently, across grades by school urbanicity, minority level, or ownership. For both grades, student response rates within participating schools varied only slightly by stratification variable category. Thus, in both grades, student response rates were 85 percent or above in all categories of all stratification variables.

5.3. Data Processing

5.3.1. Editing and Coding

Questionnaires without problems were batched after arrival and scan-editing. That is, they were physically grouped in containers holding 25 questionnaires of the same type, and the batch number of the container was associated with the 25 case IDs in the automated receipt control system and the manual log. The batch number then became the identifier for that case in all its subsequent physical movement through coding, key entry, editing, and archiving.

Batch control was maintained in both computerized and manual form. Both systems recorded the dates the batch was sent to coding, sent to keypunch, and returned from keypunch. In addition, in the manual system the number of machine-edit cycles was kept. If the machine-edit program discovered logic

Table 5-2. School and student response rates, by stratum: Grade 4, unweighted

Stratum							School			Student		
Number	Region	Urbanicity	Minority level	Certainty status	Substratum							
					Ownership	School size	Sample	Response	Rate	Sample	Response	Rate
Total							192	167	87%	7,041	6,544	93%
1	Northeast	SMA	NA	Certainty	Public	15-49	2	2	100	101	97	96
2	Northeast	SMA	NA	Certainty	Public	50+	12	6	50	288	262	91
3	Northeast	SMA	NA	Certainty	Private	15+	4	3	75	108	101	94
4	Northeast	All	NA	Noncertainty	Public	15-49	5	4	80	136	129	95
5	Northeast	All	NA	Noncertainty	Public	50+	12	10	83	465	428	92
6	Northeast	All	NA	Noncertainty	Private	15+	2	1	50	47	34	72
7	Southeast	SMA	High	Certainty	All	15+	4	4	100	206	196	95
8	Southeast	SMA	High	Noncertainty	Public	15+	9	8	89	389	352	90
9	Southeast	SMA	Low	Noncertainty	Public	15+	11	11	100	508	482	95
10	Southeast	All	All	Noncertainty	Private	15+	4	4	100	119	107	90
11	Southeast	non-SMA	High	Noncertainty	Public	15+	8	7	88	310	293	95
12	Southeast	non-SMA	Low	Noncertainty	Public	15+	9	9	100	307	274	89
13	Central	SMA	NA	Certainty	Public	15+	7	7	100	399	378	95
14	Central	SMA	NA	Certainty	Private	15+	2	2	100	99	97	98
15	Central	All	NA	Noncertainty	Private	15+	5	5	100	150	136	91
16	Central	SMA	NA	Noncertainty	Public	15-49	4	4	100	166	155	93
17	Central	SMA	NA	Noncertainty	Public	50+	13	12	92	564	528	94
18	Central	non-SMA	NA	Noncertainty	Public	15-49	8	6	75	151	135	89
19	Central	non-SMA	NA	Noncertainty	Public	50+	6	4	67	180	170	94
20	West	SMA	High	Certainty	All	15+	9	8	89	246	226	92
21	West	All	All	Noncertainty	Private	15+	4	4	100	96	93	97
22	West	SMA	High	Noncertainty	Public	15+	17	15	88	663	620	94
23	West	SMA	Low	Noncertainty	Public	15+	19	17	89	895	837	94
24	West	non-SMA	All	Noncertainty	Public	15-49	5	5	100	188	180	96
25	West	non-SMA	All	Noncertainty	Public	50+	7	5	71	218	197	90
26	All	All	All	All	All	<15	4	4	100	42	37	88

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-3. School and student response rates, by stratum: Grade 9, unweighted

Stratum					Substratum		School			Student		
Number	Region	Urbanicity	Minority level	Certainty status	Ownership	School size	Sample	Response	Rate	Sample	Response	Rate
Total							192	165	86%	3,738	3,223	86%
1	Northeast	SMA	NA	Certainty	Public	15+	17	9	53	202	170	84
2	Northeast	SMA	NA	Certainty	Private	15+	3	3	100	87	75	86
3	Northeast	SMA	NA	Noncertainty	Public	15+	13	10	77	197	175	89
4	Northeast	SMA	NA	Noncertainty	Private	15+	3	2	67	29	25	86
5	Northeast	non-SMA	NA	Noncertainty	All	15+	5	4	80	89	82	92
6	Southeast	SMA	High	Certainty	All	15+	3	3	100	81	74	91
7	Southeast	All	All	Noncertainty	Private	15+	3	3	100	62	60	97
8	Southeast	SMA	High	Noncertainty	Public	15+	9	9	100	193	158	82
9	Southeast	SMA	Low	Noncertainty	Public	15+	12	11	92	292	254	87
10	Southeast	non-SMA	High	Noncertainty	Public	15+	9	8	89	188	168	89
11	Southeast	non-SMA	Low	Noncertainty	Public	15+	9	9	100	199	167	84
12	Central	SMA	NA	Certainty	All	15+	10	10	100	222	187	84
13	Central	All	NA	Noncertainty	Private	15+	3	2	67	46	44	96
14	Central	SMA	NA	Noncertainty	Public	15+	21	19	90	391	338	86
15	Central	non-SMA	NA	Noncertainty	Public	15+	14	13	93	287	253	88
16	West	SMA	High	Certainty	All	15+	8	8	100	196	165	84
17	West	All	All	Noncertainty	Private	15+	2	1	50	17	17	100
18	West	SMA	High	Noncertainty	Public	15+	18	16	89	349	288	83
19	West	SMA	Low	Noncertainty	Public	15+	18	14	78	376	315	84
20	West	non-SMA	All	Noncertainty	Public	15+	11	10	91	222	196	88
21	All	All	All	All	All	<15	1	1	100	13	12	92

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-4. School and student response rates, by school stratification variables: Grade 4, unweighted

Stratification variable	School			Student		
	Sample	Response	Rate	Sample	Response	Rate
Total	192	167	87%	7,041	6,544	93%
Region:						
Northeast	37	26	70	1,145	1,051	92
Southeast	45	43	96	1,839	1,704	93
Central	45	40	89	1,709	1,599	94
West	61	54	89	2,306	2,153	93
All	4	4	100	42	37	88
Urbanicity:						
SMA	113	99	88	4,632	4,331	94
Non-SMA	43	36	84	1,354	1,249	92
All	36	32	89	1,055	964	91
Minority levels:						
High	47	42	89	1,814	1,687	93
Low	28	26	93	1,202	1,111	92
All	117	99	85	4,025	3,746	93
Ownership:						
Public	154	132	86	5,928	5,517	93
Private	21	19	90	619	568	86
All	17	16	94	494	459	93

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-5. School and student response rates, by school stratification variables: Grade 9, unweighted

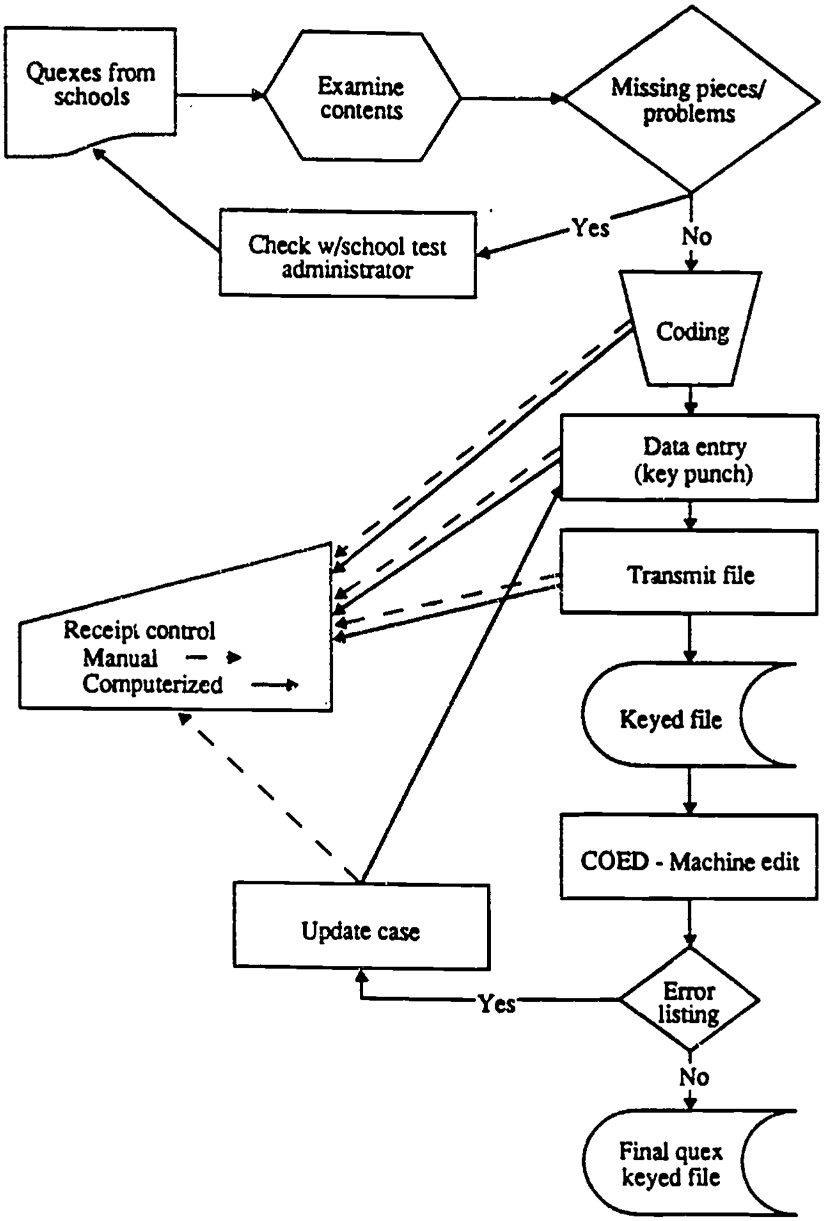
Stratification variable	School			Student		
	Sample	Response	Rate	Sample	Response	Rate
Total	192	165	86%	3,738	3,223	86%
Region:						
Northeast	41	28	68	604	527	87
Southeast	45	43	96	1,015	881	87
Central	48	44	92	946	822	87
West	57	49	86	1,160	981	85
All	1	1	100	13	12	92
Urbanicity:						
SMA	135	114	84	2,615	2,224	85
Non-SMA	48	44	92	985	866	88
All	9	7	78	138	133	96
Minority levels:						
High	47	44	94	1,007	853	85
Low	39	34	87	867	736	85
All	106	87	82	1,864	1,634	88
Ownership:						
Public	151	128	85	2,896	2,482	86
Private	14	11	79	241	221	92
All	27	26	96	601	520	87

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

or range errors, an error listing was produced. The error then had to be resolved, and the case was updated and the edit program rerun.

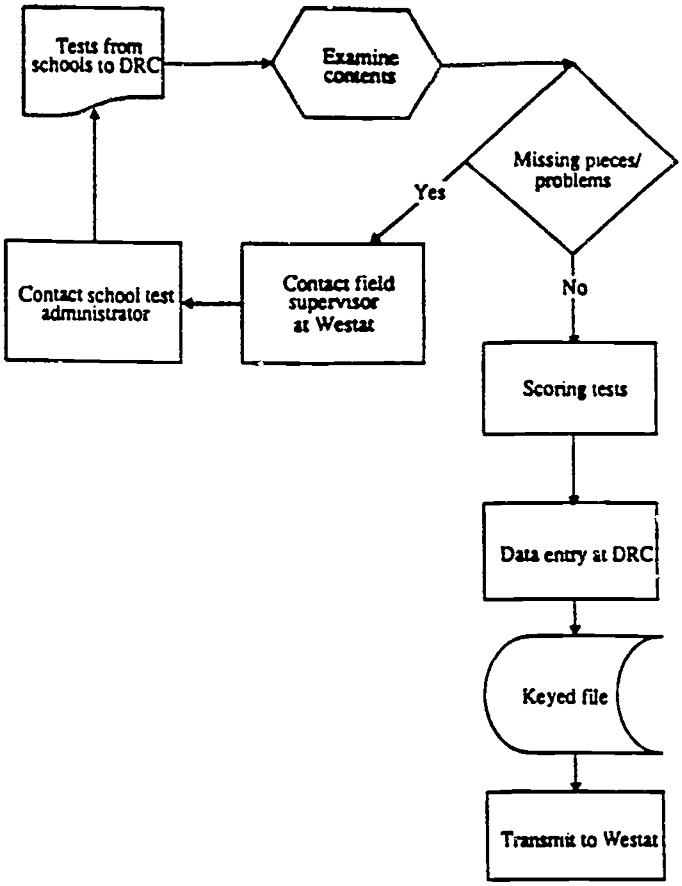
The manual and the automated receipt control systems complemented each other. The manual system had flexibility to track interim statuses of problem cases; the automated system provided for quick tabulations and aggregate summaries. Flowcharts illustrating the progress of questionnaires and reading tests through the receipt control and data processing systems are included as Exhibits 5-1 and 5-2.

Exhibit 5-1. Processing of questionnaires



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Exhibit 5-2. Processing of reading tests



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

5.3.1.1. Data Dictionairies

The coding, keying, and editing operations depend upon the data dictionairies that are produced by Westat's standard Codebook Editor, or COED, software system. The COED software system is written in COBOL and PL/1 and supports its own language for description of survey questionnaires. The input to the system, the COED source file, contains information on all data items for a particular questionnaire, including the field names, the questions, column numbers for the file to be produced, the data type, coding schemes or ranges of values, logic checks, and variable and value labels. This file is created as the questionnaire is developed and is updated throughout the survey as necessary. As the coding of the open-ended questions proceeds, for instance, further values with their labels are incorporated.

Using this file, COED's menu-driven system can produce a COBOL program to machine edit and update the data file for the questionnaire. The program runs in a batch environment, supports a hierarchical file structure, produces printed reports, and supplies a paper trail of the transactions applied to the file. Included in its output is the codebook documentation, an invaluable working document for the project and keying staff. The key entry program is derived from this source file (see Exhibit 5-3 for a sample page of a codebook). It also becomes a standard part of project documentation. Furthermore, COED is used to generate the program to produce a SAS or SPSS file of the data with formatted and labeled variable values and to run the program if desired.

Data processing staff produced 12 different codebooks: a version for each population of the School, Teacher, and Student Questionnaires -- six documents in all -- and a version for each Reading Literacy Test session, which included the variant fourth grade versions -- the additional six documents.

5.3.1.2. Coding the Questionnaires

Coding is the detailed review of the questionnaire by personnel trained to discover problems it contains. The term specifically refers to the coding into meaningful categories of responses to questions where the set of response alternatives is not specified in the questionnaire. For example, each student was asked to name the book he or she had most recently read. Each different book named was associated with a number, the number became a part of the data for that student, and a tabulation of favored books could be made. In fact, the coding of such open-ended questions (see Section 5.3.2) was delayed until after the other data were keyed and edited.

What was coded at this point was whether a missing answer was a "Not Applicable," "Not Ascertained," or "Don't Know." In addition, the coders checked that all of the skip patterns were properly followed. Answered questions that should have been skipped were set to "Not Applicable," and unanswered questions that should not have been skipped were set to "Not Ascertained." If the responses to questions that should have been skipped according to a previous answer were cogent, then the previous answer was changed. The coders were trained to resolve duplicate or confused answers as appropriate. Finally, legibility problems were fixed to the extent possible.

The coders were expected to resolve novel situations in consultation with project staff and to be consistent with each other in their decisions. Later, these same staff members were responsible for resolving problems turned up by the machine-edit program (see Section 5.3.1.3).

5.3.1.3. Key Entry and Machine Editing

At Westat

Coded questionnaires were submitted to Westat's data entry facility in batches. The facility uses a sophisticated disk data entry system that provides online editing and updating as data are keyed, controls data verification, generates production and quality reports for the data entry supervisor, and then transmits data files directly to the project processing system.

Westat's online key entry system includes programmable range checks and is a 100 percent verification system. This means that all data are entered twice by different operators and then compared. Any differences are resolved with adjudication by the supervisor for difficult cases. The coders and their supervisors work closely with the key entry facility to anticipate and forestall these problems. Backup

Exhibit 5-3. Sample page from a COED codebook

```
Q WHAT IS YOUR SOCIAL SECURITY NUMBER?
C 000000001-999999996 = SOCIAL SECURITY NUMBER
C 999999997 = REFUSED
C 999999998 = DON'T KNOW
C 999999999 = NOT ASCERTAINED
V Q2A      02 N 01 025 - 026
Q WHAT IS YOUR BIRTHDATE? --- MONTH
C 01-12 = MONTH
C 97 = REFUSED
C 98 = DON'T KNOW
C 99 = NOT ASCERTAINED [NO ENTRY]
V Q2B      02 N 01 027 - 028
Q DAY
C 01-31 = DAY
C 97 = REFUSED
C 98 = DON'T KNOW
C 99 = NOT ASCERTAINED [NO ENTRY]
V Q2C      02 N 01 029 - 030
Q YEAR
C 20-75 = YEAR
C 97 = REFUSED
C 98 = DON'T KNOW
C 99 = NOT ASCERTAINED [NO ENTRY]
V Q3       01 N 01 031
Q WHAT IS YOUR SEX?
C 1 = MALE
C 2 = FEMALE
C 7 = REFUSED
C 8 = DON'T KNOW
C 9 = NOT ASCERTAINED
V Q4       01 N 01 032
Q WHAT IS YOUR MARITAL STATUS?
C 1 = NEVER MARRIED
C 2 = MARRIED
C 3 = SEPARATED
C 4 = DIVORCED OR WIDOWED
C 7 = REFUSED
C 8 = DON'T KNOW
C 9 = NOT ASCERTAINED
V Q5A      01 N 01 033
Q DO YOU HAVE ANY DEPENDENT CHILDREN?
C 1 = YES
C * 2 = NO
C * 7 = REFUSED
C * 8 = DON'T KNOW
C * 9 = NOT ASCERTAINED
S * SKIP Q5B (CODE AS INAPPLICABLE)
V Q5B      02 N 01 034 - 035
Q IF YES, ENTER HOW MANY
C ++ = INAPPLICABLE
C 01-15 = CHILDREN
C 97 = REFUSED
C 98 = DON'T KNOW
C 99 = NOT ASCERTAINED
V Q6       01 N 01 036
Q WHAT IS YOUR RACE/ETHNICITY?
C 0 = OTHER
C 1 = AMERICAN INDIAN
C 2 = ALASKAN NATIVE
C 3 = BLACK (NOT HISPANIC)
```

SOURCE: Codebook Editor, Westat, Inc.

procedures and personnel are in place to ensure that schedules accepted by the data entry manager can be maintained.

Once a batch of questionnaires is transmitted from the keying system to the project data processing accounts, the machine-edit program is run. The purpose of machine editing is to detect and resolve as many errors as possible prior to delivering the data for more complex interfile edits and statistical data quality analyses. The errors that can be detected by machine editing are of two general types:

- Range errors, in which response values fall outside a predetermined acceptable range; and
- Logic errors, in which there is some inconsistency between response values. These include improperly followed skip patterns, identified data inconsistencies among two or more variables, and addition checks where values of a group of variables are to sum to a known value.

In range checking it is useful to distinguish between soft ranges, outside of which data values require verification but may be legal, and hard ranges, outside of which data values are surely in error. Similarly, in logic checking it is important to distinguish between improbable and impossible inconsistencies between data items.

The general machine-edit update cycle consists of the following steps:

- Execute the edit program on the file;
- Resolve errors and discrepancies;
- Perform updates; and
- Repeat the cycle until the survey data are clean.

The machine-edit program can be simple and contain only the ranges and logic checks, which include the skip patterns, built into the COED codebook, or it can be further programmed to check across files to perform more complex edits. The program produces a listing of cases with problems. Those cases are pulled and the error is resolved by the same person who coded the case. The majority of the cases are overrides (i.e., the data are out of range or inconsistent, but do match the response in the questionnaire). In some situations, the editors are authorized to make changes to the data in order to achieve consistency; in others, only supervisors are so authorized. All changes are written into the questionnaire in a color that indicates the editing phase, and the rationale is noted in the margins. The editors then write up the changes to be made on an update sheet, which is sent to data entry and keyed. Then the update records are transmitted to the project data processing account, the records are updated with the COED system, and the machine-edit program is run again.

Normally one or two such cycles is sufficient to produce a cleaned file. In the Reading Literacy Study, however, the average was three to five, because coding the open-ended questions was left until this stage.

At DRC

In addition to key entry of all student responses to the Reading Literacy Test items, scoring the open-ended writing responses (included in the Reading Literacy Tests) was the major task for DRC. This was a sophisticated process in which each essay was read by two readers independently and scored. If the scores differed, a third resolving reading was done by a task leader. Scoring was monitored closely. Daily reports were produced for each reader indicating the number of papers read, the percentage of exact, adjacent, and nonadjacent agreement with the other readers of the same papers, the tendency of the disagreement, and the score point distribution. The areas of scrutiny were inconsistency, or drift from an established standard. Throughout the project, readers scored sample papers at rangefinding meetings in order to validate and recalibrate the criteria. Retraining was ongoing to secure continued familiarity with and adherence to the scoring criteria and to prevent roomwide drift as the project progressed. Legibility issues were addressed implicitly in the open-ended question scoring process.

The scorers of the open-ended items were experienced in scoring similar questions for other large-scale assessments. Those scorers were generally recruited high school teachers who were provided training for scoring to open-ended questions for this study.

To reduce key-entry errors, DRC also used a 100 percent verification system (i.e., each test booklet was keyed independently by two operators and entries were compared before merging them to the data file), and, in addition, they used COED system codebooks written by Westat for the reading tests. DRC incorporated the range checks in their data entry program.

5.3.2. Coding Open-Ended Responses

Open-ended and other-specify questions form a class of responses that require special coding because the responses are prose and conform to no preselected response options to the questions. In an open-ended question, no response alternatives are specified, and the respondent is asked to enter a response, such as the name of the last book he or she has read. In an other-specify question, some alternatives are precoded on the question, with the expectation that these do not meaningfully represent the likely responses, and a space is provided for the respondent to write in an appropriate response.

In the IEA Reading Literacy Study, all open-ended and other-specify questions were coded, except for Student Questionnaire question 55 -- last book read. In this case, as the alternatives mounted toward 1,000, some responses were left uncoded, and the other-specify variables for this question, WASBOOKW, were all coded the same. A sample page from a codebook of the representation of a completely coded open-ended question is included (Exhibit 5-4).

In order to make the coding consistent, one coder coded all the responses for a particular question. These coded values were sent to the key entry staff, who keyed them as update records and transmitted them to the project account for updating. In other words, the process was an extension of the edit-update process.

The open-ended and other-specify questions were as follows:

- Grade 4 Student Questionnaire - Q16,Q36,Q53,Q55,Q57,Q59,Q61,Q69
- Grade 4 School Questionnaire - Q23,Q25

Exhibit 5-4. Sample codebook page for an open-ended question

Q 23

01 COMPLETE IN-SERVICE PROGRAM SPONSORED BY THE READING RESOURCE TEACHER (IN-SCHOOL PROGRAM)
02 ASSIGNED READING WITH REQUIRED SYNOPSIS AND STUDENT EVALUATION
03 IN-SERVICE PROGRAM AT DISTRICT/STATEWIDE/PROFESSIONAL LEVEL (OUTSIDE PROGRAM)
04 COURSE DEVELOP REMEDIAL READING; SKILL IMPROVEMENT; READING ENRICHMENT
05 COMBINATION OF IN-SERVICE PROGRAM AND OUTSIDE (SCHOOL DISTRICT) PROGRAMS
06 IN SPECIAL EDUCATION, REMEDIAL READING
07 SELF-ESTEEM AND SKILL DEVELOPMENT TECHNIQUES
08 COMBINATION OF ADVANCED PLACEMENT PROGRAMS AND REMEDIAL PROGRAMS
09 PROGRAM DISCONTINUED DUE TO LACK OF FINANCES; CUTBACKS OF FEDERALLY FUNDED PROGRAMS
10 TRAINING PROGRAMS FOR TEACHERS ON SKILL ENHANCEMENT

Q25

01 TRANSPORT PROBLEMS HINDER AFTER SCHOOL PROGRAMS
02 INSTRUCTIONAL EXCELLENCE
03 RECOGNIZE NEED FOR READING PROGRAMS
04 NOT A PERCEIVABLE NEED
05 FINANCIAL (BUDGET) CUTS
06 LACK OF MATERIALS
07 ABSENTEEISM
08 NEED FOR SUPPORT FROM HIERARCHY
09 CLASS OVERLOADS

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

- Grade 4 Teacher Questionnaire - Q3,Q13,Q39
- Grade 9 Student Questionnaire - Q22,Q46,Q55,Q61
- Grade 9 School Questionnaire - Q23,Q25
- Grade 9 Teacher Questionnaire - Q3,Q13,Q27

5.4. Creating the Files

5.4.1. The U.S. Files

The study produced eight U.S. files in all. For each population, reading test data from two testing sessions were combined. The file for grade 4 further combined the data for both the standard version of the reading tests and the variant version given to about 5 percent of the students. In addition, a file was created for each population for the Student, Teacher, and School Questionnaires.

After the key-entry and machine-edit process, these files were cleaned according to the range and consistency checks in the codebook. Review, analysis, and further editing of the files were then undertaken by the project director himself. These checks were less about formal consistency than about material consistency. The effort was, on the one hand, to search for improbability or nonsense, trying to be sure that aggregate and individual results made sense as a meaningful whole and that responses did not turn on ambiguity, misunderstanding, or poor response alternatives. There was further editing of cases, with many individual questionnaires reviewed. On the other hand, it was an effort to understand the results, to discover relationships among the question responses, and to begin the analysis that is ongoing.

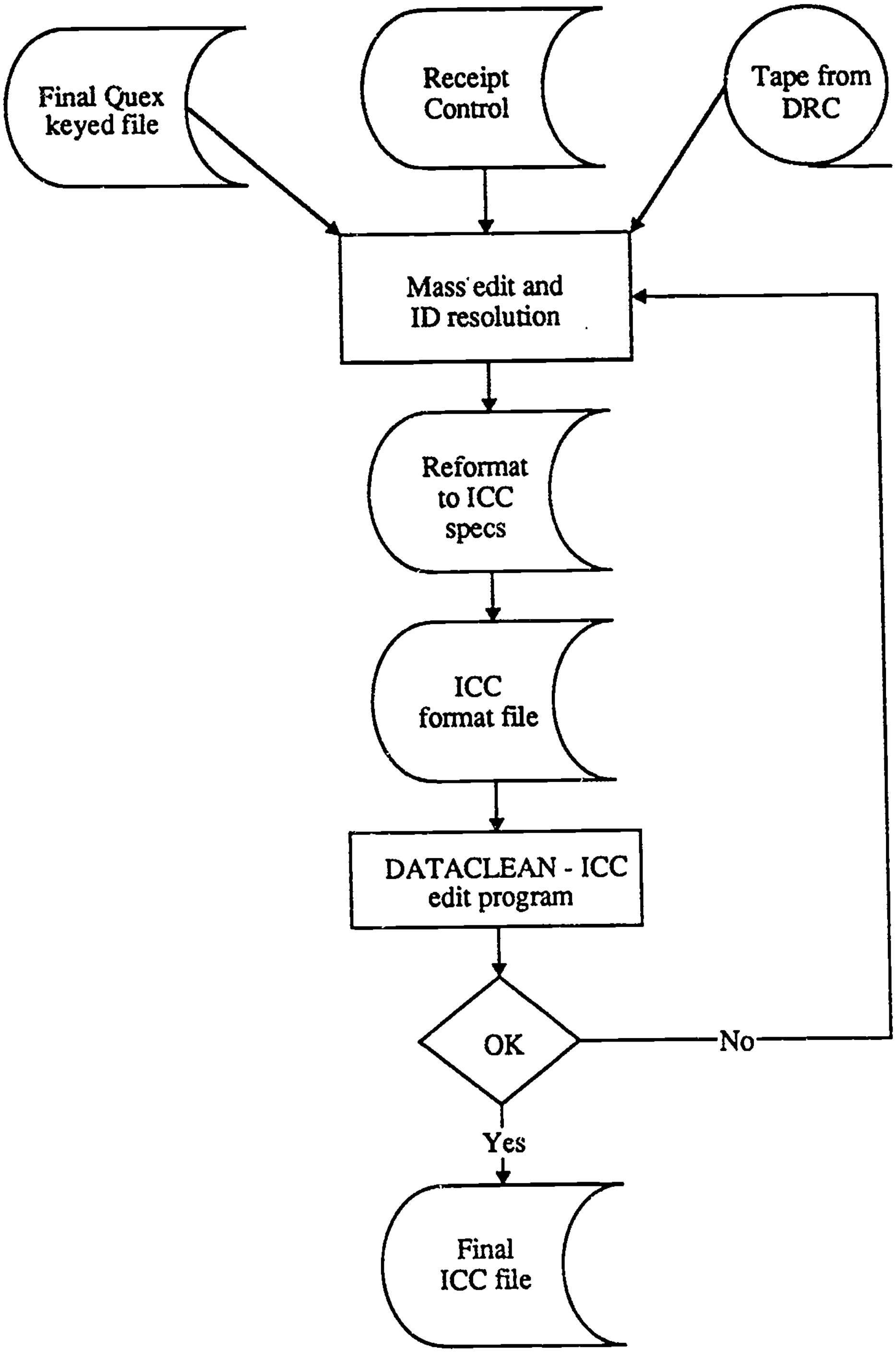
5.4.2. The International Files

The eight U.S. files were combined and reformatted in accordance with the specifications provided by ICC to produce six ICC international format files. The U.S. Teacher and School Questionnaire files were mapped onto ICC versions; the U.S. Student Questionnaire and Reading Literacy Test files were mapped onto a single ICC student file for each population (see Exhibit 5-5). While only a few of the questions in the U.S. questionnaires were asked with the same wording and response alternatives as their analogues in the ICC version, the data, nonetheless, were to go to the ICC in the format of its questionnaires.

The ICC supported its questionnaires with software for data entry, record editing, range checks, ID checks across files, and logic and consistency checks, including skip patterns and intra- and interfile checks. These checks were a subset of those included in Westat's codebooks. When the data were converted to ICC format and these checking programs were run, almost all of the errors occurred in cases where a prescribed range was violated by a legitimate, if unusual, value, or a consistency check was violated by a combination of such values. Essentially the data did not require further editing in order to conform to ICC standards.

The substantive tasks were twofold: first, to determine how to map the U.S. questions to the ICC questions, and second, how to map the responses to the U.S. questions to those of the ICC questions. The relationship between the pairs, or among the sets of questions and their responses, varied from the

Exhibit ⁵-5. File creation



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

obvious to the subtle and complex. The ICC consistency checks provided an implicit check on the correctness of the mapping.

A basic match had to be maintained between the different sets of not ascertained, not applicable, and other missing values. Similarly, the statuses of present, absent, and excluded for a student session were mapped to corresponding values for student and booklet. U.S. questionnaire skip patterns absent in the ICC format complicated the mapping considerably. Values from the skip determining variable were combined with those of variables within the skipped sequence to determine the appropriate ICC response.

5.5. Weighting

Since the secondary stratification was applied only to the schools in the initial sample of NAEP PSUs, after weighting up the characteristics of the schools in the sampled PSUs by the inverse of the probabilities of selection of those PSUs, sampling error was introduced in the estimates of the substratum totals. Since the time that the design was set, it has been possible to tabulate the entire QED file by the characteristics that define the substrata. This made it possible to adjust the sample weights so that the number of schools in the selected sample would weight up to the number of schools in the QED tape within each substratum -- a straightforward poststratification procedure.

The enrollments in the sampled schools were multiplied by the school weights and compared with estimated enrollments for the fourth and ninth grades produced by the Current Population Survey (CPS). The differences were judged to be large enough that a second adjustment to the sampling weights was made so that the estimated enrollments in the two grades would equal the CPS estimates within each NAEP region.

The two weight adjustments automatically correct for school nonresponse to the survey. In making the first adjustment, the weighted number of sampled schools was adjusted to equal the number of schools listed in the QED file, with no account taken of the number of schools that had closed. This handling of closed schools was considered appropriate since there was no opportunity to include schools newly opened after data collection for the QED file ended.

The student weights within each school reflect both the subsampling of classrooms in the school and the individual student nonresponse within the school. That is, the school weight was multiplied by the number of classrooms in the school and divided by the number of classrooms sampled. This weight was multiplied by the number of students in the selected classrooms and divided by the number of responding students to produce the student weights.

The distribution of weights after adjustment is shown by substrata in Tables 5-6 and 5-7. Note that the range in weights within the substratum is never more than twice the average weight. The last substratum for each class represents the schools with an estimated enrollment of less than 15 students in the class. These two substrata were sampled thinly to conserve costs, so one expects their average weights to be high. However, the weighted sum of students in substratum 26 for the fourth grade is only 3 percent of the total, so that the contribution to the average is small from the large weights there. For the ninth grade, substratum 21 weights up to only about 1 1/2 percent of the total students.

Table 5-6. Weights by substrata: Grade 4

Substratum	Maximum weight	Minimum weight	Average	Sum of weights
1	644.2	596.5	620.4	60,247
2	1031.3	633.0	889.9	231,655
3	998.9	451.9	666.0	64,723
4	944.6	674.3	850.3	124,004
5	907.2	113.4	572.3	245,555
6	494.7	494.7	494.7	16,821
7	877.4	565.3	714.4	141,318
8	1203.4	242.8	573.3	207,809
9	628.4	327.0	461.4	224,020
10	744.1	269.5	519.5	52,061
11	753.8	203.3	427.4	117,354
12	554.7	214.1	333.2	91,740
13	728.7	208.2	415.5	160,344
14	769.1	270.7	436.9	40,215
15	892.5	640.9	707.3	98,517
16	1029.6	489.8	644.7	101,438
17	679.8	177.7	506.1	269,221
18	621.2	512.5	570.3	83,026
19	805.8	555.7	691.1	121,170
20	1035.9	228.9	496.0	142,455
21	569.6	470.8	496.8	46,841
22	1013.3	187.6	514.2	305,071
23	975.3	279.2	458.8	388,262
24	395.7	353.6	369.1	66,485
25	844.7	671.2	744.8	146,182
26	4048.1	2576.0	3100.2	110,397
Total				3,656,929

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-7. Weights by substrata: Grade 9

Substratum	Maximum	Minimum	Average	Sum of weights
1	2,483.9	671.1	1,378.7	213,508
2	2,628.8	1,376.7	1,868.7	134,824
3	2,071.8	1,180.6	1,663.7	289,340
4	1,019.3	329.6	674.4	16,516
5	3,562.4	263.6	1,147.8	63,120
6	1,145.6	376.6	844.8	60,409
7	1,251.0	518.5	834.5	47,657
8	1,723.3	316.2	797.2	117,513
9	1,436.4	671.0	958.6	239,497
10	2,208.1	228.3	1,318.3	219,710
11	978.6	588.9	800.1	133,714
12	1,576.8	534.3	1,064.5	187,838
13	1,592.2	678.0	1,154.4	217,428
14	1,786.1	671.9	983.5	327,937
15	1,445.4	418.6	935.5	229,549
16	1,837.8	256.5	1,219.8	224,310
17	610.7	610.7	610.7	10,383
18	3,038.2	343.4	710.4	185,294
19	2,742.3	326.2	1,375.9	380,447
20	1,563.5	679.5	1,044.0	199,549
21	4,599.9	4,599.9	4,599.9	55,199
Total				3,553,741

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

5.6. Estimation of Averages, Ratios, and Proportions

Averages, ratios, and proportions were all estimated by the same method. The method is described here in terms of the average scaled narrative reading score. The estimated average over all students in the given grade is the weighted sum of the sample scores divided by the weighted sum of all sampled students taking the test. The same rule applies if the sum is taken over any subset of the sample, say, males, or Hispanics, or students who watch television more than 2 hours per day. For example:

AST_WGT_{ijk} = adjusted student weight for the kth student in the jth school in the ith PSU (or variance replicate as defined in the next section)

SS_NAR_{ijk} = narrative reading score of the kth student in the jth school in the ith PSU (or variance replicate)

Then,

$$T(SS\_NAR) = \sum_{S} ASTD\_WGT_{ijk}*SS\_NAR_{ijk} \tag{5.1}$$

is an estimate of the total narrative reading scores for the subset of i, j, and k defined by the subset s. Similarly,

$$T(ASTD\_WGT) = \sum_{S} ASTD\_WGT_{ijk} \tag{5.2}$$

is the corresponding estimate of the number of students. The estimated average score of the students in the subset is the ratio of T(SS_NAR) to T(ASTD_WGT).

The same formulation is appropriate for any ratio of one variable to another if one replaces an estimate of the number of students in the denominator by the weighted sum of the variable serving as the base of the ratio; for example, hours of TV watching per hour spent on homework. The formulation also works for estimating the proportion of students having a given characteristic, such as having a single parent. In that case, the variable in the numerator is one if the student has a single parent and zero otherwise. The weighted sum in the numerator is then the sum of the weights for the students with the given characteristic.

The subset s can also be defined generally to include students with specified characteristics or students enrolled in schools with specified characteristics.

5.7. Estimation of Standard Errors

The sample was designed so that standard errors could be estimated using the "ultimate cluster" method (see Hansen, Hurwitz, and Madow 1953). The ultimate cluster is an aggregation of students that reflects the gains in precision from stratification and the loss in precision from clustering of the students within classrooms or within schools. For students in schools that are not in PSUs that were selected with certainty, the appropriate cluster is the PSU, since the aggregate for the PSU takes into account the stratification and allows for variation between PSUs within strata and between schools within the PSUs. For PSUs selected with certainty, the appropriate cluster is the school. There is no contribution from the variation among PSUs since the PSU was selected with certainty.

Standard errors for the descriptive statistics were computed by the jackknife method (Rust 1985) using Westat's WESVAR software and the ultimate clusters described above. To use this method, the noncertainty PSUs were grouped into pairs within the substrata, and within the certainty PSUs the schools were grouped into pairs. Each member of the pair (i.e., each half-sample) could contain more than one school. A summary of the pairing for both grade 4 and grade 9 is shown in Table 5-8.

Table 5-8. Summary of variance strata for jackknife estimation

Region	Urbanicity	Certainty	Minority	Number of variance strata	
				Grade 4	Grade 9
Northeast	MSA	Certainty	All	5	5
		Noncertainty	All	2	2
	Non-MSA	Noncertainty	All	1	1
Southeast	MSA	Certainty	High	2	1
		Noncertainty	High	1	1
		Noncertainty	Low	1	1
	Non-MSA	Noncertainty	High	1	1
		Noncertainty	Low	1	1
Central	MSA	Certainty	All	4	4
		Noncertainty	All	3	3
	Non-MSA	Noncertainty	All	2	2
West	MSA	Certainty	High	4	5
		Noncertainty	High	2	2
		Noncertainty	Low	2	2
	Non-MSA	Noncertainty	All	2	2
Total				33	33

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

To compute the jackknife estimate, the quantities defined in equations (5.1) and (5.2) in Section 5.6 are computed for the whole file (or for a given subset defined by s), and the ratio or average or proportion is computed from them. Then, one member of the pair in the first variance replicate is selected at random. That member of the pair is given zero weight, and the weight of the other member is doubled. The quantities defined in (5.1) and (5.2) as well as the average or ratio or proportion are recomputed for this set of weights. This constitutes the first replicate estimate. Call it E_1 and denote the overall estimate computed above by E. Repeat this process for each of the variance replicates. Then, the standard error (se(E))of the estimate, E, is found by

$$se(E) = \sqrt{\sum_i (E_i - E)^2} \tag{5.3}$$

where the summation is over all of the variance replicates.

There are 36 noncertainty NAEP PSUs in the 12 original strata. The noncertainty PSUs in variance estimation strata were created by pairing adjacent noncertainty PSUs in the same stratum whenever possible. There were 18 variance estimation strata in the noncertainty PSUs. One of these for grade 4 estimates contained only one PSU and was collapsed with the succeeding PSU in the ordered sequence to form a replicate pair.

The contribution to variance of the certainty strata was estimated by pairing schools in the certainty strata. Variance estimation strata were formed by pooling responding schools in the same region with the same type of control (public or private) and same enrollment class. Pairs (and triplets, where necessary) were formed within these classes. The sample of grade 4 schools in certainty strata consisted of 32 responding schools in 12 PSUs. The grade 9 sample contained 33 responding schools in 13 certainty PSUs. Schools for both grades were grouped into 15 variance strata with two or three schools in each.

In two cases in grade 4 and three cases in grade 9, it was convenient to have three members (PSU. - schools) within one variance stratum. Replacing the total estimate for the variance stratum by two times the estimate from one member (or, alternatively, two times the sum of the estimates from the other two members) biases the estimate of variance upward by a small amount. This was avoided by using Function 3 of the WESWGT proprietary Westat package. When there are three members within the variance stratum, say a, b, and c, one can form three estimates of the stratum from taking 1.5 times a + b, or 1.5 times a + c, or 1.5 times b + c. Function 3 chooses two of these estimates at random and, in effect, forms two replicates representing the i^{th} variance stratum, say i(1) and i(2) where i represents the stratum with three members. The variance is computed as in (1), above, adding over all of the replicates, including the additional ones created by splitting the three members.

The jackknife estimation method serves quite generally for most estimates and the estimate, E, can take many forms. It is known not to be efficient in some circumstances for the estimates of position statistics such as medians or percentiles. However, empirical research (Hansen 1989) suggests that for multistage samples of PSUs, schools, and students, it has sufficient reliability for such position measures for the total student population.

Jackknife estimates of standard errors were computed for the subclasses of family composition, language spoken, ethnicity, father's education, mother's education, gender, whether the student lives with a nuclear family or an extended family, region of residence, and degree of urbanization. Other variables could have been chosen, but these are likely to be used in many analyses of the data. The jackknife estimates are shown in Table 5-9 for grade 4 and in Table 5-10 for grade 9. The estimated standard errors differ somewhat by the reading scale used and by the variable defining the subgroup. There is a general tendency for standard errors to decrease as the sample size, n, increases, but the relationship is not linear. The fact that the sample is clustered (all of the students in a classroom were taken into the sample) causes the variable categories with large numbers of students in them to have relatively larger standard errors than can be accounted for by the sample size. With large subgroups there are, on the average, many students in a classroom that are members of the subgroup and hence the effect of the intraclass correlation (see Hansen, Hurwitz, and Madow 1953) is magnified in comparison with small subgroups with an average of just a few students in the classroom.

5.8. Estimation and Generalization of Design Effects

The design effect (DEFF) is the ratio of the variance of a statistic (square of the standard error), taking into account the stratification and clustering in the design, to the variance of the statistic that would have been achieved if the sample had been drawn as a simple random sample (i.e., without stratification or clustering). Except for binomial variables (and then with some limiting assumptions), it is not feasible to reconstruct the variance that could have been achieved under simple random sampling.

It has become customary to compare the achieved variance with the variance computed by ignoring the design, that is, using the data drawn from the design but considering those data as a simple random sample. This approximation contains the positive effect of stratification, but ignores the effect of clustering. Since the effect of clustering tends to dominate the difference between the design variance and the simple random sample variance, the approximation yields estimates of the design effects that are useful in evaluating the design.

Table 5-9. Jackknife estimates of standard errors for grade 4

Variable	Category	n	Standard errors		
			Narrative	Document	Expository
FCOMP	No Parents	159	9.43	6.95	8.34
FCOMP	Step Parents Only	209	5.85	6.59	5.01
FCOMP	Mother Only	671	5.29	4.37	4.03
FCOMP	Mother & Stepfather	428	5.04	3.45	5.53
FCOMP	Father Only	224	5.86	6.08	4.97
FCOMP	Father & Stepmother	165	7.38	6.64	7.97
FCOMP	Mother & Father	3,590	3.13	2.55	2.75
FCOMP	Odd Groupings	802	4.89	4.26	4.54
LANG	Home Eng./Fam. Eng.	4,657	3.13	2.63	3.06
LANG	Home Eng./Fam. Other	86	7.80	6.40	7.60
LANG	Home Other/Fam. Eng.	1,004	3.20	2.90	3.37
LANG	Home Other/Fam. Other	501	5.75	4.82	5.13
WASETH	Asian	246	8.39	7.14	7.21
WASETH	American Indian	195	11.74	8.57	7.43
WASETH	Hispanic	541	3.94	5.29	5.14
WASETH	White	4,219	2.26	2.11	2.63
WASETH	Black	1,047	4.28	3.06	4.82
WASFED	Unknown	203	9.68	7.35	8.26
WASFED	Less than H.S.	607	6.11	4.77	4.21
WASFED	High School	1,454	3.47	2.69	3.42
WASFED	Some College	1,058	4.10	3.27	2.96
WASFED	College/University	2,926	3.36	3.07	3.44
WASMED	Unknown	57	10.32	8.09	10.15
WASMED	Less than H.S.	547	4.72	4.13	4.50
WASMED	High School	1,631	4.15	3.10	3.20
WASMED	Some College	1,274	4.41	3.22	3.66
WASMED	College/University	2,739	3.06	3.02	3.09
WASSEX	Male	3,153	3.63	3.02	3.18
WASSEX	Female	3,095	3.10	2.81	3.03
XTND	Nuclear Family	4,016	2.61	2.47	2.74
XTND	Extended Family	2,232	3.64	2.57	3.39
REGION	Northeast	1,008	9.77	9.67	9.57
REGION	Southeast	1,622	6.47	4.25	5.65
REGION	Central	1,568	5.74	4.36	5.40
REGION	West	2,050	3.54	2.74	3.39
WACTYC	Rural	1,099	7.62	6.38	5.92
WACTYC	Small Town	1,290	6.18	6.12	6.34
WACTYC	50k-100k City	774	7.56	6.75	7.16
WACTYC	50k-100k Suburb	512	7.31	6.98	6.29
WACTYC	100k-500k City	808	8.22	8.10	6.85
WACTYC	100k-500k Suburb	641	10.14	6.78	9.24
WACTYC	Over 500k City	432	15.04	11.04	13.91
WACTYC	Over 500k Suburb	644	10.23	9.20	9.57

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-10. Jackknife estimates of standard errors for grade 9

Variable	Category	n	Standard errors		
			Narrative	Document	Expository
FCOMP	No Parents	58	16.00	9.18	17.16
FCOMP	Step Parents Only	190	10.45	6.43	7.13
FCOMP	Mother Only	422	6.69	5.47	7.48
FCOMP	Mother & Stepfather	318	7.05	7.09	7.83
FCOMP	Father Only	83	10.80	11.02	12.11
FCOMP	Father & Stepmother	114	13.66	11.27	12.57
FCOMP	Mother & Father	1,945	5.54	3.99	6.29
FCOMP	Odd Groupings	169	12.35	9.95	12.20
LANG	Home Eng./Fam. Eng.	2,480	4.96	4.04	5.86
LANG	Home Eng./Fam. Other	56	12.11	9.52	14.11
LANG	Home Other/Fam. Eng.	388	6.47	6.90	8.17
LANG	Home Other/Fam. Other	285	9.01	7.81	12.30
WBSETH	Asian	114	12.12	9.19	12.55
WBSETH	American Indian	89	17.63	12.27	21.25
WBSETH	Hispanic	269	11.29	7.97	10.58
WBSETH	White	2,338	4.50	3.73	5.36
WBSETH	Black	399	11.60	9.49	13.06
WBSFED	Elementary	39	20.42	16.10	20.34
WBSFED	Junior High School	82	13.54	13.45	16.21
WBSFED	Some H.S.	238	8.40	8.79	9.26
WBSFED	High School	1,044	5.49	4.24	6.52
WBSFED	Some College	622	7.08	5.90	7.12
WBSFED	College/University	1,138	5.03	4.17	5.65
WBSMED	Elementary	29*	16.24	13.12	20.28
WBSMED	Junior High School	71	11.59	11.98	10.91
WBSMED	Some H.S.	246	9.74	7.48	8.78
WBSMED	High School	1,104	6.09	4.59	7.54
WBSMED	Some College	781	5.48	4.95	6.81
WBSMED	College/University	970	4.81	3.79	5.31
WBSSEX	Male	1,583	6.23	4.89	7.50
WBSSEX	Female	1,626	4.99	3.97	5.74
XTND	Nuclear Family	2,691	4.90	3.79	5.68
XTND	Extended Family	518	7.54	5.54	8.15
REGION	Northeast	524	15.68	10.52	17.70
REGION	Southeast	878	7.70	7.37	10.02
REGION	Central	819	8.22	7.40	10.00
REGION	West	988	8.58	6.09	9.33
WBCTYC	Rural	635	7.53	6.99	10.11
WBCTYC	Small Town	831	7.41	6.21	7.86
WBCTYC	50k-100k City	320	25.44	15.99	26.46
WBCTYC	50k-100k Suburb	166	28.18	20.49	30.90
WBCTYC	100k-500k City	268	12.36	10.76	15.82
WBCTYC	100k-500k Suburb	259	18.36	11.09	19.97
WBCTYC	Over 500k City	257	26.34	23.19	36.34
WBCTYC	Over 500k Suburb	473	14.01	10.28	16.15

*Although NCES does not normally publish estimates based on samples smaller than 30, in this case the estimates are standard errors (not statistics of substantive interest). Furthermore, the estimates are part of an overall presentation to indicate trend. They should not be used as reliable estimates in their own right.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Since the estimated design effects show, primarily, the effect of clustering, they tend to be small (in the neighborhood of 1.0) for characteristics that do not differ greatly from cluster to cluster such as the gender of students in public schools. However, school policies, neighborhood environments, and instructional methods may combine to cause variation in the differences between the sexes in test scores, thus causing a substantial design effect. Also, design effects tend to be small for small subsets of the population and large for large subsets. The reason, as explained above, is that the clusters tend to contain larger numbers of the members of the subset.

The design effects for the three scales, and the same population subgroups as for the variances, are given in Tables 5-11 and 5-12. One way to use the design effects is to divide the actual sample size, n, by the design effect to achieve an "effective" sample size, that is, the size of a simple random sample that would have produced the same precision as the design sample size. For example, 1,047 grade 4 students were black. The design effect for the narrative reading scale for this subgroup was estimated to be 2.45, so the effective sample size was about 427. It should be remembered, when making such interpretations, that the DEFF estimates are subject to a substantial amount of sampling error since the number of schools producing members of the subclass is small. Design effects of less than 1.0 typically are associated with small subgroup sizes and with characteristics that are thinly distributed over the entire sample, that is, that are not clustered. In general, because of the sampling error, these estimates should be considered as being near 1.0.

A few characteristics have unusually large design effects. They include the Northeast region and cities with over 500,000 population for both grade 4 and 9 and cities with from 50,000 to 100,000 population for grade 9. These large values indicate a homogeneity within schools and a lack of homogeneity between schools in the strata from which these students were drawn. The sample sizes, in terms of schools, are so small, however, that one can not generalize too broadly from these data.

Although the design effects clearly have an effect on the standard errors, a substantial part of their size is related to sample size. It seems possible, then, that one might be able to make estimates of standard errors, as functions of sample size, that would be sufficiently accurate for most analytic purposes. Various transformations of both subgroup sample size and the standard errors were tried in order to find a linear relationship between the transformed standard errors and the transformed subgroup sample sizes. It was found that the inverse of the standard errors was approximately a linear function of the cube root of subgroup sample size for variables that are well distributed over the population, that is, for variables that are not identified with one or more specific geographic areas, such as region or urbanicity. Those variables, of course, were part of the stratification of the two universes, so that there is, for example, only one category of "region" that contains all of the sampled schools in the Southeast. In general, if a variable category does not cross major stratum boundaries it was left out of the attempt to generalize standard errors.

Figures 5-1 and 5-2 show the relationship between the inverses of the subgroup standard errors and the cube roots of the subgroup sample sizes for the variables that are not geographic in nature (that is, excluding region and urbanicity) for the narrative scales of both grades 4 and 9. (The corresponding figures for the other two scales were similar and are not shown.)

Table 5-11. Estimated design effects for grade 4

Variable	Category	n	Design effect		
			Narrative	Document	Expository
FCOMP	No Parents	159	1.56	1.17	1.66
FCOMP	Step Parents Only	209	0.87	1.93	1.12
FCOMP	Mother Only	671	1.92	2.05	1.63
FCOMP	Mother & Stepfather	428	1.30	0.89	2.39
FCOMP	Father Only	224	0.86	1.27	0.98
FCOMP	Father & Stepmother	165	1.00	1.35	1.63
FCOMP	Mother & Father	3,590	4.00	3.61	4.39
FCOMP	Odd Groupings	802	2.13	2.70	2.75
LANG	Home Eng./Fam. Eng.	4,657	4.90	4.90	6.77
LANG	Home Eng./Fam. Other	86	0.74	0.71	0.75
LANG	Home Other/Fam. Eng.	1,004	1.10	1.35	1.94
LANG	Home Other/Fam. Other	501	2.00	1.94	2.26
WASETH	Asian	246	1.66	1.53	1.74
WASETH	American Indian	195	2.68	2.16	2.10
WASETH	Hispanic	541	1.09	2.64	2.75
WASETH	White	4,219	2.45	3.02	4.66
WASETH	Black	1,047	2.45	2.15	5.02
WASFED	Unknown	203	1.84	1.77	1.96
WASFED	Less than H.S.	607	2.52	2.28	1.98
WASFED	High School	1,454	2.04	1.87	2.93
WASFED	Some College	1,058	1.96	1.75	1.55
WASFED	College/University	2,926	3.59	4.02	5.31
WASMED	Unknown	57	0.71	0.72	1.19
WASMED	Less than H.S.	547	1.52	1.69	1.98
WASMED	High School	1,631	3.14	2.58	2.84
WASMED	Some College	1,274	2.79	2.14	2.80
WASMED	College/University	2,739	2.71	3.65	3.97
WASSEX	Male	3,153	4.33	3.99	4.89
WASSEX	Female	3,095	3.38	4.12	4.63
XTND	Nuclear Family	4,016	2.99	3.68	4.67
XTND	Extended Family	2,232	3.46	2.61	4.62
REGION	Northeast	1,008	11.38	14.20	14.77
REGION	Southeast	1,622	7.40	4.83	8.78
REGION	Central	1,568	5.59	4.65	7.18
REGION	West	2,050	2.73	2.24	3.60
WACTYC	Rural	1,099	7.01	6.73	5.88
WACTYC	Small Town	1,290	5.11	7.59	8.75
WACTYC	50k-100k City	774	4.79	5.47	6.33
WACTYC	50k-100k Suburb	512	3.51	4.19	3.24
WACTYC	100k-500k City	808	5.96	8.25	6.49
WACTYC	100k-500k Suburb	641	7.01	4.60	8.46
WACTYC	Over 500k City	432	10.70	7.98	12.61
WACTYC	Over 500k Suburb	644	7.47	8.38	9.28

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-12. Estimated design effects for grade 9

Variable	Category	n	Design effect		
			Narrative	Document	Expository
FCOMP	No Parents	58	1.51	0.85	1.11
FCOMP	Step Parents Only	100	1.33	0.91	0.62
FCOMP	Mother Only	422	2.08	1.94	2.20
FCOMP	Mother & Stepfather	318	1.98	2.65	1.94
FCOMP	Father Only	83	1.32	1.12	1.11
FCOMP	Father & Stepmother	114	2.01	2.93	1.70
FCOMP	Mother & Father	1,945	6.32	4.54	6.81
FCOMP	Odd Groupings	169	2.46	2.33	2.43
LANG	Home Eng./Fam. Eng.	2,480	6.37	5.94	7.55
LANG	Home Eng./Fam. Other	56	0.73	0.75	1.04
LANG	Home Other/Fam. Eng.	388	1.88	2.73	2.39
LANG	Home Other/Fam. Other	285	3.05	3.40	4.31
WBSETH	Asian	114	1.81	1.42	1.66
WBSETH	American Indian	89	2.96	1.92	5.08
WBSETH	Hispanic	269	4.62	2.93	3.30
WBSETH	White	2,338	5.39	5.26	6.31
WBSETH	Black	399	6.09	6.45	7.01
WBSFED	Elementary	39	1.58	1.43	1.27
WBSFED	Junior High School	82	2.20	2.32	2.80
WBSFED	Some H.S.	238	1.93	3.20	2.33
WBSFED	High School	1,044	3.52	3.07	4.25
WBSFED	Some College	622	3.55	3.22	2.81
WBSFED	College/University	1,138	3.14	3.01	3.39
WBSMED	Elementary	29*	0.67	0.70	0.93
WBSMED	Junior High School	71	1.46	2.08	1.41
WBSMED	Some H.S.	246	2.58	2.34	2.15
WBSMED	High School	1,104	4.63	3.81	6.22
WBSMED	Some College	781	2.66	2.96	3.33
WBSMED	College/University	970	2.28	1.91	2.29
WBSSEX	Male	1,583	6.37	5.17	7.59
WBSSEX	Female	1,626	4.53	4.13	5.01
XTND	Nuclear Family	2,691	7.00	5.77	7.81
XTND	Extended Family	518	3.01	2.47	3.24
REGION	Northeast	524	13.99	9.23	15.67
REGION	Southeast	878	5.69	7.83	8.52
REGION	Central	819	5.94	6.53	6.46
REGION	West	988	7.95	5.36	8.48
WBCTYC	Rural	635	4.48	4.57	6.45
WBCTYC	Small Town	831	5.13	5.02	4.88
WBCTYC	50k-100k City	320	21.31	12.04	18.59
WBCTYC	50k-100k Suburb	166	13.09	8.94	13.85
WBCTYC	100k-500k City	268	3.91	4.24	5.27
WBCTYC	100k-500k Suburb	259	9.19	6.02	10.51
WBCTYC	Over 500k City	257	20.42	20.96	26.97
WBCTYC	Over 500k Suburb	473	9.10	7.50	11.34

*Although NCES does not normally publish estimates based on samples smaller than 30, in this case the estimates are design effects (not statistics of substantive interest). Furthermore, the estimates are part of an overall presentation to indicate trend. They should not be used as reliable estimates in their own right.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 5-1. Transformed narrative scale standard errors: Grade 4

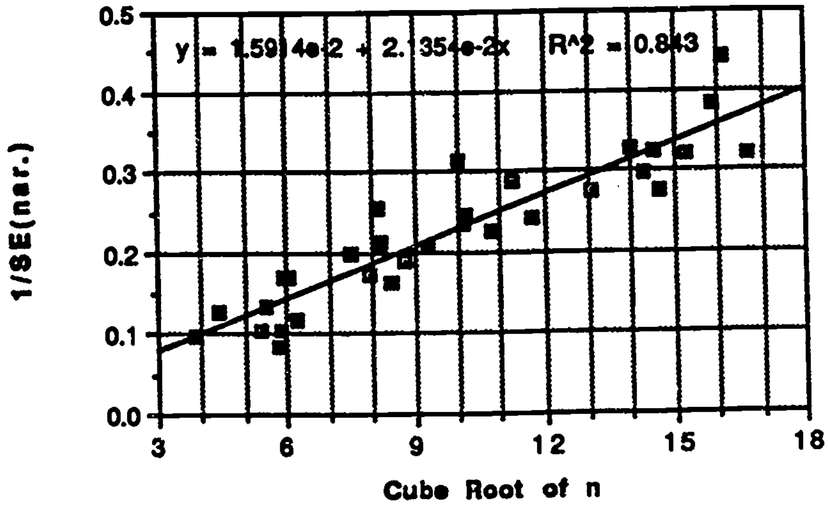
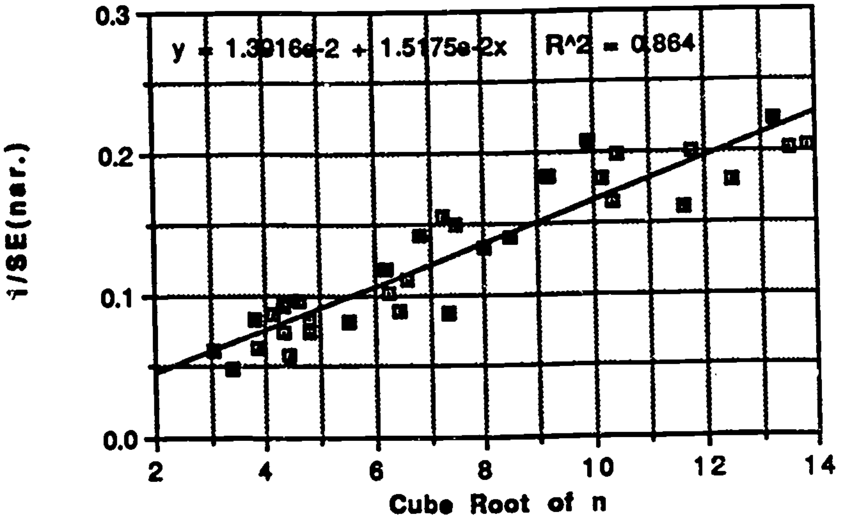


Figure 5-2. Transformed narrative scale standard errors: Grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

These line fits are quite good, the squares of the correlation coefficients being 0.84 and 0.86, indicating that about 85 percent of the variation on the transformed standard errors is accounted for by variation in the subgroup sample size, n. For the document scales the R squares were 0.85 and 0.88 for grades 4 and 9, respectively, and for the expository scales the R squares were 0.86 and 0.77, respectively. These are all considered quite good, indicating that the standard errors may be estimated by the linear transformation shown.

The generalized estimates, when converted to the untransformed scales, are shown for both grades 4 and 9 in Tables 5-13 and 5-14. These estimates may be used, with linear interpolation between adjacent subclasses, as needed, in lieu of computing the standard errors from the data. They are not appropriate, however, for subclasses that do not cross-cut all of the strata, such as region and urbanicity. Those variables require separate estimation.

Table 5-13. Generalized standard errors for grade 4

Subclass size (n)	Narrative scale	Document scale	Expository scale
100	8.7	7.2	7.4
200	7.1	5.9	6.3
300	6.3	5.3	5.7
400	5.8	4.9	5.3
500	5.4	4.6	5.0
600	5.1	4.3	4.7
700	4.9	4.1	4.5
800	4.7	4.0	4.4
900	4.5	3.8	4.2
1000	4.4	3.7	4.1
1200	4.1	3.5	3.9
1400	3.9	3.4	3.7
1600	3.8	3.2	3.6
1800	3.6	3.1	3.5
2000	3.5	3.0	3.4
2500	3.3	2.8	3.2
3000	3.1	2.7	3.0
3500	2.9	2.5	2.9
4000	2.8	2.4	2.8
4500	2.7	2.3	2.7
5000	2.6	2.3	2.6

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 5-14. Generalized standard errors for grade 9

Subclass size (n)	Narrative scale	Document scale	Expository scale
100	11.9	9.8	12.3
200	9.8	8.0	10.5
300	8.7	7.1	9.5
400	8.0	6.5	8.8
500	7.5	6.0	8.3
600	7.1	5.7	7.9
700	6.7	5.4	7.6
800	6.5	5.2	7.4
900	6.2	5.0	7.1
1000	6.0	4.9	6.9
1200	5.7	4.6	6.6
1400	5.5	4.4	6.3
1600	5.2	4.2	6.1
1800	5.0	4.0	5.9
2000	4.9	3.9	5.7
2500	4.6	3.6	5.4
3000	4.3	3.4	5.1
3500	4.1	3.3	4.9
4000	3.9	3.1	4.7
4500	3.8	3.0	4.5
5000	3.7	2.9	4.4

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The standard errors estimated from this relationship are, of course, subject to errors in the estimation of the true relationship. But the individual estimates of the standard errors are subject to a substantial amount of sampling error also. Thus, in some instances, more credibility can be attached to the generalized standard errors than to the individually estimated standard errors. To provide some measure of the reliability of the estimates, the standard deviation around the fitted curves (in untransformed units) was computed. This approximates the standard deviation of the standard error. For grade 4 the standard deviations were 0.74 for the narrative scale, 0.78 for the document scale, and 0.83 for the expository scale. The corresponding figures for grade 9 were 1.83, 1.48, and 2.54, the larger figures being at least partially attributable to the smaller sample size of grade 9.

References

Hansen, M.H. (1989). Comparison of jackknife estimates of standard errors of the mean and median. Memorandum to Eugene Johnson, November 30, 1989.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample survey methods and theory*. Vol. 1. New York: John Wiley & Sons.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys, *Journal of Official Statistics*, 4, 381-397.

6. CONFIDENTIALITY

6.1. Introduction

Since the study data were to be delivered to the International Coordinating Center (ICC), and ultimately were to be released as a public use data file in the U.S., it was very important to ensure that the data were in a form that did not breach any confidentiality provisions. Specifically, we needed to ensure that the collection and release of the data would not permit identification of the individual schools and students that participated in the study. This chapter describes how this confidentiality was ensured.

6.2. Data Files

Two types of data files were created for the IEA Reading Literacy Study: the U.S. files and the international files. The study produced eight U.S. files in all, four files for each population. For each population, the U.S. files, which to a large extent followed the test and questionnaire layout, consisted of the following files:

- A student file containing the student responses to the Reading Literacy Test items;
- A student file containing the student background information from the Student Questionnaire, as well as the teacher- reported race/ethnicity and the teacher-assigned student reading literacy levels;
- A Teacher Questionnaire file; and
- A School Questionnaire file.

The eight U.S. files were combined and reformatted in accordance with the IEA specifications to produce six international files. For each population, the two U.S. student files (i.e., the Reading Literacy Test item response and Student Questionnaire files) were mapped onto a single IEA file. Additionally, the U.S. Teacher and School Questionnaire files were mapped onto corresponding IEA versions. Information describing the differences between the U.S. and the IEA files is available from Westat.

6.3. Public Use Data Files

The U.S. international files were designed to permit the IEA, or any individual or research organization with an interest in the IEA Reading Literacy Study, to perform comparative analyses. More than 30 countries participated in the IEA Reading Literacy Study. To facilitate comparative analyses of the data from their national studies, the IEA developed extensive data processing procedures -- data from each country were machine edited, errors identified and reconciled, and internationally defined constructs were derived. These procedures were designed to ensure the consistency of data across countries participating in the study.

As part of the agreement to participate in the IEA Reading Literacy Study, each participating country, including the U.S., had granted the IEA Headquarters permission to release its data to individuals or organizations desiring to perform secondary analyses. To avoid disclosure problems, the U.S. files submitted to IEA were considered as public use data files, and extensive analyses were performed to ensure that individual respondents would not be identified. Westat also employed additional security procedures to ensure the confidentiality of the IEA data. The following sections include a description of Westat's security procedures, followed by a description of the confidentiality analyses.

6.4. Security Procedures

Assuring the confidentiality of data is an important component of the overall data security procedures at Westat. For the IEA Reading Literacy Study, data security referred to protection of data against accidental or intentional disclosure to unauthorized persons or unauthorized modifications or destruction. Because the data files contained sensitive data about students, teachers, and schools, security measures were established to prevent the data from being lost, stolen, or otherwise subjected to unauthorized access. The data also were protected from hardware or software failures, from catastrophes, and from unauthorized use.

Westat security procedures used to assure confidentiality of data involved three levels of data protection:

- Administrative controls;
- Physical controls; and
- Technical controls.

6.4.1. Administrative Controls

Administrative controls were necessary to ensure that explicit procedures for securely handling confidential data were in place and were understood by the staff processing the data. These administrative controls entailed the following components:

- **Modifying the IEA Student Name Form.** The IEA-proposed Student Name Form required that the name of each participating student be printed on the form. To ensure confidentiality, we devised a five-copy noncarbon form that included the student names on only the first copy, which was retained in the schools. The other four copies, which identified students only by ID, were shipped to Westat and DRC.

- **Assigning security responsibility.** The security procedures at the computer site was the overall responsibility of Westat's director of data processing.

- **Signing confidentiality agreement.** As a condition of employment, Westat personnel associated with survey efforts, including data processing, operations, field, and professional staff, all work under the terms of an individually signed confidentiality agreement (Exhibit 6-1). Thus, all staff working for the IEA Reading Literacy Study had already signed a confidentiality agreement.

Exhibit 6-1. Westat confidentiality agreement

WESTAT, INC.

EMPLOYEE OR CONTRACTOR'S ASSURANCE OF CONFIDENTIALITY OF SURVEY DATA

Statement of Policy

Westat is firmly committed to the principle that the confidentiality of individual data obtained through Westat surveys must be protected. This principle holds whether or not any specific guarantee of confidentiality was given at time of interview (or self-response), or whether or not there are specific contractual obligations to the client. When guarantees have been given or contractual obligations regarding confidentiality have been entered into, they may impose additional requirements which are to be adhered to strictly.

Procedures for Maintaining Confidentiality

1. All Westat employees and field workers shall sign this assurance of confidentiality. This assurance may be superseded by another assurance for a particular project.

2. Field workers shall keep completely confidential the names of respondents, all information or opinions collected in the course of interviews, and any information about respondents learned incidentally during field work. Field workers shall exercise reasonable caution to prevent access by others to survey data in their possession.

3. Unless specifically instructed otherwise for a particular project, an employee or field worker, upon encountering a respondent or information pertaining to a respondent that s/he knows personally, shall immediately terminate the activity and contact her/his supervisor for instructions.

4. Survey data containing personal identifiers in Westat offices shall be kept in a locked container or a locked room when not being used each working day in routine survey activities. Reasonable caution shall be exercised in limiting access to survey data to only those persons who are working on the specific project and who have been instructed in the applicable confidentiality requirements for that project.

   Where survey data have been determined to be particularly sensitive by the Corporate Officer in charge of the project or the President of Westat, such survey data shall be kept in locked containers or in a locked room except when actually being used and attended by a staff member who has signed this pledge.

5. Ordinarily, serial numbers shall be assigned to respondents prior to creating a machine-processible record and identifiers such as name, address, and Social Security number shall not, ordinarily, be a part of the machine record. When identifiers are part of the machine data record, Westat's Manager of Data Processing shall be responsible for determining adequate confidentiality measures in consultation with the project director. When a separate file is set up containing identifiers or linkage information which could be used to identify data records, this separate file shall be kept locked up when not actually being used each day in routine survey activities.

6. When records with identifiers are to be transmitted to another party, such as for keypunching or key taping, the other party shall be informed of these procedures and shall sign an Assurance of Confidentiality form.

7. Each project director shall be responsible for ensuring that all personnel and contractors involved in handling survey data on a project are instructed in these procedures throughout the period of survey performance. When there are specific contractual obligations to the client regarding confidentiality, the project director shall develop additional procedures to comply with these obligations and shall instruct field staff, clerical staff, consultants, and any other persons who work on the project in these additional procedures. At the end of the period of survey performance, the project director shall arrange for proper storage or disposition of survey data including any particular contractual requirements for storage or disposition. When required to turn over survey data to our clients, we must provide proper safeguards to ensure confidentiality up to the time of delivery.

8. Project directors shall ensure that survey practices adhere to the provisions of the U.S. Privacy Act of 1974 with regard to surveys of individuals for the Federal Government. Project directors must ensure that procedures are established in each survey to inform each respondent of the authority for the survey, the purpose and use of the survey, the voluntary nature of the survey (where applicable) and the effects on the respondents, if any, of not responding.

PLEDGE

I hereby certify that I have carefully read and will cooperate fully with the above procedures. I will keep completely confidential all information arising from surveys concerning individual respondents to which I gain access. I will not discuss, disclose, disseminate, or provide access to survey data and identifiers except as authorized by Westat. In addition, I will comply with any additional procedures established by Westat for a particular contract. I will devote my best efforts to ensure that there is compliance with the required procedures by personnel whom I supervise. I understand that violation of this pledge is sufficient grounds for disciplinary action, including dismissal. I also understand that violation of the privacy rights of individuals through such unauthorized discussion, disclosure, dissemination, or access may make me subject to criminal or civil penalties. I give my personal pledge that I shall abide by this assurance of confidentiality.

____
Signature

- **Developing formal procedures.** Formal procedures were developed to assure proper handling and disposition of paper and machine-readable data. Data transfer was controlled and formally handled by coordination with National Center for Education Statistics (NCES).

6.4.2. Physical Controls

Physical security included the use of key-card locks on the doors, fire prevention precautions through halon systems, and onsite fireproof vault protection of stored data files as well as offsite backup storage of data files.

6.4.3. Technical Controls

Technical security included controlling who could access the data and what actions users could take once they had access. Some of the data files for the study resided on MicroVax computers at Westat. The VAX/VMS operating system provided a sound system for preventing unauthorized access to the system and to all sensitive disk and tape files. In order to gain access to the system, each user was assigned a username and a password. The password was known only to the user and stored in an encoded form that could not be read by other users, even the system manager. Passwords were changed periodically to prevent unauthorized access by other users.

6.5. Confidentiality Analysis

Legislation passed by the Congress in 1988 required that NCES strengthen its efforts to ensure confidentiality, and mandated severe penalties for failure to do so. The problem faced by study staff was that in the IEA Reading Literacy Study student responses needed to be linked to information collected from school principals and teachers of sampled classrooms. The capability to link school/teacher and student information could result in problems related to disclosure (i.e., schools, unlike students, might be identified by matching against various national files that are in the public domain).

To determine whether or not schools sampled in the IEA Reading Literacy Study could be identified by matching against two national files, Quality Education Data (QED) and Common Core Data (CCD), four steps were conducted (details of these steps are discussed in Sections 6.5.1 through 6.5.4):

1. The variables comparable between the Reading Literacy Study and the QED and CCD files were identified;

2. Classification variables were selected to categorize the IEA sampled schools;

3. Euclidean distance measures were created for each pair of IEA schools and QED schools and for each pair of IEA and CCD schools in the corresponding subsets defined by the classification variables identified in step 2; and

4. For each IEA sampled school, the closest school (i.e., the school with the minimum distance measure) was identified. If more than one school had the same minimum distance, the school was categorized as "not matched." If there was a unique closest

school, additional variables (available on Westat's file but not on IEA files) were used to determine whether the schools matched.

The analyses showed that none of the schools in the IEA sample could be identified using the QED or CCD files. The details of the analyses are discussed in the following sections.

6.5.1. Identifying Comparable Variables

Comparable variables that existed (or could be constructed from existing variables) on the IEA, QED, and CCD files are shown in Table 6-1.

Table 6-1. Comparable variables on the IEA, QED, and CCD files: Grades 4 and 9

IEA variable	QED variable	CCD variable*
Grade (4 or 9)	Grade (4 or 9)	Grade (4 or 9)
Number of teachers	Number of teachers	Number of teachers
Number of students	Number of students	Number of students
Number of 4th grade students (Pop A)	Not available	Number of 4th grade students
Number of 9th grade students (Pop B)	Not available	Number of 9th grade students
Library (Y/N)	Library (Y/N)	Not available
Locale	Locale	Locale
Type	Type	Type

*CCD variables apply to public schools only.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Thus, comparable variables were found for a total of eight IEA variables on the QED or CCD data files. Of these, five variables existed on both QED and CCD data files, whereas, three variables existed on only one of the two national files. The variable "type" had two categories (public and private) and was consistently coded across all files except for the CCD file, which had public only. Each of the data files, however, used a different scale for recording the type of locale in which the school was located.

The IEA file used a four-point scale:

1. Rural;
2. Small town;
3. Large town; and
4. City of 1 million persons or more.

The QED file used a three-point scale:

1. Urban--Central city;
2. Suburban--Area surrounding central city, but still located within the counties constituting the MSA; and
3. Rural--Area outside any MSA.

The CCD file used a seven-point scale:

1. Large central city (400,000 or more population or 6,000 or more persons per square mile);
2. Mid-size central city;
3. Urban fringe of large city (within MSA of 1 above and defined as urban by U.S. Census Bureau);
4. Urban fringe of mid-size city (within MSA of 2 above and defined as urban by U.S. Census Bureau);
5. Large town (outside of MSA, 25,000 or more population, and defined as urban by U.S. Census Bureau);
6. Small town (outside of MSA, under 25,000 population but greater than 2,499, and defined as urban by U.S. Census Bureau); and
7. Rural (population of less than 2,500 and defined as rural by U.S. Census Bureau).

For the purpose of this analysis, the IEA locale variable was transformed so that the values increased in the same direction as the QED and CCD values. Thus, the revised IEA locale variable had the following values:

1. City of 1 million persons or more;
2. Large town;
3. Small town; and
4. Rural area.

This revised variable was used as the input to the standardization procedure described above. No manipulation was performed on the QED and CCD variables prior to standardization.

6.5.2. Selecting Classification Variables

Based on the assumption that the probability of matching a school will increase if matching is done within well-defined subgroups, classification variables were selected to categorize the IEA sampled schools. Thus, each subset of the IEA schools was "matched" with all schools in the QED or CCD files. Moreover, two alternative approaches were examined:

1) Type (public/private) and grade (4th/9th) were selected as the classification variables and Euclidean distance measures for each pair of IEA schools and QED schools and for each pair of IEA and CCD schools in the corresponding subsets were estimated.

2) All categorical variables (i.e., type, grade, library, and locale) were used as the classification variables, and Euclidean distance measures for each pair in the corresponding subsets were estimated.

The locale variables required special treatment when classifying the schools into categories for the analyses in which only continuous variables were included in the distance measures. Because some of the QED locality variables were missing, the unknown localities were included in the pool of "unknown" schools for each of the analyses. Since there are four levels of locale in the IEA variable but only three levels in the QED variable, the analyses were conducted by comparing each level of the QED variable with two levels of the IEA variable. The levels compared are indicated in Table 6-2.

Table 6-2. Comparison of IEA and QED locale variables for analyses using locale as a classification variable

IEA variable			QED variable	
Original	Revised	Label	Code	Label
4	1	City of 1 million	1	Urban + Unknown
3	2	Large town		
3	2	Large town	2	Suburban + Unknown
2	3	Small town		
2	3	Small town	3	Rural + Unknown
1	4	Village		

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The CCD locale variable was first transformed to a three-level variable corresponding to QED categories. This transformation was achieved by combining the following CCD categories (Table 6-3).

Table 6-3. Transforming the CCD locale variable to corresponding QED locale variables

QED category	CCD categories
1	1-2
2	3-4
3	5-7

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

6.5.3. Estimating Euclidean Distance Measures

Prior to calculating the Euclidean distance measures for each pair of schools, the values of the analysis variables identified were standardized. SAS PROC STANDARD was used for the standardization with MEAN=0 and STD=1.

Separate standardization computations were performed for each of the six files used in the analyses:

- Fourth Grade CCD file
- Ninth Grade CCD file

- Fourth Grade QED file
- Ninth Grade QED file

- Fourth Grade IEA file
- Ninth Grade IEA file

The Euclidean distance measure for a pair of schools was constructed by squaring the differences between corresponding variables associated with the two schools, summing the squared differences, and taking the square root of the result. Thus, the difference between IEA school I and known (QED or CCD) school K is given by the formula:

$$D^{IK} = [\Sigma\ (V^{IJ} - V^{KJ})^2]^{1/2}$$

Where:

N is the number of variables used in the analysis;

V_{IJ} is the value of variable j for school I in the IEA file; and

V_{KJ} is the value of variable j for school K in the QED or CCD file.

6.5.4. Determining Closest Schools

For each IEA sampled school, the closest school (the school with the minimum distance measure) was identified. If more than one school had the same minimum distance, the school was categorized as "not matched." If there was a unique closest school, additional variables (available on Westat's files but not on IEA files) were used to determine whether the schools matched. As a first step to determining this match, the state identifications were compared. If the pair of closest schools were within different states, the schools were also categorized as "not matched." If the same state was indicated for the pair of closest schools, the schools were then compared on additional variables available on Westat's files. Table 6-4 summarizes the number of IEA schools analyzed, the number of QED or CCD schools compared, the number of closest schools with matched state identification, and the actual number of schools matched.

As shown in Table 6-4, for some of the closest schools (i.e., pair of schools with minimum measure of distance) the state identification matched (i.e., the IEA school and the QED or CCD school were in the same state); nevertheless, in all of these cases additional information revealed that the schools were different. Thus, on the basis of the analyses performed, it is shown that schools on the IEA files cannot be identified by matching them to schools on QED or CCD files using descriptive variables on the IEA files.

Table 6-4. Number of IEA schools analyzed, QED and CCD schools compared, closest schools with matched state identification, and schools matched for each grade and national file

Grade	File	Classification*	Number IEA schools	Number CCD or QED schools	Number schools with matched state	Number schools matched
4	CCD	1	141	49,446	2	0
4	QED	2	141	65,230	0	0
4	QED	3	25	65,230	0	0
4	CCD	4	64	49,446	4	0
4	CCD	5	87	49,446	14	0
4	CCD	6	75	49,446	15	0
4	QED	7	60	65,230	0	0
4	QED	8	4	65,230	0	0
4	QED	9	84	65,230	0	0
4	QED	10	13	65,230	0	0
4	QED	11	75	65,230	0	0
4	QED	12	18	65,230	0	0
4	QED	13	4	65,230	0	0
4	QED	14	2	65,230	0	0
4	QED	15	3	65,230	0	0
4	QED	16	2	65,230	0	0
4	QED	17	0	65,230	0	0
4	QED	18	1	65,230	0	0
9	CCD	1	147	18,813	12	0
9	QED	2	147	24,038	0	0
9	QED	3	18	24,038	0	0
9	CCD	4	75	18,813	7	0
9	CCD	5	86	18,813	14	0
9	CCD	6	71	18,813	11	0
9	QED	7	75	24,038	0	0
9	QED	8	2	24,038	0	0
9	QED	9	86	24,038	0	0
9	QED	10	8	24,038	0	0
9	QED	11	71	24,038	0	0
9	QED	12	16	24,038	0	0

*Classifications are as follows:

1 = Public only from CCD
2 = Public only
3 = Private only
4 = Small town and village
5 = Large town and small town
6 = City and large town
7 = Has library, rural and unknown, public
8 = Has library, rural and unknown, private
9 = Has library, suburban and unknown, public
10 = Has library, suburban and unknown, private
11 = Has library, urban and unknown, public
12 = Has library, urban and unknown, private
13 = No library, rural and unknown, public
14 = No library, rural and unknown, private
15 = No library, suburban and unknown, public
16 = No library, suburban and unknown, private
17 = No library, urban and unknown, public
18 = No library, urban and unknown, private

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

PART II. MEASURING READING LITERACY

The organizing framework for this section is derived explicitly from the first objective of this study, as put forth by the International Steering Committee, which was "to develop internationally valid instruments for measuring reading literacy suitable for establishing internationally comparable literacy levels in each of the participating countries." To determine if this objective was accomplished, it is necessary to evaluate what was done, how it was done, and ultimately to use this evaluation to arrive at a valid interpretation of the data within the U.S. context.

To accomplish this goal, this part of the technical report has been divided into the following chapters:

- **Chapter 7: The IEA Reading Literacy Test.** *Marilyn Binkley.* A description of the IEA Reading Literacy Test. This will provide a description of the framework, the format, and item types included in the instrument.
- **Chapter 8: The Scaling Procedures.** *Nadir Atash, Marilyn Binkley.* A description and evaluation of the scaling procedures used. This initial look at both the procedures and resulting distribution of scores will provide insight into the meaning of the scores.
- **Chapter 9: Estimates of the Reading Proficiency of U.S. Students.** *Marilyn Binkley, Keith Rust, Nadir Atash.* A description of U.S. students' reading proficiency. This is a close look at the resulting distribution of student performance with some comparison to international and intranational results.
- **Chapter 10: Evidence Supporting the Validity of the IEA Reading Literacy Cognitive Instruments.** *Marilyn Binkley, Nadir Atash, Keith Rust.* A discussion of evidence pertaining to the validity of the instrument, internationally and within the U.S.

7. THE IEA READING LITERACY TEST

7.1. Introduction

The purpose of this chapter is to provide the reader with an overview of the reading literacy cognitive instruments. The chapter provides information on the scope and specification of the test as well as some description and discussion of the test items. Actual test instruments are included as attachments to this report.

7.2. The IEA Framework

Within the context of the IEA Reading Literacy Study, reading literacy was defined as

> *...the ability to understand and use those written language forms required by society and/or valued by the individual. Literacy occurs in a variety of language contexts (e.g., school, home, work, and religious or civic institutions). Reading literacy involves both a range of competencies and a set of habits or practices, arrayed along various dimensions such as reading "stop" signs to being aware of nuances in complex philosophic texts; from reading only what is necessary to pursuing one's further learning and recreation through books and journals.*[1]

From this definition, an operational set of test specifications was developed. These specifications were based on a "...simple three-domain classification - Narrative Prose, Expository Prose, and Documents" and "...six 'skills' or 'mental constructs' which cut across these domains and refer to the kind of mental processing which it is presumed the students use in arriving at their answers" (Elley 1988). Thus, the items can be classified in a two-dimensional grid shown in Exhibit 7-1.

Exhibit 7-1. IEA reading literacy item framework

Domain	Skills assessed					
	Verbatim	Paraphrase	Main theme	Inference	Locate information	Follow directions
Narrative Expository Document						

7.3. Domain Classifications

For the purposes of this study, Elley defined the three text types identified as distinct domains in the following manner.

---

[1] Dr. Warwick Elley is the chairman of the IEA International Steering Committee. The definition is from his 1992 IEA publication *How in the World Do Students Read?*

Narrative Prose *refers to continuous text materials in which the writer's aim was to tell a story, whether fact or fiction. They are normally designed to entertain or involve the reader emotionally; they are written in the past tense, and usually have people or animals as their main theme.*

Expository Prose *refers to continuous text materials designed to describe or explain something. The subjects of such texts are usually things, but they may be written in the present or the past, the style is typically impersonal, highlighting such features as definitions, causes, classifications, functions, contrasts and examples, rather than a moving plot with climax.*

Documents *refer to structured, tabular texts, such as forms, charts, labels, graphs, lists, and sets of instructions where the reading requirements typically involve locating information or following directions, rather than continuous reading of connected text.*

Seven criteria were established for the selection of passages: "That they are unfamiliar to students; that they are suitable for all countries, languages, sexes; that they will not date quickly; that they are well written and interesting; that they present new information, such that students cannot answer without reading the passage; that they range in length and in difficulty level; and that they provide a diversity of topics and styles" (IEA Coordinating Committee internal document 1989, 33).

Passages and exemplars of documents were submitted by participating countries. Some of these passages were actual texts that students would encounter in their reading, while others were constructed especially for the assessment. In this process, an effort was made to avoid those topics that were extremely biased with regard to particular cultures. The texts used in the expository and narrative domains represent a reasonable range of topics, with a slightly heavier emphasis on animal-related texts.

As noted in Table 7-1, the passages ranged in length from short to moderate. None were exceedingly long, although a few did have sufficient length to develop character or comparisons. The grade 4 narrative texts varied from 292 words to 706 words; grade 9, from 402 words to 1,130 words. In both cases, these texts were somewhat shorter than those that students in the U.S. are likely to read as part of either classwork or leisure activities.

Table 7-1. Number of words per passage by grade and text type

Grade 4				Grade 9			
Narrative		Expository		Narrative		Expository	
The Bird and the Elephant	292	Postcard	56	Killing the Fox	402	Marmots	222
Grandpa	310	Quicksand	141	A Shark Makes Friends	452	Paracutin	260
A Shark Makes Friends	452	Walrus	207	Mute	686	A Woman Learns To Read	282
No Dogs Is Not Enough	706	Marmots	222	Magician's Revenge	718	Smoke	364
		Trees	383	Listen to the Angel's Laughing	1130	The Promise of the Laser	760

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The expository texts also ranged from short to moderate--in grade 4, from 56 words to 383 words; in grade 9, from 222 to 760 words. None of these passages would be considered sufficiently long, either in word length or in conceptual range, to constitute a chapter from a textbook. None give enough information to formulate full understanding of the concepts related to the theme of the passage. However, these passages do have the characteristic qualities of articles, presenting a limited argument or a theme.

The documents were short displays of information in tabular formats, maps, or graphs that were specifically developed for this assessment. While fairly representative of the range of documents one might find in the students' environments, they were not within any context. Therefore, the purpose for referring to the documents was artificial.

7.4. The Skills Assessed

For this assessment, Elley identified six skills that cut across the domains. They were defined in the following ways:

- "**Verbatim** items require the student merely to match the words of the item with those of the text.
- "**Paraphrase** items require the student to choose or compose an answer which is explicitly stated in the text, but is expressed in words different from that of the item.
- "**Main theme** items require the student to identify the main theme or underlying message of the text or some specified part of it.
- "**Inference** items require the student to draw an inference or generalization from the text about some character or event. The information required is not explicitly stated in the text.
- "**Locate information** items require the student to search and find some specified information contained in a structured document.
- "**Follow direction** items require the student to follow the directions contained in a structured document."

The distribution of items across skills assessed by domain and grade is shown in Table 7-2. Each domain was represented by more than 20 items designed to yield reliable test scores at each grade level.

However, it was decided not to report across skill areas because there were insufficient numbers of items to report on each skill reliably. As noted in Table 7-2, students in grade 4 were asked only 4 main theme questions, 11 verbatim questions, 12 inference questions, 16 paraphrase questions, and 23 locate information questions. Students in grade 9 were asked 7 verbatim questions, 11 main theme questions, 13 follow direction questions, 18 paraphrase questions, 19 locate information questions, and 21 inference questions.

Table 7-2. Number of items by reading literacy domain, skill assessed, and grade

Domain	Skills assessed													
	Verbatim		Paraphrase		Main theme		Inference		Locate information		Follow directions		Total items	
	Grade 4	Grade 9	Grade 4	Grade 9	Grade 4	Grade 9	Grade 4	Grade 9	Grade 4	Grade 9	Grade 4	Grade 9	Grade 4	Grade 9
Narrative prose	3	1	9	9	2	4	8	15	-	-	-	-	22	29
Expository prose	8	6	7	9	2	7	4	4	-	-	-	-	21	26
Document	-	-	-	-	-	-	-	2	23	19	-	13	23	34
Total Items	11	7	16	18	4	11	12	21	23	19	-	13	66	89
Estimated Reliability	0.747	0.600	0.767	0.795	0.496	0.661	0.688	0.719	0.763	-	-	0.615	-	-

-Not available.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

7.5. Item Format

Three item formats were used--open ended, fill in the blank, and multiple choice. As noted in Table 7-3, two of the items at each grade level were open-ended questions that required students to formulate and write an answer to the question. Four of the fourth grade and 20 of the ninth grade items were fill-in-the blank items, and the remainder were multiple choice. That table also displays the distribution of item types across domains for each grade level.

Table 7-3. Number of passages and types of items by reading literacy domain and grade

Domain	Grade 4				Grade 9			
	Number of passages	Number of multiple choice items	Number of fill in the blank items	Number of open-ended items	Number of passages	Number of multiple choice items	Number of fill in the blank items	Number of open-ended items
Narrative	4	22	0	1	5	29	0	1
Expository	5	21	0	1	5	26	0	1
Document	6	19	4	0	9	14	20	0
Total	15	62	4	2	19	69	20	2

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

7.6. Scoring the Items

Scoring the items was done according to answer keys and guidelines provided by the IEA International Coordinating Center. Both multiple choice and fill-in-the-blank items were scored as either right or wrong. The IEA also provided scoring rubrics for the open-ended items. As seen in Exhibit 7-2, the rubrics, developed item-by-item, were fairly specific about acceptable responses.

Exhibit 7-2. Sample original scoring guide

A. Why did the parents ask grandpa back to the table?

9 = no response

1 = gives an unacceptable response
gives response that does not include reason for parents' change in attitude

2 = gives an acceptable response
"They realized they had been selfish."
"They were embarrassed (ashamed) after watching their son."
"They put themselves in his place and realized how hard it was for him."
"They learned from their son's activity what it could be like to be an old person."

B. What might be some way scientists could study blue whales?

9 = no response

1 = unacceptable response
"There are fish in my school."

2 = gives one way
"Follow them around."
"Capture a blue whale."

3 = gives two or more distinctive ways: give one way with some elaboration.
"Scientists could put radios on whales and follow them around."

The open-ended items were not to be included in the international scales. Therefore, within the United States, we chose to score them a second time according to a slightly more elaborate rubric that had generic characteristics that were similar from item to item. The general rubric is provided as Exhibit 7-3.

A discussion of these open-ended items and a comparison of results related to differences in the scoring rubric are presented as a separate paper in *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study.*

7.7. Summary Comments

All sampled students took all items of this test. The instrument closely resembled a standardized multiple choice reading test that has been customarily administered in the United States over the last 40 years. Consistent with this genre of standardized tests, passages varied in length, questions tapped a variety of skills, and the test itself was of a similar time duration to those commonly encountered in the U.S. Directions to students were also very similar to those that accompany these types of tests in the U.S.

Exhibit 7-3. IEA general rubric for open-ended questions

9 = No response	
0 = Off Task	Responses not related to the question.
1 = Unsatisfactory	These responses indicate miscomprehension of the question or the passage. They often contain incomplete, incorrect, or fragmented information.
2 = Partial	These responses demonstrate only some comprehension. They give information on only one part of a question or do not anchor the response in the text. When elaboration is required, these responses only give text information.
3 = Essential	These responses demonstrate adequate comprehension. Although they contain essential information, either there are a few specific references to the text or there is little elaboration.
4 = Extensive	These responses demonstrate rich comprehension. They contain complete, relevant information that is internally consistent and related. They also contain specific references to the text and, where called for, elaboration based on background knowledge.

8. THE SCALING PROCEDURES

8.1. Introduction

Given the richness of the potential data available and the large number of cognitive items included in the test (see Table 7-4), it would be impractical to report item-by-item comparisons. The use of a summary statistic to characterize reading performance in a meaningful way would make it possible to present more readily usable information. With the assistance and guidance of the IEA Technical Advisory Committee, the International Steering Committee developed procedures for creating international scaled scores based on the Rasch model, the one-parameter item response theory model.

This chapter provides a description of the procedures used by the International Steering Committee in developing the scales for the IEA Reading Literacy Study and examines issues related to how well these scales and the procedures used reflect standard practice and expectations of performance in the United States. The chapter is divided into the following sections:

- An overview of the international scaling procedures;
- Using the Rasch model with the U.S. data;
- Procedures used as they pertain to the U.S. data, including selection of the calibration sample and estimation of abilities within the U.S. data; and
- The scaled scores as a measure of U.S. students' reading proficiency.

8.2. Overview of the International Scaling Procedures

The International Coordinating Center (ICC) at the University of Hamburg performed all tasks relating to scaling of the Reading Literacy Tests (i.e., calibrated items and estimated student abilities) using the one-parameter IRT (Rasch model) procedure. Calibration of items and estimation of abilities were performed separately for each of the three reading literacy domains (narrative, expository, and document). Item difficulties were estimated on the basis of responses of a random sample of students selected from all participating countries. This international calibration sample consisted of 10,790 students for grade 4 and 10,772 students for grade 9.

The ICC deleted a total of six items for grade 4 and seven items for grade 9 that did not fit the international calibration sample, and they were deleted from consideration. Rasch analysis was performed within each participating country, setting the item difficulties derived on the international calibration sample as known parameters. Item fit was also examined within each participating country. If an item was found not to fit the Rasch model in a given country, that item was not included in estimating student abilities within the country under consideration. Based on the invariance properties of the Rasch model (i.e., examinee ability estimation is independent of the particular set of items administered from a calibrated pool), however, the ICC derived reading literacy ability estimates for students within each participating country and placed them on a common scale. For ease of use, the logit scale was transformed such that the international mean and standard deviation were 500 and 100, respectively, for each reading literacy domain.

Since the international mean and standard deviation were arbitrarily set, the scale scores across the domains are not equated. Similarly, the scale scores across the two populations are not equated either.

While the U.S. National Steering Committee has decided to report results that are consistent with the international scales, they also have examined these scales in light of standard practice within the U.S. The construction of these scores as they relate to reporting U.S. results is discussed in the following section.

8.3. Using the Rasch Model with the U.S. Data

Instead of using an Item Response Theory (IRT) model such as Rasch to create scaled scores, a number of other alternatives, including a norm-referenced metric and a percentage-correct metric, could have been chosen. However, IRT allows one to summarize easily student performance on the three reading literacy domains without sacrificing a great deal of useful information because of the following attributes:

- Item parameters (e.g., difficulty) are not dependent on the particular examinees sampled.
- Comparisons of students among countries would not necessarily require the same set of items for all countries. Increased test validity could, therefore, be achieved if certain items not working in one country were dropped from consideration when estimating student abilities within that country.
- Better estimates of test reliability could be obtained because standard error of measurement (SEM) does not need to be interpreted on the basis of parallel test forms. More importantly, one does not have to assume that the variance of error of measurement is the same for all examinees.
- How an examinee might perform when confronted with a given test item could be estimated. This information would then be useful in interpreting what students can and cannot do in relation to the test items under consideration.

While three distinct IRT alternatives for summarizing student performance are available-- one-, two-, and three-parameter IRT models -- on the basis of ease of use, desired statistical qualities, and wide acceptability in countries participating in the study, the one-parameter IRT model (Rasch model) was selected. Although this is not consistent with the scaling procedures used in the National Assessment of Education Progress (NAEP), where a three-parameter model is the method of choice, the Rasch model offers many advantages if certain underlying assumptions are met.

IRT procedures postulate that an examinee's performance on a test can be explained by a latent trait or ability; response to each item is considered to be a manifestation of that latent trait. "An item response model specifies a relationship between observable examinee test performance and the unobservable traits or abilities assumed to underlie performance on the test" (Hambleton and Swaminathan 1986, 9). This relationship is described by an item characteristic curve (ICC), which is a mathematical function relating the probability of success on an item to the ability measured by the item set or test that contains it. Different IRT models specify different mathematical relationships between the observed response (i.e., examinees' test performance) and the unobservable trait or ability.

The Rasch model specifies the following relationship:

$$P_i(\theta) = \frac{e^{D\bar{a}(\theta - b_i)}}{1 + e^{D\bar{a}(\theta - b_i)}}$$

where $P_i(\theta)$ is the probability that a randomly selected examinee with ability θ answers item i correctly, b_i is item difficulty, $\bar{a}$ is the common level of item discrimination, and D is a scaling factor.

To establish whether this model is valid with regard to the U.S. data, we have examined four interrelated underlying assumptions of the Rasch model -- unidimensionality, local independence, the shape of item characteristic curves as reflected in fit statistics, and the invariance of item and ability estimates. Violations of these assumptions would be reflected in the fit statistics provided by the statistical computer programs designed for Rasch analysis. (Note: the computer program BIGSCALE was used in this study).[1]

The assumption of **unidimensionality** implies that only one trait or ability accounts for examinees' test performance. In practice, one does not assume that other factors (e.g., motivation, anxiety, test-taking skills, traits other than the one being measured) have no impact on test performance. What is assumed is that the trait or ability under consideration is the dominant factor in explaining examinees' test performance.

Numerous methods are available to assess the validity of the assumption of unidimensionality. Several of these methods were applied to the IEA Reading Literacy data and are discussed in Chapter 10. In general, these findings indicate that overall the unidimensionality assumption is met for each of the reading literacy domains, although the degree of the departure from unidimensionality varies by domain. This degree of departure from unidimensionality, as well as departures from other assumptions of the model, is reflected in the fit statistics produced by BIGSCALE and are discussed below.

The assumption of **local independence** states that an examinee's performances on different test items are statistically independent. "For this assumption to be true, an examinee's performance on one item must not affect, either for better or for worse, his or her responses to any other items in the test" (Hambleton and Swaminathan 1986, 23). For a set of test data to meet the assumption of local independence, at least two conditions must be met:

- The test must be unidimensional. It has been shown that performance across test items at a given ability will not be statistically independent when more than one ability are being measured by the test; and
- The order of presentation of test items must not affect test performance.

---

[1] BIGSCALE is a Rasch-model rating scale analysis computer program. *A User's Guide to BIGSCALE* (1989), by B.D. Wright, J.M. Linacre, and M. Schultzz, is available from MESA Press in Chicago.

As previously mentioned, the findings from the U.S. data indicate that the unidimensionality assumption is met for each of the reading literacy domains. Assessing order effects, on the other hand, is quite difficult in a given testing situation. To a limited extent, this assumption was empirically studied in the context of the U.S. portion of the IEA Reading Literacy Study. The findings do suggest a minor order effect. This minor effect is evident only in the expository scale in grade 4. Although there might be situations where items within a passage might violate the assumption of local independence, given the small magnitude of the order effect and the magnitude of the fit statistics, within the U.S. data a serious violation of the assumption of local independence is not indicated.

The Rasch model assumes that the **item characteristic curves** take the form of parallel logistic distribution functions. Although one could plot and visually examine the curves for each item in the IEA Reading Literacy Study, there are better tests of fit to the model. These fit statistics, which take into account all departures from the assumptions of the model, are provided as part of the item calibration process. The output from the BIGSCALE program provides all the necessary fit statistics. These are included in Tables 8-1 and 8-2. The column "COUNT" indicates the number of examinees correctly responding to an item, and the column "SAMPLE" refers to the number of examinees with valid responses for an item. Thus, by dividing "COUNT" by "SAMPLE," the proportion of examinees correctly responding to the item (i.e., p-value) can be obtained.

The column "CALIBRTN" (calibration) indicates the item's difficulty parameter on the logit scale. The logit values have been suffixed by "A" to indicate that item values have been anchored at the values obtained for the international calibration sample. The column "ERROR" represents the standard error of the item difficulty parameter.

Two types of fit statistics are provided -- "INFIT" and "OUTFIT." While both are measures of model fit -- the degree to which the observed data agree with predicted values based on the model -- the infit statistic is more sensitive to unexpected responses by people whose abilities are around the item's difficulty value. In contrast, the outfit statistic is more sensitive to responses by people whose abilities are some distance (on the logit scale) from the item difficulty value. MNSQ is the mean-square INFIT (or OUTFIT) statistics, with the expected value equal to 1. Values substantially less than 1 may indicate dependency in the data, while values substantially greater than 1 may indicate random error (noise).

The column labeled "DISPLACE" represents the difference between the anchored value (based on a best fit of the data to the international calibration sample) and the item difficulty estimate resulting from a best fit of the data to the model based on the U.S. sample of students. The optimal fit for the international calibration sample may not necessarily produce item parameters that may be considered optimal for the U.S. sample. DISPLACE shows the departure from optimality for the U.S. sample relative to the international calibration sample.

Inspection of both these fit statistics in Tables 8-1 and 8-2, which range from -25.9 to 25.7, reveals that, generally, they are within acceptable ranges when one considers the attributes of the sample design as well as the U.S. sample size. For example, the Table 8-1 narrative section includes fit statistics of 20.1 and 25.1. To evaluate these statistics one must take into account the large sample sizes (in excess of 6,000 for grade 4 and 3,000 for grade 9) and the sampling design. Both may contribute to the inflation of these fit statistics. Although the design effect for the Rasch fit statistics has not been estimated, it is known that the design effect for estimating the population mean is about 6 for grade 4 and about 8 for grade 9 and that the design effects for estimating regression coefficients are typically around 2. While the design effects for estimating population mean and regression coefficients may not be used directly to estimate the design effect for Rasch fit statistics, it may, nevertheless, be supposed that the design effects for these fit statistics are somewhere between 2 and 6.

Table 8-1. Fourth grade IRT item statistics

DOMAIN	NUM	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFIT	DISPLACE*
NARRATIVE										
BIRD	1	3810	5812	.73A	.03	.9	-5.3	.9	-5.4	.27
	2	4088	5812	.02A	.03	1.2	11.7	1.5	12.3	-.14
	3	2711	5812	1.31A	.03	1.1	7.0	1.1	5.7	-.21
	4	4539	5812	-.73A	.04	1.1	3.1	1.0	-.2	-.37
	5	5399	5812	-1.96A	.05	.9	-3.0	.8	-2.8	*
NO DOG	2	3742	5812	.95A	.03	1.1	3.6	1.1	3.3	.42
	3	4645	5812	-.17A	.03	.9	-6.0	.9	-3.9	.34
	4	4367	5812	.57A	.03	.9	-7.4	.8	-8.1	.68
	5	3850	5812	1.34A	.03	1.1	6.1	1.1	6.3	.90
	6	4861	5812	-.17A	.03	.7	-15.8	.6	-13.9	.61
SHARK	1	5035	5812	-1.01A	.04	.8	-7.8	.6	-7.2	.14
	2	4263	5812	-.79A	.04	1.5	20.1**	2.0	15.5	-.86
	3	4589	5812	-.71A	.04	1.1	4.7	1.0	.5	-.27
	4	4242	5812	-.37A	.04	1.1	6.3	1.1	2.7	-.37
	5	3689	5812	-.26A	.04	1.4	18.2	1.6	12.8	-.94
GRANDPA	1	4006	5812	-.06A	.03	1.1	6.6	1.1	3.6	-.32
	3	3834	5812	.35A	.03	.9	-3.2	.9	-4.4	-.08
	4	4290	5812	.35A	.03	.9	-5.2	.9	-5.0	.41
	5	4318	5812	-.03A	.03	.9	-8.0	.7	-10.0	.08
	6	2674	5812	.64A	.03	1.4	25.1**	1.7	23.0**	-.96
EXPOSITORY										
POSTCARD	1	6076	6325	-1.92A	.05	.6	-11.6	.4	-9.9	.52
	2	6157	6325	-3.04A	.08	1.0	.4	1.2	1.1	*
WALRUS	1	5879	6325	-1.42A	.04	.7	-12.5	.6	-7.7	.41
	2	5952	6325	-1.46A	.05	.5	-17.7	.4	-12.1	.53
	3	4976	6325	-.43A	.03	.9	-6.0	.8	-6.5	*
	4	4995	6325	-.46A	.03	.9	-4.9	.8	-4.8	-.04
	5	4958	6325	-.17A	.03	.9	-6.8	.9	-4.3	.20
	6	3694	6325	.68A	.03	1.0	.0	1.0	.6	-.13
QUICKSAND	2	5256	6334	-.83A	.04	1.0	-1.6	.8	-4.9	-.10
	3	5702	6333	-1.04A	.04	.7	-11.4	.5	-11.2	.38
MARMOTS	1	3398	6334	.63A	.03	1.1	5.2	1.1	2.9	-.44
	2	2574	6334	.88A	.03	1.2	15.1	1.2	12.4	-.85
	3	2467	6334	1.40A	.03	1.0	4.3	1.1	4.5	-.41
	4	2734	6334	1.36A	.03	1.0	.6	1.1	2.9	-.23
TREES	1	4663	6334	.27A	.03	.8	-17.1	.7	-14.0	.31
	2	3079	6334	1.34A	.03	1.0	3.4	1.1	4.5	*
	3	3409	6334	1.45A	.03	1.0	-1.9	1.0	.1	.40
	4	2137	6334	1.81A	.03	1.1	4.2	1.1	5.7	-.28
	5	4526	6334	.95A	.03	1.0	-3.6	.9	-3.2	.81
DOCUMENT										
ISLAND	1	5803	6302	-.51A	.04	.6	-17.3	.6	-11.2	.66
	2	6038	6302	-1.64A	.05	.7	-7.7	.7	-4.4	.38
	4	5112	6302	-.05A	.03	1.0	-.2	1.0	1.1	.16
MARIA	1	5044	6302	-1.15A	.05	2.1	24.7**	2.7	18.7	-1.41
	2	4047	6302	.04A	.03	1.4	22.6**	1.7	17.7	-.92
	3	5266	6302	-.75A	.04	1.2	6.4	1.1	2.7	-.37
BOTTLE	1	6079	6316	-.97A	.04	.4	-24.1	.3	-16.3	.83
	2	4705	6315	.73A	.03	.8	-16.4	.7	-13.3	.45
	3	6095	6316	-1.55A	.05	.5	-14.7	.4	-10.1	.57
	4	5880	6316	-.36A	.04	.5	-25.9**	.4	-18.5	.84
BUSES	1	5737	6299	-1.10A	.05	.8	-5.9	.6	-6.9	.09
	2	4398	6262	.54A	.03	.9	-9.8	.8	-8.4	*
	3	1712	6305	2.65A	.03	.9	-6.7	.8	-6.8	-.14
	4	2832	6295	.96A	.03	1.2	14.9	1.3	13.1	-.90
CONTENTS	1	5882	6316	-1.42A	.05	.9	-2.3	1.0	-.4	.10
	3	5939	6316	-.76A	.04	.5	-21.7**	.4	-15.8	.69
TEMPERATURE	1	4868	6316	-.40A	.04	1.3	12.2	1.4	7.9	-.54
	2	2161	6316	2.42A	.03	1.0	1.0	1.2	6.6	.04
	3	3004	6316	1.64A	.03	1.0	-.6	1.0	2.4	-.05
	4	3008	6316	1.50A	.03	1.0	2.2	1.1	4.2	-.19
	5	5317	6316	.20A	.03	.7	-19.5**	.6	-15.0	.58

*Values less than 0.04.

**High fit statistics (see pages 112 and 115).

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 8-2. Ninth grade IRT item statistics

DOMAIN	NUM	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFIT	DISPLACE*
NARRATIVE										
FOX	2	2915	3162	-1.59A	.07	.9	-2.4	.8	-2.0	.12
	3	2429	3162	.56A	.04	1.0	-.5	.9	-1.5	.66
	4	2923	3162	-.13A	.05	.8	-11.1	.6	-9.0	1.16
	5	2845	3162	-1.28A	.06	1.1	2.7	2.2	8.9	.11
MUTE	1	1679	3162	1.22A	.04	1.1	4.6	1.1	3.5	-.07
	2	1811	3162	1.03A	.04	1.0	-.4	1.0	-.4	*
	3	2180	3162	-.11A	.05	1.1	4.6	1.2	3.0	-.51
	4	1275	3162	1.88A	.04	1.3	13.2	1.5	12.8	-.12
	5	1948	3162	.59A	.04	1.1	2.5	1.1	3.0	-.22
SHARK	2	2971	3173	-2.03A	.08	1.0	.7	.9	-.4	*
	3	2952	3173	-1.66A	.07	.8	-4.0	.6	-4.2	.19
	4	2990	3173	-1.64A	.07	.7	-7.2	.6	-3.6	.38
	5	2807	3173	-1.38A	.06	1.1	2.3	1.2	2.2	-.18
REVENGE	1	2255	3173	-.64A	.05	1.4	11.1	1.3	3.8	-1.04
	2	1691	3173	.88A	.04	1.2	9.9	1.3	8.6	-.41
	3	2168	3173	-.03A	.05	1.1	4.6	1.1	2.2	-.47
	4	1972	3173	.98A	.04	1.0	-2.4	.9	-1.9	.19
	5	2080	3173	.31A	.04	1.1	3.8	1.1	2.9	-.28
	6	2184	3173	.35A	.04	1.1	2.6	1.0	1.0	*
	7	2505	3173	-.49A	.05	1.0	-1.5	.9	-1.5	-.18
ANGELS	1	1774	3173	1.02A	.04	1.0	-2.5	.9	-2.8	-.12
	2	2194	3173	.35A	.04	.9	-6.9	.7	-7.6	*
	3	2193	3173	.49A	.04	.9	-6.4	.8	-6.9	.13
	5	2244	3173	.54A	.04	.9	-7.2	.7	-7.6	.27
	6	2317	3173	.39A	.04	.8	-9.5	.7	-8.9	.27
	7	2317	3173	.40A	.04	.9	-6.0	.8	-5.9	.28
EXPOSITORY										
MARMOTS	1	2656	3177	-.60A	.05	.9	-4.4	.9	-2.4	.26
	2	2488	3177	-1.05A	.05	1.4	10.4	1.6	6.6	-.69
	3	2116	3177	-.18A	.05	1.4	15.0	1.8	14.7	-.50
	4	2593	3177	-.41A	.05	.9	-5.2	.8	-3.1	.28
LASER	1	2920	3177	-1.55A	.06	.8	-4.7	.8	-2.3	.23
	2	2557	3177	-.17A	.05	.8	-11.1	.6	-9.1	.41
	3	2507	3177	-.95A	.05	1.3	7.8	1.3	3.7	-.50
	4	2010	3177	1.01A	.04	1.0	.3	1.0	.7	.52
	5	1986	3177	.64A	.04	1.0	1.4	1.0	.8	.12
	6	2547	3177	-.97A	.05	1.1	2.3	.8	-2.7	-.41
	7	1796	3177	.79A	.04	1.0	1.1	1.1	1.8	-.05
	9	2202	3177	.39A	.04	1.0	-2.3	.9	-1.7	.25
	10	2263	3177	-.19A	.05	1.0	-2.2	.9	-3.0	-.21
PARACUTIN	1	2193	3192	.22A	.04	1.1	3.0	1.1	2.6	.05
	2	2653	3192	-1.03A	.05	1.2	5.6	1.6	6.8	-.22
	3	2861	3192	-1.04A	.05	.8	-6.8	.7	-5.0	.39
	5	2310	3192	.53A	.04	1.1	2.9	1.1	3.5	.55
	6	2297	3192	.03A	.04	1.0	-.1	1.0	-.3	.05
SMOKE	1	1772	3192	.95A	.04	1.0	1.5	1.0	1.5	.05
	2	1482	3192	1.33A	.04	1.0	-1.2	1.0	.5	-.05
	3	1524	3192	.97A	.04	.9	-3.3	.9	-2.8	-.34
	4	2109	3192	-.10A	.04	1.2	7.0	1.2	4.2	-.45
	5	2345	3192	-.06A	.04	.9	-7.3	.8	-6.3	.06
	6	1341	3192	1.43A	.04	1.0	-.8	1.0	.9	-.19
DOCUMENT										
CARD	1	3254	3308	-.71A	.06	.3	-21.8**	.2	-15.7**	1.08
	3	3254	3308	-1.81A	.09	.4	-10.0	.6	-3.5	.71
	4	3148	3308	-1.72A	.08	1.1	.9	1.6	3.8	*
	5	2662	3307	-.18A	.05	1.2	5.9	1.4	6.1	-.25
	6	2010	3310	.71A	.04	1.3	14.0	1.5	12.5	-.55
	7	3137	3309	-1.49A	.08	1.0	-.7	1.2	1.4	.13
RESOURCES	1	2480	3301	-.81A	.06	2.3	22.6**	2.9	16.5**	-1.79
	2	2542	3306	1.19A	.04	.8	-11.2	.8	-9.6	.79
	3	1854	3305	2.23A	.04	1.1	6.7	1.2	5.8	.79
JOB VACANCIES	1	2602	3310	.07A	.05	1.1	2.6	1.0	.7	-.12
	2	2856	3310	-.36A	.05	1.0	-1.0	.9	-1.4	.08
LYNX	1	2714	3310	-.23A	.05	1.2	6.3	1.5	7.5	-.17
	2	1568	3310	1.93A	.04	1.0	1.0	1.1	2.9	.04
	3	1878	3310	1.55A	.04	1.1	4.3	1.1	4.5	.12
BUS SCHEDULE	1	1961	3322	.14A	.05	1.8	25.7**	2.2	19.7**	-1.44
	2	2329	3322	.66A	.04	1.0	.0	1.0	-.5	-.06
	3	1838	3321	1.25A	.04	1.0	2.7	1.1	2.7	-.25
DIRECTIONS	1	2744	3320	-.32A	.05	1.1	2.1	.9	-1.5	-.21
	2	2742	3321	.06A	.05	.8	-7.6	.6	-8.4	.17
	3	2366	3296	.66A	.04	.9	-3.2	.9	-2.8	*
WEATHER	1	2948	3321	-1.35A	.07	1.6	8.4	1.3	2.8	-.82
	2	2669	3321	.36A	.04	.9	-2.5	1.0	-.3	.29
	3	2851	3320	-.80A	.06	1.2	5.0	1.1	1.1	-.45
	4	1719	3312	1.19A	.04	1.1	6.7	1.2	5.8	-.49
TEMPERATURE	1	3016	3322	-.95A	.06	1.0	.5	1.0	.4	*
	2	2685	3322	.53A	.04	.8	-9.7	.8	-6.9	.46
	3	2565	3322	.46A	.04	1.0	-1.6	1.0	-.1	.18
	4	2722	3322	-.03A	.05	1.0	.0	1.0	.2	*
	5	3146	3322	-1.34A	.07	.8	-4.1	.8	-2.5	.25
ASPIROL	1	3179	3322	-1.94A	.09	1.0	-.4	.6	-3.2	*
	2	2988	3322	-.76A	.06	.9	-2.7	.8	-3.5	.09
	3	1895	3322	1.82A	.04	.9	-4.3	.9	-3.9	.41

*Values less than 0.04.

**High fit statistics (see pages 112 and 115).

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Therefore, corrections that account for these two attributes would yield adjusted fit statistics of 2.12 and 2.65.[2] These would not typically be considered to be of an acceptable magnitude. Further examination of Table 8-1 for grade 4, after taking into consideration the sampling attributes, reveals that two items on the narrative scale and six items on the document scale would have fit statistics that are considered high (i.e., adjusted fit greater than 2.0). For grade 9 (Table 8-2), three items on the document scale are considered high. Overall, however, it is reasonable to conclude that the data seem to adequately fit the one-parameter Rasch model.

Additional evidence of whether item characteristic curves are parallel may be found through an examination of the index of item discrimination (i.e., point-biserial correlation between the item score and domain score). The item discrimination index for grades 4 and 9 test items are displayed in Tables 8-3 and 8-4.

With a few exceptions, these tables indicate that the point-biserial correlations tend to be high. For grade 4, typical values are 0.53, 0.45, and 0.42 for narrative, expository, and document items, respectively. For grade 9, the corresponding figures are 0.53, 0.48, and 0.49. To estimate the degree of heterogeneity among these correlations, the quartile deviation was computed.[3] The quartile deviations for grade 4 are 0.026, 0.025, and 0.059 for narrative, expository, and document items, respectively. For grade 9, the corresponding figures are 0.085, 0.044, and 0.038. Because these quartile deviations are not particularly high, it is reasonable to conclude that the items are homogeneous in terms of their discrimination and thus that the assumption of equal item discrimination is tenable.

The assumption of **invariant item and ability estimates** states that estimates of item difficulties for a given set of items obtained from two different samples of examinees would be statistically equivalent, as would estimates of abilities for a given set of examinees obtained from two different sets of items administrated from a calibrated pool. This invariance aspect of the Rasch model was found to be quite useful in the context of the IEA Reading Literacy Study. Item-by-country interactions were checked on the basis of this property. If an item was found to behave significantly differently in one or more countries, the item was flagged. After a careful review of item statistics across participating countries, a total of six items from grade 4 and seven items from grade 9 were dropped. Table 8-5 presents the initial and final number of items for each domain and for each grade.

---

[2] To account for the above mentioned sampling attributes, we wished to transform the fit statistics so that the actual sample size would function as if it were a simple random sample of 400 students. To do this, we assumed that the fit statistic is inversely proportional to the square root of the sample size and is directly proportional to the square root of the design effect, which for this example was assumed to be 6 and 8 for grades 4 and 9, respectively.

For example, to transform the observed fit statistic of 25.1, we performed the following calculations:

$$\begin{aligned} \text{Adjusted Fit} &= \text{Observed Fit} \times \ (\sqrt{400}/\sqrt{6000 \times 6} \\ &= 25.1 \times 0.1054 \\ &= 2.65 \end{aligned}$$

[3] The Quartile Deviation (QD) was calculated as follows:

$$QD = \frac{Q3-Q1}{2}$$

where Q3 represents the 75th percentile, and Q1 represents the 25th.

Table 8-3. Item correlations with total score: Grade 4

Item	Point-biserial correlation	Item	Point-biserial correlation
Narrative		Document	
Bird1	0.579	Island1	0.311
Bird2	0.424	Island2	0.246
Bird3	0.514	Island4	0.325
Bird4	0.569	Maria1	0.335
Bird5	0.428	Maria2	0.428
Dog2	0.505	Maria3	0.421
Dog3	0.501	Bottle1	0.318
Dog4	0.543	Bottle2	0.515
Dog5	0.537	Bottle3	0.325
Dog6	0.546	Bottle4	0.376
Shark1	0.522	Bus1	0.443
Shark2	0.464	Bus2	0.549
Shark3	0.514	Bus3	0.464
Shark4	0.539	Bus4	0.472
Shark5	0.564	Content1	0.302
Grandpa1	0.529	Content3	0.379
Grandpa3	0.585	Temp1	0.415
Grandpa4	0.529	Temp2	0.414
Grandpa5	0.599	Temp3	0.439
Grandpa6	0.456	Temp4	0.430
		Temp5	0.465
Expository			
Card1	0.281		
Card2	0.187		
Walrus1	0.368		
Walrus2	0.400		
Walrus3	0.524		
Walrus4	0.508		
Walrus5	0.460		
Walrus6	0.494		
Sand2	0.489		
Sand3	0.408		
Marmot1	0.493		
Marmot2	0.427		
Marmot3	0.420		
Marmot4	0.454		
Trees1	0.570		
Trees2	0.447		
Trees3	0.516		
Trees4	0.387		
Trees5	0.476		

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 8-4. Item correlations with total score: Grade 9

Item	Point-biserial correlation
Narrative	
Fox2	0.364
Fox3	0.431
Fox4	0.308
Fox5	0.190
Mute1	0.496
Mute2	0.550
Mute3	0.574
Mute4	0.370
Mute5	0.537
Shark2	0.336
Shark3	0.401
Shark4	0.372
Shark5	0.412
Reveng1	0.624
Reveng2	0.461
Reveng3	0.561
Reveng4	0.557
Reveng5	0.525
Reveng6	0.497
Reveng7	0.560
Angel1	0.572
Angel2	0.717
Angel3	0.592
Angel5	0.583
Angel6	0.602
Angel7	0.554
Expository	
Marmot1	0.409
Marmot2	0.434
Marmot3	0.343
Marmot4	0.429
Laser1	0.350
Laser2	0.522
Laser3	0.441
Laser4	0.490
Laser5	0.474
Laser6	0.554
Liter1	0.504
Liter3	0.479
Liter4	0.563
Parac1	0.436
Parac2	0.332
Parac3	0.351
Parac5	0.382
Parac6	0.459
Smoke1	0.488
Smoke2	0.570
Expository (continued)	
Smoke3	0.563
Smoke4	0.488
Smoke5	0.565
Smoke6	0.503
Document	
Card1	0.215
Card3	0.165
Card4	0.167
Card5	0.323
Card6	0.330
Card7	0.176
Resourc1	0.343
Resourc2	0.483
Resourc3	0.464
Job1	0.377
Job2	0.344
Lynx1	0.254
Lynx2	0.401
Lynx3	0.363
Bus1	0.359
Bus2	0.428
Bus3	0.432
Direct1	0.438
Direct2	0.505
Direct3	0.465
Weather1	0.407
Weather2	0.329
Weather3	0.436
Weather4	0.412
Temp1	0.281
Temp2	0.422
Temp3	0.366
Temp4	0.357
Temp5	0.290
Aspirol1	0.384
Aspirol2	0.396
Aspirol3	0.502

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 8-5. Initial and final number of items relating to each reading literacy domain for each grade

Domain	Grade 4			Grade 9		
	Initial	Dropped	Final	Initial	Dropped	Final
Narrative	22	2	20	29	3	26
Expository	21	2	19	26	2	24
Document	23	2	21	34	2	32
Total	66	6	60	89	7	82

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

To determine whether item difficulties obtained from the International Calibration Sample and the U.S. sampled students are equivalent, we examined "DISPLACE" values (Tables 8-1 and 8-2)[4]. The findings indicate that while for most items the values are small, indicating equivalence, some items have extremely high values, implying that the departure from the optimum for the U.S. sample relative to the international calibration sample is high for these items. To illustrate, using a critical value of 1.0 (i.e., the difference in optimal solution between the U.S. and the international calibration sample is one logit or more), in grade 4 two items and in grade 9 four items were flagged. Although this is very few items for the U.S., we are not aware of how many items would be similarly flagged in other countries. Therefore, it would be appropriate to examine the international data sets to determine why some items have high "displace" values for certain countries. This information would be extremely useful in conducting future international comparative studies.

8.4. Procedures Used as They Pertain to the U.S. Data

8.4.1. Selection of the Calibration Sample

The International Steering Committee, with the advice of the IEA Technical Advisory Committee, chose to calibrate the items by selecting an international sample of students (herein referred to as the "calibration sample") and performing the item calibration within this sample. The calibration sample consisted of 10,790 students for grade 4 and 10,772 students for grade 9. Thus, for each population, about 360 randomly selected examinees from each participating country were included in the calibration sample, giving each country an equal weight in the item calibration phase.

Two alternative approaches for calibration might have been considered. One approach would have been to combine all students from all nations participating in the study and perform the scaling on this "super population" of students. However, the enormity of the sample size and the difficulty in studying country-by-item interactions would make this approach infeasible. Alternatively, one country might have been designated as a reference country, and item parameters obtained in this country could

---

[4] The column labeled "DISPLACE" represents the difference between the anchored value (based on a best fit of the data to the international calibration sample) and the item difficulty estimate resulting from a best fit of the data based on the U.S. sample of students. The optimal fit for the international calibration sample may not necessarily produce item parameters that may be considered optimal for the U.S. sample. "DISPLACE" shows the departure from the optimum for the U.S. sample relative to the international calibration sample.

have been anchored to all other participating countries. However, the arbitrary nature of selecting the reference country would make this approach infeasible as well.[5]

Therefore, the approach selected by the International Steering Committee was the best pragmatic alternative. However, the decision to represent each country in the calibration sample equally, in spite of the vast differences in the size of the eligible populations in each country, could be considered as a separate issue.

8.4.2. Estimation of Abilities Within the U.S. Data

Using the BIGSCALE computer program, item calibration was performed by the ICC on the international calibration sample separately for each of the three domains. The calibration phase involved three stages. In the first stage, all items within a domain were subjected to Rasch analysis. In the second stage, fit statistics were examined for each item to identify items not fitting the Rasch model. Six items for grade 4 and seven items for grade 9 were found not to fit the Rasch model. These items were deleted from the analysis. In the third stage, items not deleted in stage 2 were subjected to Rasch analysis again. The item statistics (i.e., Rasch difficulty values) obtained in stage 3 were defined as "known" item parameters for subsequent analyses.

Using the BIGSCALE computer program and specifying the item difficulties obtained previously (as known item parameters), estimation of examinees' abilities were performed by the International Coordinating Center within each participating country including the U.S. However, prior to estimating abilities, item fit statistics were examined for each country. If for a given country an item was found not to fit the Rasch model, that item was dropped for the country under consideration. For the U.S. data, however, no items were deleted at this stage of the analysis. Thus, for each country, estimation of examinees' abilities were based on items not deleted (i.e., items found to fit the Rasch model for the country under consideration). On the basis of the invariance properties of the Rasch model, using different subsets of items to estimate examinees' abilities did not pose a problem. In fact, this was one of the reasons why IRT procedures were applied in the context of the IEA Reading Literacy Study.

While this approach produced a satisfactory international model, it does not necessarily follow that at the national level the same would hold equally as well for each country. For example, within the U.S., given the large "DISPLACE" values for some of the items (as mentioned in Section 8.3.), it appears that some items do function differently in the U.S. as compared with the international calibration sample. Consequently, this test may not have produced item parameters that can be considered optimal for the U.S. sample. This raises the question as to whether the scale is meaningful for the U.S.

8.5. The Scaled Scores as a Measure of U.S. Students' Reading Proficiency

The IEA International Steering Committee arbitarily set the scale such that the international mean was equal to 500 and the international standard deviation was equal to 100. Because these scores are on an interval scale, it is possible to perform common mathematical manipulations (estimate means, variances, correlation and regression coefficients) that are necessary for the analyses. However, because the scales for each domain were constructed separately, it is impossible to make comparisons across the

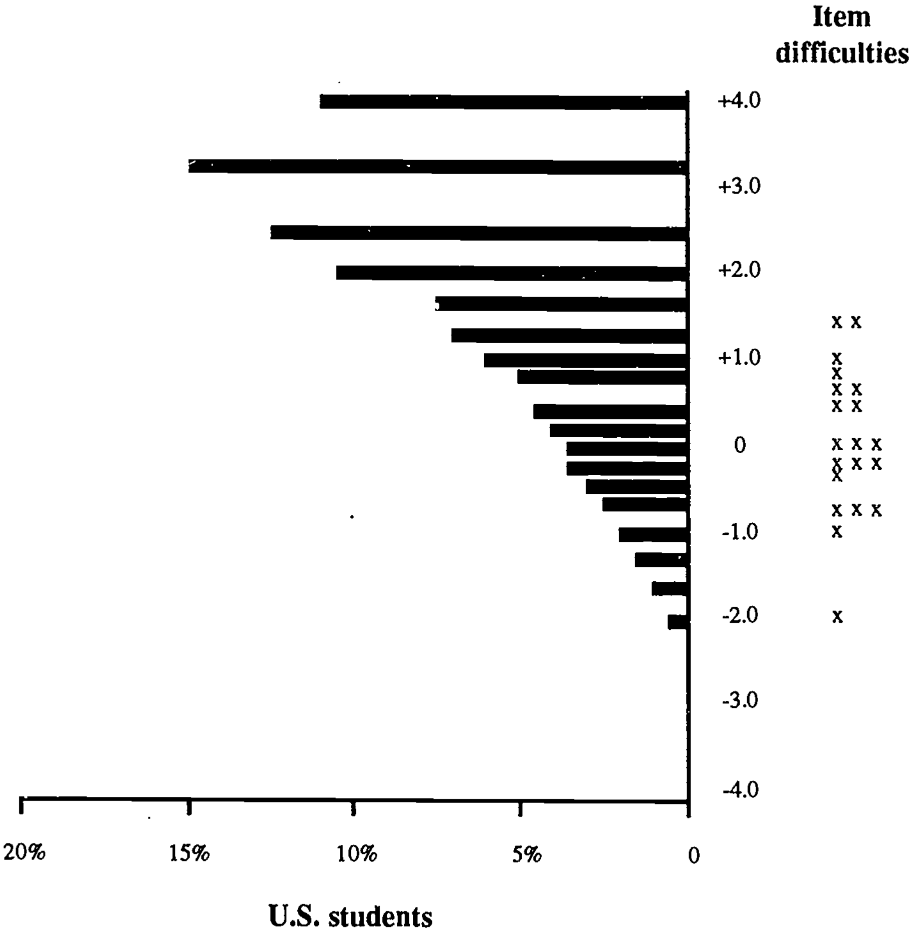
---

[5] For example, if Country A were to be selected as the reference and an item was found not to fit the Rasch model in that country, the most prudent decision would be to drop that item from calibration. However, the same item might fit the Rasch model in other countries. Thus, the choice of the reference country may have affected item parameters, and, consequently, ability estimates.

scales, in spite of the fact that they have the same arbitrarily selected mean and standard deviation. Similarly, one cannot compare scores across the two populations.

In contrast, because the scaled scores are based on the Rasch logit values that are on the same scale as the item difficulties, it is possible to describe what students can and cannot do in relation to the test items under consideration. For example, Figure 8-1 presents an item difficulty distribution and examinee ability distribution for grade 4 narrative test items. As one can observe, the ability of the U.S. fourth grade students to read narrative texts is skewed to the left (negatively skewed) and has no tail at the upper end of the scale, indicating a high degree of ceiling effect. That is, many U.S. students got correct answers to all, or almost all, items, making it difficult to obtain a reliable assessment of their position on the scale. Since the ability and item difficulty are on the same scale, it is apparent that the overlap between the two distributions is not great -- more than half of the students have abilities greater than the most difficult item.

Figure 8-1. Grade 4 narrative domain "BIGSCALE" Rasch analysis

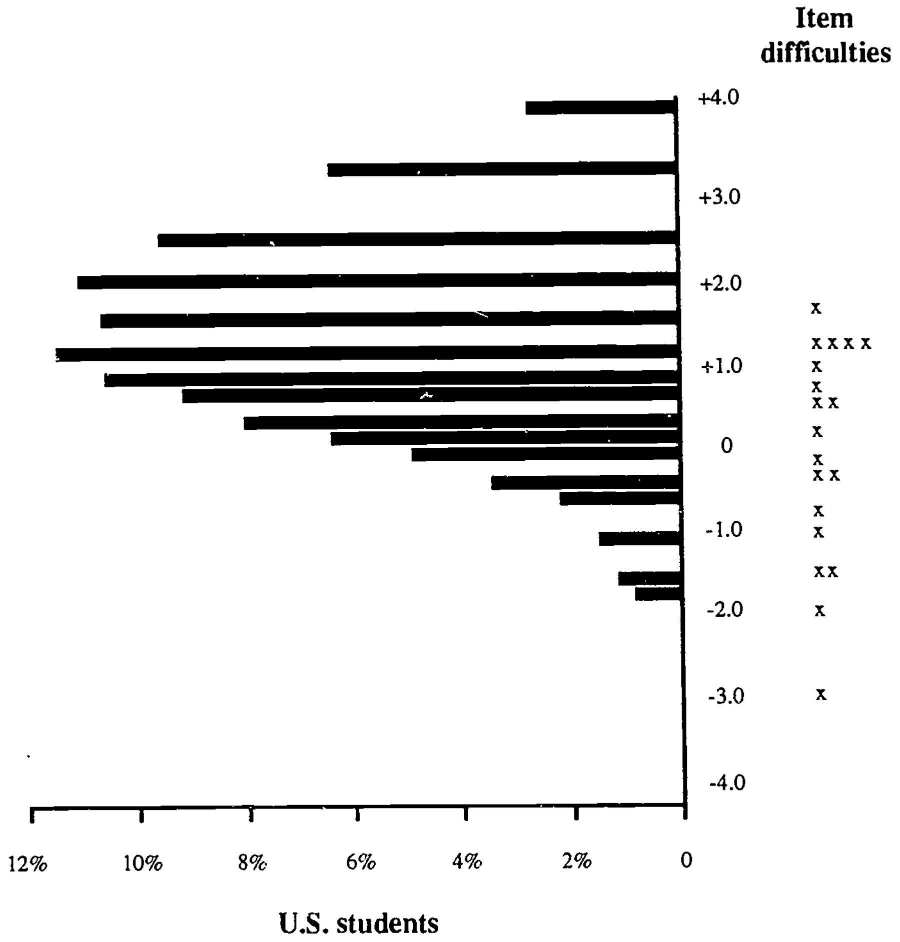


NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The ability distribution of the U.S. fourth grade students to read expository texts appears to be more symmetrical than for narrative texts (Figure 8-2). In comparison to the narrative distribution, ceiling effects do not seem to be as marked. However, similar to the narrative texts, a large proportion of U.S. fourth grade students have abilities greater than the most difficult item on the test.

Figure 8-2. Grade 4 expository domain "BIGSCALE" Rasch analysis

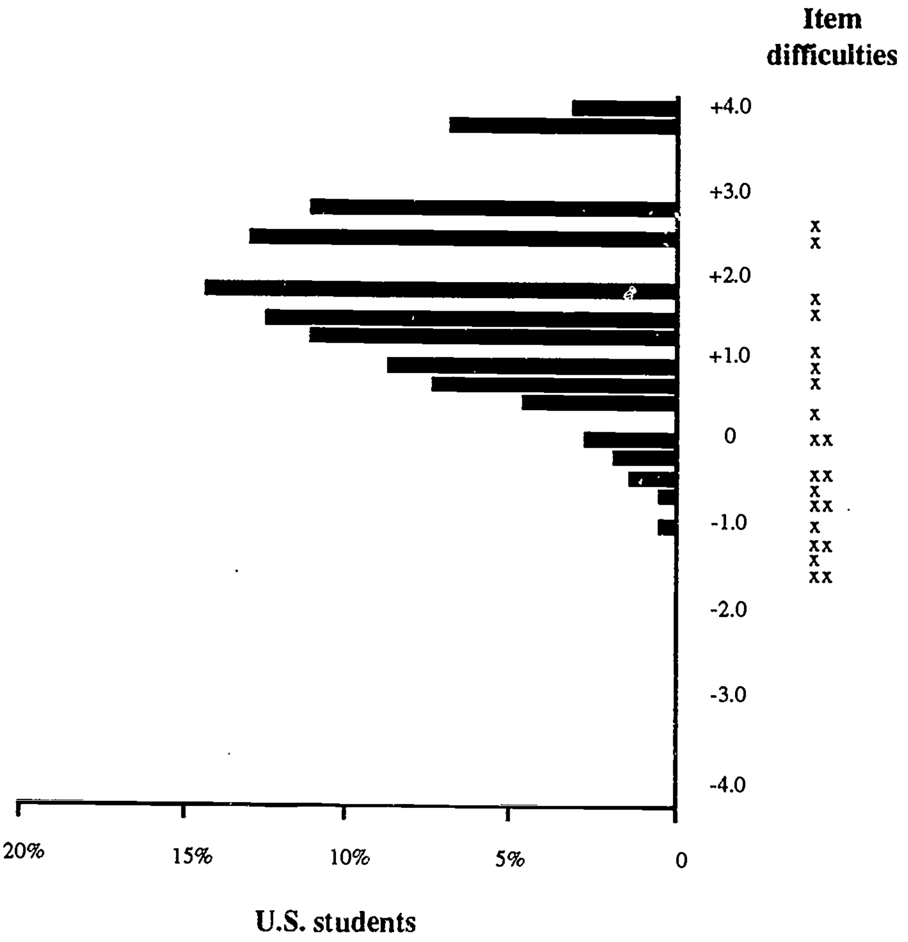


NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The ability distribution of the U.S. fourth grade students to read documents is more symmetrical as compared to the expository texts (Figure 8-3). As compared to the narrative texts there is a less marked ceiling effect. Additionally, the overlap in item and ability distributions is greater than those for the other two domains.

Figure 8-3. Grade 4 document domain "BIGSCALE" Rasch analysis



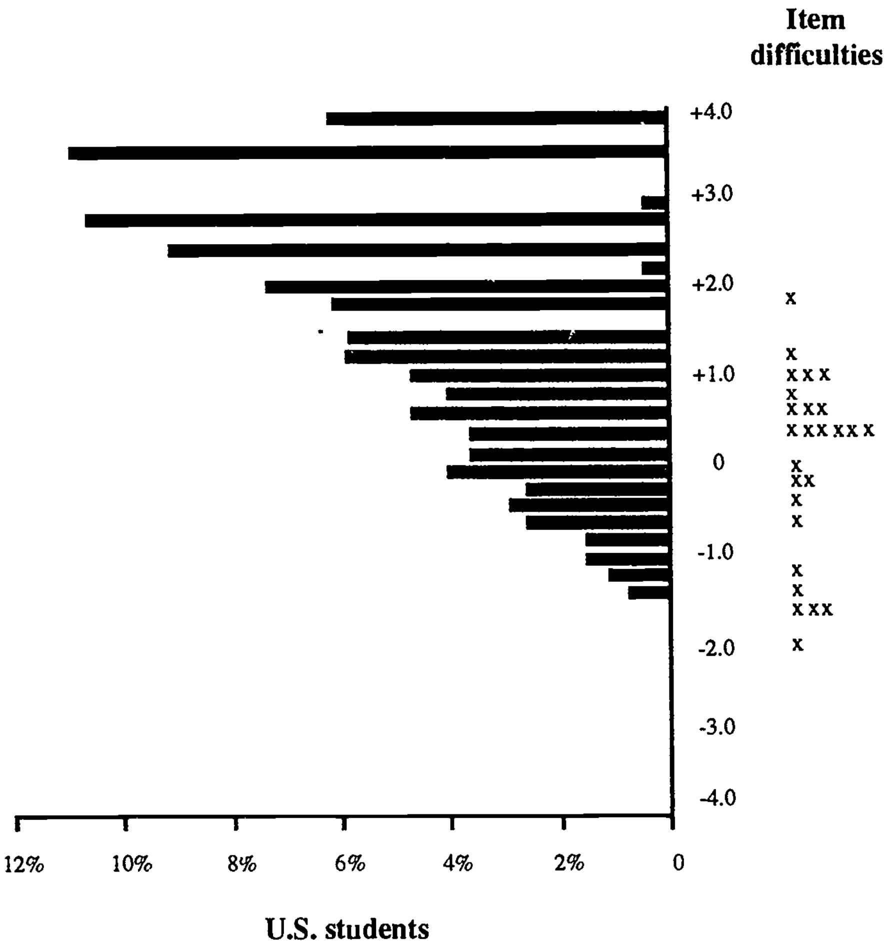
NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The distribution of abilities and item difficulties for the U.S. ninth grade students (Figures 8-4 to 8-6) are similar to the fourth grade distributions described above. Thus, generally the Reading Literacy Test items were easy for the U.S. students.

The Rasch model can also be used to predict the probability of an examinee with a given ability level correctly responding to an item, information that can be used to describe what students can and cannot do. For the purpose of illustration, consider the performance of the following four fourth grade students: (1) a fourth grade student whose ability places him/her at about the 75th percentile (ability = 3 logits) within the U.S.; (2) a median student (ability = 2.2 logits) within the U.S.; (3) a fourth grade student whose ability places him/her at about the 25th percentile (ability = 1.0 logits) within the U.S.; and (4) a fourth grade student whose ability places him/her at about the 10th percentile (ability = -1.0 logits) within the U.S. These performance ratings are all relative to the following three items: (1) the most difficult item (difficulty = 1.93 logits); a typical item (difficulty = 0.0 logit); and (3) an easy item (difficulty = -2.0 logits). Table 8-6 displays estimates of probabilities for each item-examinee combination.

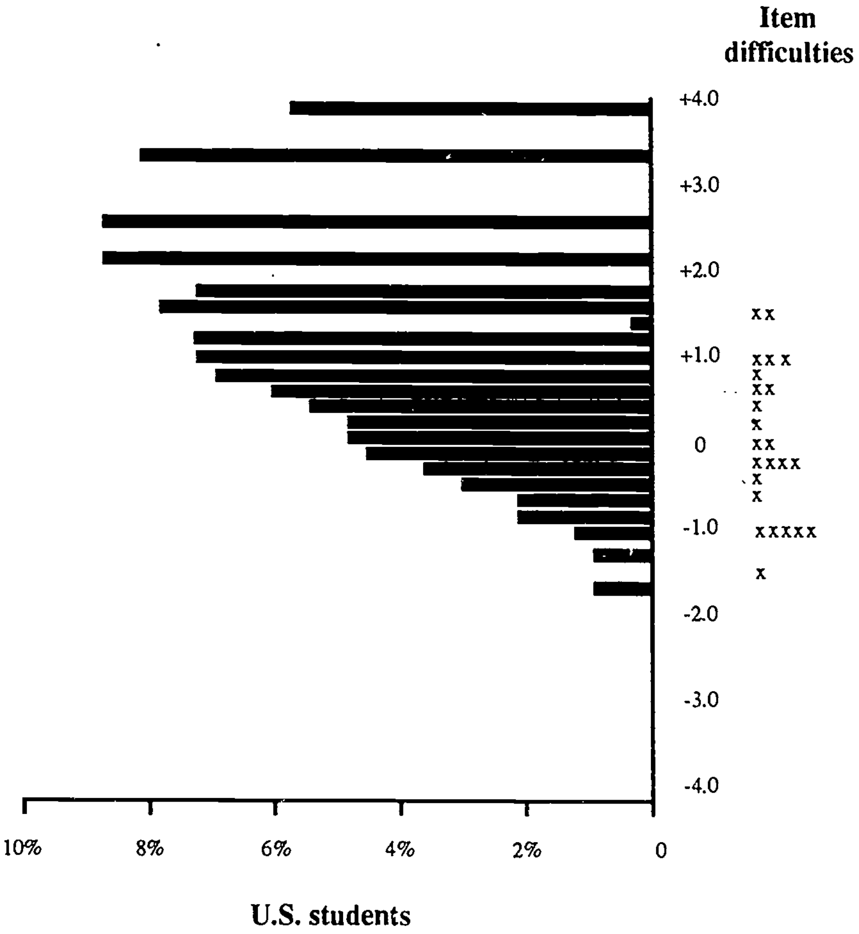
Figure 8-4. Grade 9 narrative domain "BIGSCALE" Rasch analysis



NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

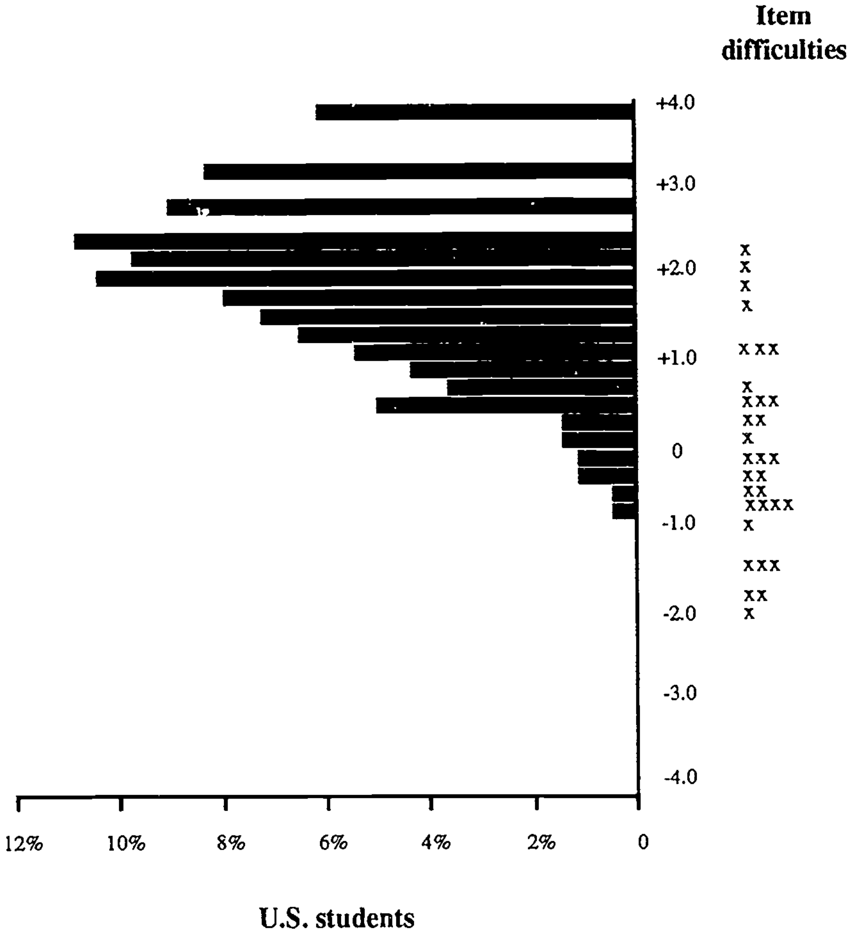
Figure 8-5. Grade 9 expository domain "BIGSCALE" Rasch analysis



NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 8-6. Grade 9 document domain "BIGSCALE" Rasch analysis



NOTE: Measure is the logit score for item difficulty and student ability.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 8-6. Predicted probability of correctly responding to three items with varying difficulties by examinees with different ability levels: Grade 4

Student ability within the U.S.	Item difficulty		
	Most difficult item	Typical item	Easy item
75th percentile	0.67	0.90	0.98
50th percentile	0.56	0.77	0.96
25th percentile	0.23	0.55	0.89
10th percentile	0.07	0.16	0.56

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The probabilities provided above indicate that (1) the high ability students (students at or above the 75th percentile) have a high probability of responding correctly to even the most difficult item on the test; (2) average students within the U.S. have a 56 and 77 percent chance of correctly responding to the most difficult and typical items, respectively; (3) low ability students (students at about the 25th percentile) have a low probability of responding to the most difficult items, but these students have a 55 and 89 percent chance of correctly responding to typical and easy items, respectively; (4) very low ability students (students at the 10th percentile) have difficulties in responding correctly to the most difficult and typical items, although even these students have a 56 percent chance of correctly responding to the easy items. Based on these observations, one can conclude that the test items are easy for the U.S. students, and the easy items are not contributing significantly in measuring the U.S. students' abilities.

8.6. Summary Comments

Throughout this chapter, we have examined the scaling procedures used internationally as they relate to the U.S. data. While the procedures themselves were straightforward and represent perhaps the best international compromise, we can conclude on the basis of five different empirical tests that these reading tests are not necessarily optimal for describing U.S. students' reading proficiency. While we would conclude that the test items on each scale are fairly unidimensional, and that there is a reasonably high degree of local independence, the item characteristic curves as reflected in fit statistics and the invariant item and ability estimates indicate some items are behaving differently in the U.S. than in the international calibration sample. Similarly, we noted ceiling effects on each of the scales and a very high probability that students at the 25th percentile in the U.S. could easily pass easy items (0.89 at grade 4), were likely to pass typical items (0.55 at grade 4), and had a slightly less than one in four chance of passing the most difficult item. Given these characteristics of the scales, we turn to examining the estimates of the reading proficiency of U.S. students.

9. ESTIMATES OF THE READING PROFICIENCY OF U.S. STUDENTS

9.1. Introduction

The objectives of this chapter are three-fold:

- To provide estimates of the reading proficiency of fourth and ninth grade students in the U.S.;
- To compare estimates of the mean reading proficiency of subpopulations defined by the student's gender, race/ethnicity, region of the country, and size of community; and
- To compare the distribution of the reading proficiencies of U.S. students to the international means and standard deviations.

These estimates and comparisons will be presented for each of the three reading literacy domains (narrative, expository, and document) using the IRT scaled scores.

To provide the reader with a broad picture of the reading proficiencies of U.S. students as measured by this assessment, we have selected a few demographic variables that are commonly used in other studies (e.g., the National Assessment of Educational Progress (NAEP)) to report achievement results. This presentation will make it possible to contrast these results with what might have been expected given results from other assessments, specifically the NAEP reading assessments.

The U.S. national portion of IEA Reading Literacy Study included two samples from which population estimates can be derived. Because the probability of selection was not constant across students, sampling weights were developed to derive the appropriate population values. All estimates reported here were derived by employing these sampling weights. Furthermore, because the sampling design employed in the U.S. was a multistage, highly stratified, clustered sampling design, a jackknife procedure was required to derive the standard errors associated with the population estimates.

This chapter is organized in two primary sections. The first presents the distribution of the domain-level proficiencies for each population. The second describes the reading performance of the U.S. students by the selected demographic variables of gender, race/ethnicity, region, and community size. For each of these sets of demographic subgroups, the mean performances of subgroups are compared with each other, the international mean, and mean values from the 1992 NAEP. Also presented are comparisons of the distributions of scores by demographic subgroups to further enhance understanding of student performance.

9.2. Domain-Level Reading Proficiencies of U.S. Students

9.2.1. Means, Standard Deviations, and Shape of Score Distributions for Each Population

The average reading proficiencies of fourth and ninth grade students in the U.S. for each of the three reading literacy domains are presented in Table 9-1. Because the reading proficiency results are based on samples of schools and students, they are subject to sampling error. The standard errors, which indicate the amount of sampling error associated with the estimate of the population means, can be used to construct confidence intervals. For example, the 95 percent simple confidence interval for fourth grade narrative ranges from 549.3 to 561.1 (that is, the mean plus or minus twice the standard error). If the study were replicated and similar confidence intervals constructed, it can be said that approximately 95 percent of such confidence intervals would capture the true population mean.

Table 9-1. Mean reading proficiency of U.S. students in grades 4 and 9, by reading literacy domains

Domain and grade	Sample size	Mean	Standard error*	Standard deviation	Effect size
Grade 4					
Narrative	6,248	555.2	2.94	95.7	+0.55
Expository	6,248	539.4	2.85	70.9	+0.39
Document	6,248	550.9	2.57	81.0	+0.51
Grade 9					
Narrative	3,209	541.9	4.98	97.5	+0.42
Expository	3,209	543.5	5.71	105.8	+0.44
Document	3,209	530.4	3.97	82.0	+0.30

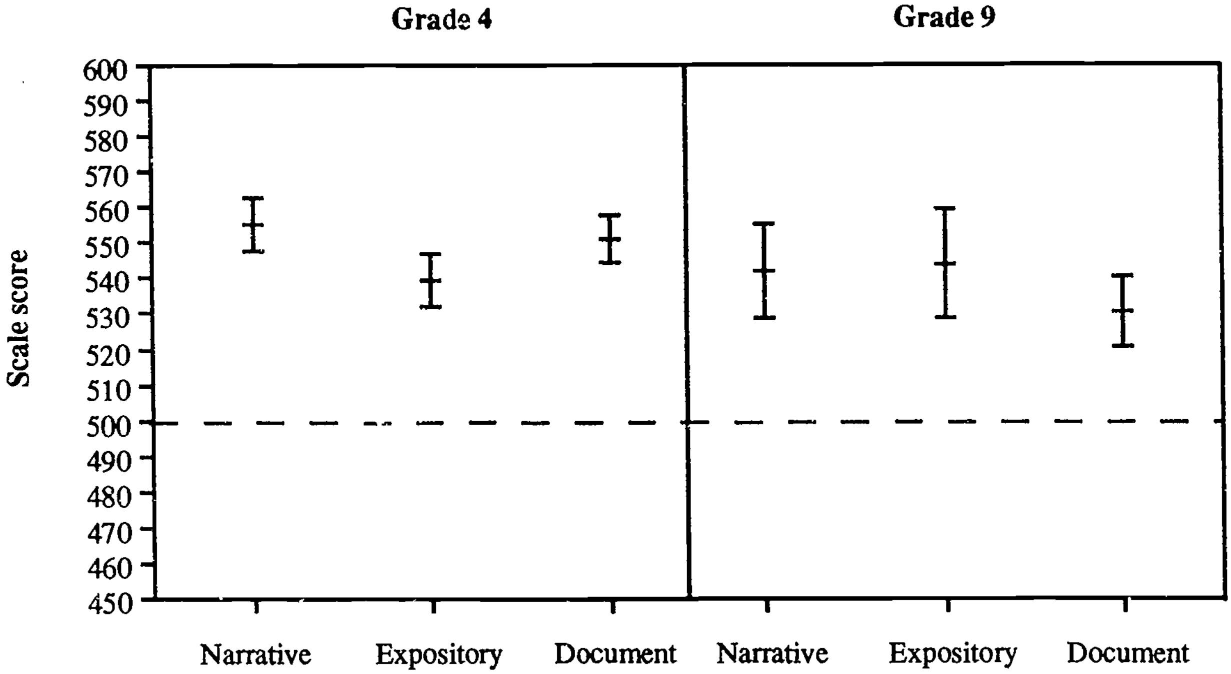
*Standard errors were estimated using a jackknife replication method.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

To identify relative strengths and weaknesses of U.S. students in broad terms, we compared the effect sizes, that is, the deviations of the U.S. means from the international means measured in units of standard deviations of the international scale. Table 9-1 also shows the effect sizes by domain for each grade. Effect sizes measure the differences between two distributions having similar standard deviations but differing means. The effect size for each of the six distributions is given as the difference between the U.S. national mean and the international mean of 500, divided by the international standard deviation (100). The effect sizes shown are generally high, ranging from 0.30 to 0.55. Relative to the respective international distribution, the strongest performance for grade 4 is in the narrative domain and for grade 9 is in the expository domain. Relative to the international distributions, the weakest performance for U.S. students is in the grade 9 document domain.

Figure 9-1 shows the **simultaneous** 95 percent confidence intervals by the reading literacy domains for fourth and ninth grade students, comparing the six means simultaneously with their respective international means of 500, using a Bonferroni multiple comparison procedure with alpha = 0.05/6 for each comparison. Thus, the confidence interval limits are given by the mean plus and minus 2.64 times the standard error in each case. The performance of the U.S. students is shown to be above the international mean for each domain, for both grades. For each population, we can state with 95 percent confidence that none of the three U.S. means are below the respective international mean. This is because, using this procedure, on 95 percent of occasions over repeated sampling these six confidence intervals will all capture the respective population means simultaneously.

Figure 9-1. Simultaneous 95 percent confidence intervals for mean reading proficiency, by reading literacy domains: Grades 4 and 9



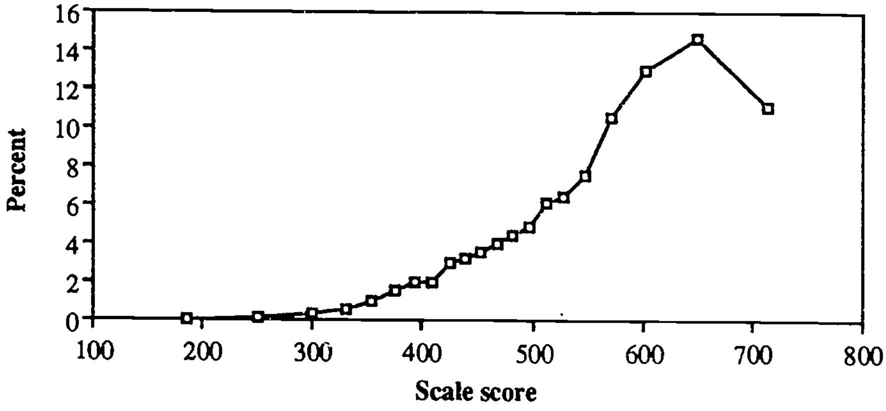
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figures 9-2 and 9-3 present the distributions of scale scores by specific domain and grade. These figures show the percentage (weighted) of students obtaining each score. For example, Figure 9-2 indicates that the distribution of scaled scores for the U.S. fourth grade students in the narrative domain is negatively skewed (i.e., with a long tail at the lower end of the distribution). The fact that there is no tail at the upper end of the distribution indicates a high degree of ceiling effect (i.e., about 11 percent of the students received the maximum score by correctly responding to all the narrative items). In contrast, the distribution of scaled scores for the U.S. ninth grade students in the narrative domain (Figure 9-3) is relatively more symmetric, flatter (i.e., the peak of the distribution is not as high), and exhibits a lesser degree of ceiling effect (i.e., about 7 percent of students received the maximum score).

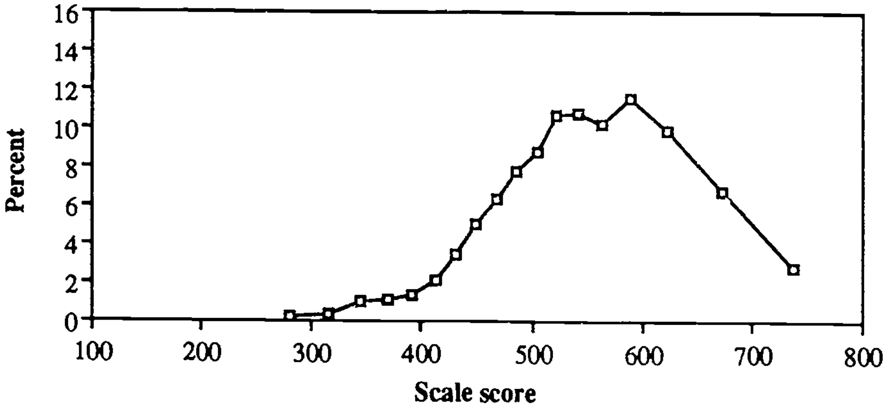
The distribution of scaled scores for the U.S. fourth grade students in the expository domain appears more symmetric than the distribution of scores for that population in the narrative domain (Figure 9-2). The ceiling effect, as represented by the percentage of students receiving the highest possible score, is also lower (i.e., only 3 percent of students received the maximum score). The distribution of scaled scores for the U.S. ninth grade students in the expository domain, as compared to the corresponding distribution for fourth grade, is flatter and shows a slightly higher degree of ceiling effect (i.e., about 5 percent of students have received the maximum score). This difference in "flatness" of the two expository scaled score distributions is also reflected in the substantial difference in the two measures of dispersion (i.e., standard deviation) -- for grade 4, the standard deviation was 79.9; for grade 9, it was 105.8. Because the international distribution has a standard deviation of 100, it can be concluded that the U.S. distribution in the expository domain is flatter (i.e., more dispersed) at grade 9 than at grade 4, relative to the aggregate of other countries.

Figure 9-2. Distributions of scaled scores by reading literacy domain: Grade 4

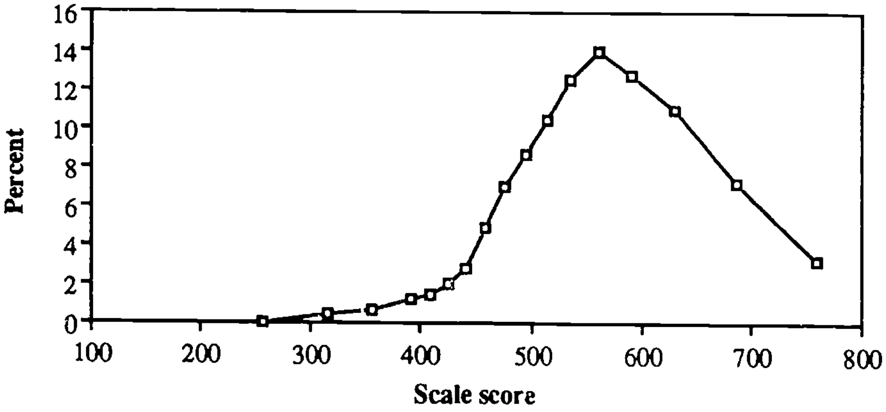
Narrative



Expository



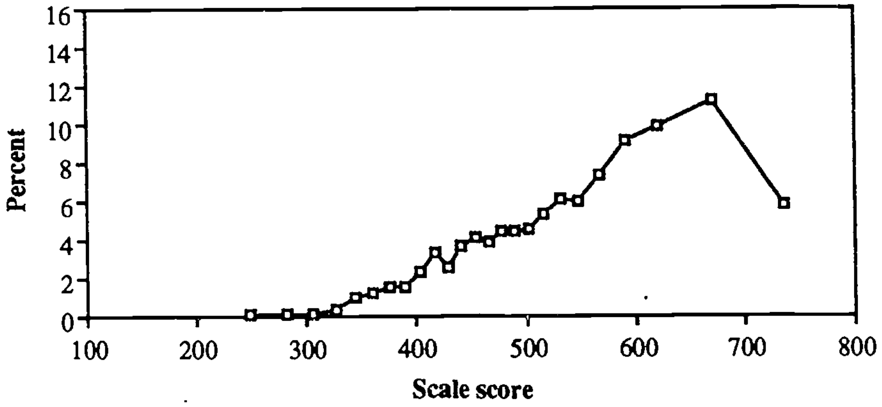
Document



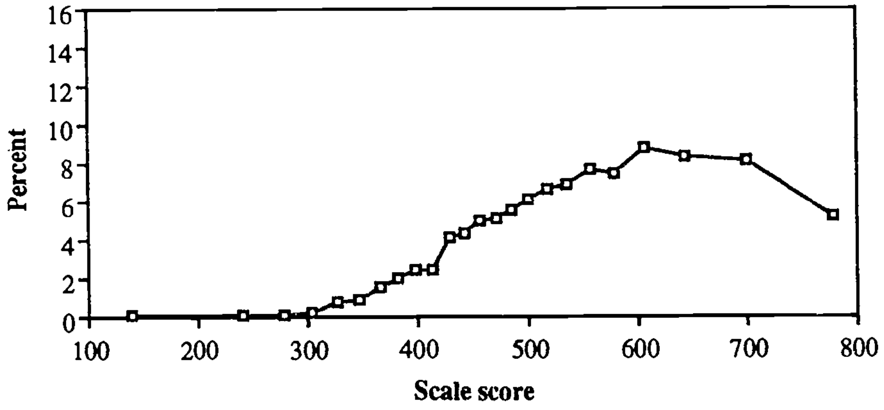
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-3. Distributions of scaled scores by reading literacy domain: Grade 9

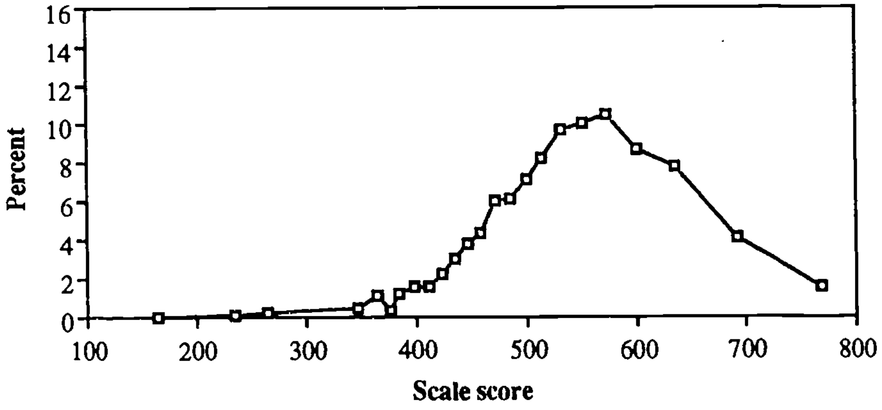
Narrative



Expository



Document



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The distributions of scaled scores for the document domain are more nearly symmetric for both populations and show low ceiling effect -- about 4 percent of grade 4 students and 2 percent of grade 9 students received the maximum scores.

Based on the shapes of these distributions we conclude that ceiling effects, especially for the fourth grade narrative domain, may attenuate the relationship between narrative reading literacy domain scores and explanatory variables of interest. Subpopulation differences may be masked because ceiling effects may vary across subpopulations of interest.

9.2.2. Percentiles and Percentages of Students Above the U.S. and International Means for Each Grade

When the distributions of scores depart from symmetry, as is the case for the fourth grade narrative domain, the mean may not be the most appropriate statistic to summarize the central tendency of a distribution. Thus, it may also be instructive to examine the percentile points of the distributions of reading literacy scaled scores. Accordingly, Table 9-2 gives each of the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentile points for each of these distributions. The median, or the 50th percentile, is the point that divides the distribution into two parts such that an equal number of scores fall above and below that point.

Table 9-2. Percentile for distribution of scores for each grade and domain

Percentile	Grade 4			Grade 9		
	Narrative	Expository	Document	Narrative	Expository	Document
5th	392.8	411.7	424.3	389.5	381.6	402.5
10th	423.8	431.1	458.3	415.8	427.6	434.0
25th	480.8	485.5	494.9	477.0	471.4	485.4
50th	548.3	542.5	536.9	548.4	536.9	532.4
Mean .	555.2	539.4	550.9	541.9	543.5	530.4
75th	649.9	589.4	591.5	621.1	607.1	573.5
90th	714.4	621.9	629.4	670.1	700.5	635.8
95th	714.4	671.9	686.1	736.3	778.2	691.6

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In calculating these percentile points, two technical difficulties were encountered. First, the SAS computer program generally considered appropriate for estimating these points did not handle sampling weights. Second, the scaled scores were not continuous, and the gap between adjacent scores was not constant. Given that the scaled scores were derived by a nonlinear transformation of the raw score, correctly identifying the real limits for each scale score was difficult.

To overcome these problems, the percentile points were estimated on the weighted raw score distributions. For example, the test for the grade 4 narrative domain contained 20 items, and so the raw score distribution ranged from 0 to 20. The percentile points on the raw score metric were converted to scaled scores using appropriate transformations. Because the raw scores were contiguous (i.e., all integer values between 0 and 20 were possible raw scores), there was no problem in identifying the correct real limits. In calculating the percentile points, we assumed that the cases falling within an interval were evenly distributed across the interval. This assumption, which is necessary for calculation of the percentile points and is common practice, usually has little impact on the determination of percentile points.[1] Note that, for the fourth grade narrative domain, 11 percent of students had received the maximum score. Thus, the 90th and 9th percentiles for this distribution are both estimated to be 714.4, the maximum score.

It is also instructive to estimate the proportion of U.S. students above the international mean (500). These percentages are substantially above 50 percent (Table 9-3), indicating that, overall, a large majority of U.S. students performed above the international mean. For both populations the U.S. students' performance was higher than the international mean, although the figures for grade 4 were slightly better than those for grade 9.

Table 9-3. Percentage of U.S. students scoring above the international and national means

Domain	Grade 4		Grade 9	
	International	National	International	National
Narrative	69.4	49.3	65.4	49.4
Expository	71.5	52.0	65.3	45.5
Document	71.2	48.2	67.6	52.6

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

A comparison of percentages of students above the international mean and the corresponding figures for the U.S. mean shows that the percentage of U.S. students scoring above the international mean is much greater than of those scoring above the U.S. mean. This is as expected given the large effect sizes of the U.S. means relative to the international distribution.

Theoretically, if the distribution of scores were symmetric or nearly symmetric the percentage of students scoring above the national mean would be close to 50 percent. If the distribution were negatively skewed, the percentage of students scoring above the U.S. national mean would be greater than 50 percent. Conversely, if the distribution of scores were positively skewed (i.e., had a long tail at the upper end of the distribution), the percentage of students scoring above the mean would likely be less than 50 percent. Because the estimates of the percentages above the U.S. mean are based on samples, their sampling variability needs to be considered. The 95 percent simple confidence intervals for each of these estimates all include 50 percent, indicating that the observed deviations from 50 percent could be attributed to sampling variability.

---

[1] The percentiles reported here are a little different from those reported internationally for the U.S. results. The IEA reported percentiles were based on the scaled scores where the scaled scores were interpolated to avoid the problem of "gaps" in the scaled scores. We consider that the assumption of an even distribution across each interval is more tenable for the raw scores than for scaled scores, but this is a matter of opinion.

9.3. Domain-Level Means and Distributions of Scale Scores by Selected Demographic Variables

We cannot present here all of the subpopulation estimates that are possible using the U.S. data for the IEA Reading Literacy Study. Nevertheless, to provide the reader with a broad picture of reading proficiencies of U.S. students, as measured by this assessment, we have selected a few demographic variables that are commonly used to report achievement results. All results reported in this chapter are weighted (i.e., incorporate the sampling weights derived to obtain unbiased population estimates).

Tables 9-4, 9-6, 9-8, and 9-10 present the average reading proficiency of the subpopulations of interest for the two populations of U.S. students. The standard errors associated with estimating the mean reading proficiencies, provided in parentheses, show that the precision of estimating the subpopulation means varies considerably across groups. For example, because the standard error for estimating the average reading proficiency for males is generally small, it can be concluded that the precision of the estimated mean proficiency score for this group is high. In contrast, the standard errors for the mean reading proficiencies for Asian and Pacific Islanders and American Indians are generally large, indicating that the precision of estimating mean proficiency for these subpopulations is relatively low. These findings are consistent with the variations in subsample size for these various demographic subgroups.

Because more than two categories of demographic variables were involved, with the exception of gender which had only two categories, it is possible to conduct more than one test of statistical significance between categories within a family defined by a given demographic variable. To avoid the problem of inflating the chance of falsely concluding that a difference exists, beyond the nominal alpha level, appropriate statistical techniques were used. Specifically, a Bonferroni multiple comparison procedure, with alpha equal to 0.05 divided by the total number of pairwise comparisons possible within the family, was used to ascertain statistical significance associated with the differences among the observed mean proficiencies. For example, because each one of the five race/ethnicity categories can be compared to each other, a total of 10 comparisons are possible. Thus, a Bonferroni multiple comparison procedure with alpha = 0.05/10 = 0.005 for each comparison was used to ascertain statistical significance of differences between mean proficiencies among the race/ethnicity subpopulations. This multiple comparison procedure was applied separately within each domain and grade so that no attempt has been made to express the statistical significance of subpopulation differences summarized across these six combinations simultaneously. Thus, looking ahead to Table 9-6, we do not claim to be 95 percent certain that for no grade or domain do either black or Hispanic students have higher mean achievement than white students. For each grade and domain separately, we are 95 percent certain that whites have a higher mean proficiency than both blacks and Hispanics. In each case the standard error of the difference was calculated directly using the jackknife procedure, thus appropriately reflecting the impact of the covariance between the subgroup means induced by the sample design and weighting procedures.

The lower portions of Tables 9-4, 9-6, 9-8, and 9-10 summarize the comparisons among the subpopulations of interest and their statistical significance. Because only the subgroups within a given demographic variable were compared to each other, comparisons among subgroups defined by two different demographic variables (e.g., males and students in the Northeast) were not included. For the purpose of illustration, let us assume we want to compare fourth grade Hispanic and white students in terms of their mean narrative scores (see Table 9-6). To determine the statistical significance of the difference in the subgroup means (570.0 for white students, 527.8 for Hispanics), we first locate the subgroup "HI" (Hispanic) for the narrative domain within the rows of the lower portion of Table 9-6, then

"WH" (white) within the columns. In this lower section of the table, the intersection of the third row and the first column shows a "1," indicating that the mean for fourth grade Hispanic students was significantly lower than that of white (non-Hispanic) students on the narrative scale. To use these tables correctly, always locate the first group within the rows of the table, then the second group within the columns.

Each table is followed by a corresponding set of graphs (Figures 9-4, 9-7, 9-10, and 9-13). These graphs compare the mean proficiencies of the U.S. subpopulations with the international mean of 500, with confidence intervals constructed around the estimated means. Because a number of comparisons to the international mean were involved for each grade, the simultaneous confidence intervals were constructed using a Bonferroni multiple comparison procedure with overall alpha equal to 0.05.

Thus, for example, when comparing the 15 means for each grade derived for the five racial/ethnic groups for each of the three domains, the confidence interval for each mean was obtained using alpha equal to 0.5/15. Thus, the confidence interval in each case is given by the mean plus and minus 2.93 times the standard error.

Generally, comparing the central tendencies (means) of two distributions does not provide a full description of how these distributions may differ from each other. Thus, it is also instructive to examine the distributions of scale scores within each category of the demographic variables under consideration. Figures 9-5, 9-6, 9-8, 9-9, 9-11, 9-12, 9-14, and 9-15 compare the shapes of the distribution of scale scores for the various demographic subgroups for each grade and reading domain.

To place the comparisons among subgroups somewhat in context, and as a guide to the validity of the results of the Reading Literacy Study, the subgroup comparisons are contrasted with analogous comparisons from the 1992 National Assessment of Educational Progress (NAEP) reading assessment. Because the Reading Literacy Study and NAEP scales are different, it is not possible to equate strictly the results of the two studies. Since both studies are assessments of reading proficiency, meaningful parallels between the results of the two studies can be drawn, once adjustments have been made for the differences in location and spread of the two scales.

By presenting subgroup mean differences in units of the population standard deviation (S.D.) from each study separately, we can contrast IEA Reading Literacy Study and NAEP results in terms of these roughly comparable units. Note that the NAEP results are for grades 4 and 8, rather than for grades 4 and 9.

Subgroup comparisons for the Reading Literacy Study and NAEP are contrasted for gender (Table 9-5), race/ethnicity (Table 9-7), and region (Table 9-9). No such comparison is possible for community size, as the two studies do not use comparable classifications of community size.

9.3.1. Gender

The means for both males and females for each domain and grade were substantially and significantly above the international mean of 500 in each case, as shown in Table 9-4 and Figure 9-4. The confidence intervals in Figure 9-4 are constructed using an alpha level of 0.05/6, with each confidence interval given as the mean plus and minus 2.64 times the standard error.

In comparing male and female students, we see that the means are very similar for the document domain at both grades. For the expository domain, females have somewhat higher estimated means than males, but, as shown in Table 9-4, these differences are not statistically significant. For the narrative domain, however, the mean for females is substantially higher than that for males at each grade, and these differences are both statistically significant.

Table 9-4. Mean reading proficiency by gender: Grades 4 and 9

Gender	Grade 4			Grade 9		
	Domain			Domain		
	Narrative	Expository	Document	Narrative	Expository	Document
Males	546.1 (3.6)	535.1 (3.2)	552.1 (3.0)	530.3 (6.2)	541.3 (7.5)	530.0 (4.9)
Females	564.4 (3.1)	543.9 (3.0)	549.6 (2.8)	553.7 (5.0)	545.8 (5.7)	530.7 (4.0)

Domain/gender		Significances of mean differences (α = 0.05 for each grade by domain)			
		Grade 4		Grade 9	
		M	F	M	F
Narrative	M	-	l	-	l
	F	h	-	h	-
Expository	M	-	ns	-	ns
	F	ns	-	ns	-
Document	M	-	ns	-	ns
	F	ns	-	ns	-

ns = no significant differences; l = row group lower than column group; h = row group higher than column group. For a discussion of how to use this table, see pages 134 and 135.

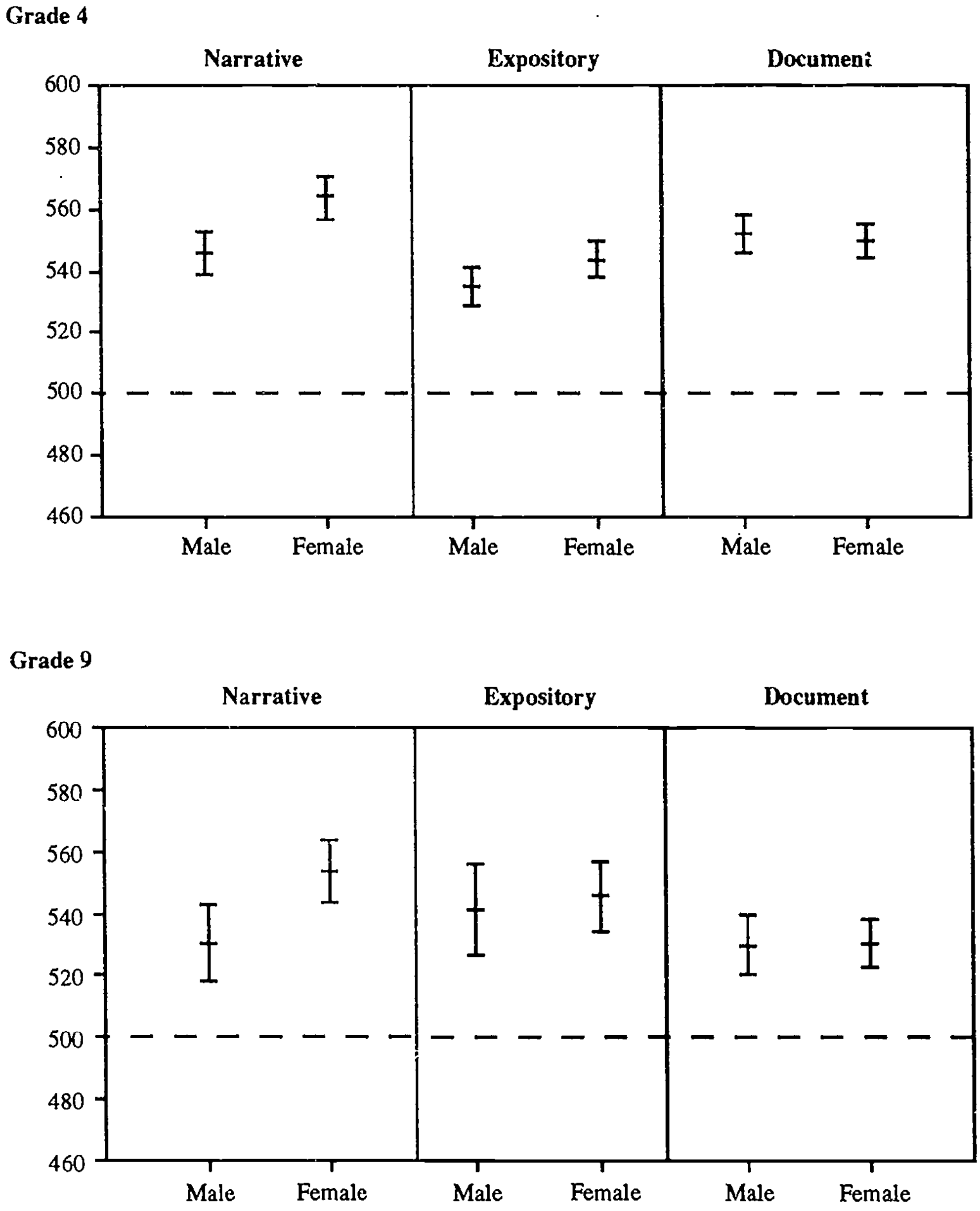
Key to gender subpopulations: M = Males; F = Females

NOTE: Standard errors, which appear in parentheses, were estimated using a jackknife replication method.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Gender differences in distributional shape are typically small (Figures 9-5 and 9-6). Relatively speaking, the largest difference between males and females was observed in the fourth grade narrative domain, and there is some difference between the ceiling effects, especially at grade 4. The advantage for females appears to be primarily among the more proficient students, and the differences reflected in the mean scores may be understated because of the variation in ceiling effects. Figure 9-5 shows that, for grade 4, 13 percent of females correctly responded to all narrative scale items, compared to 9 percent of males. For grade 9, the corresponding percentages are 7 percent for females and 5 percent for males (Figure 9-6).

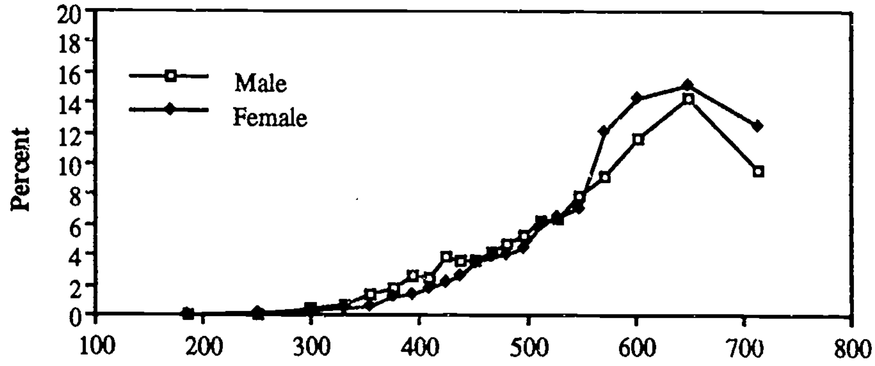
Figure 9-4. Simultaneous 95 percent confidence intervals for mean reading proficiencies, by reading literacy domain and gender: Grades 4 and 9



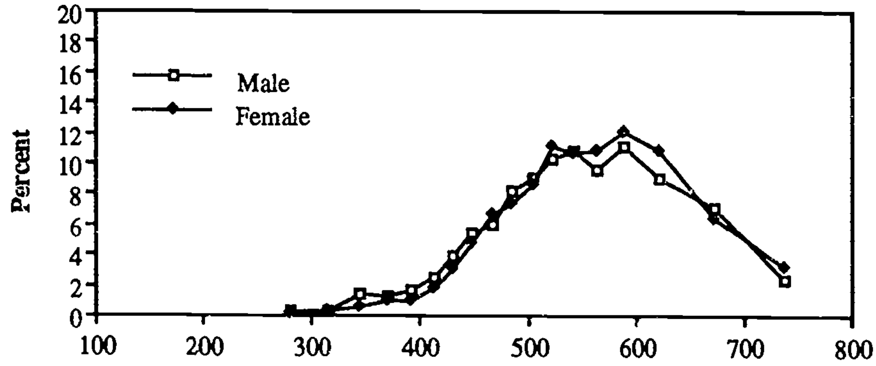
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-5. Distributions of scaled scores by gender: Grade 4

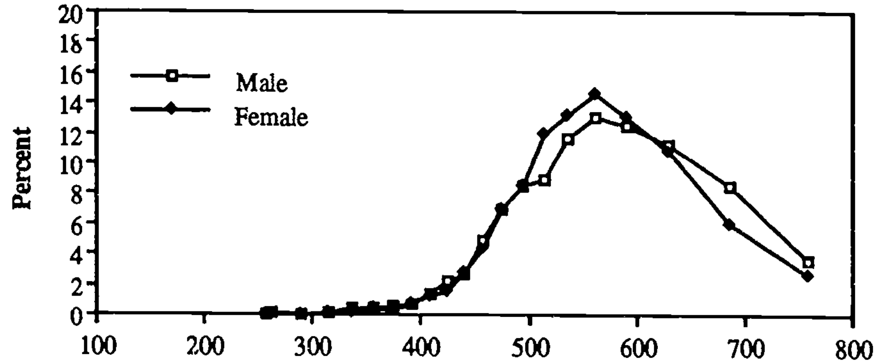
Narrative



Expository



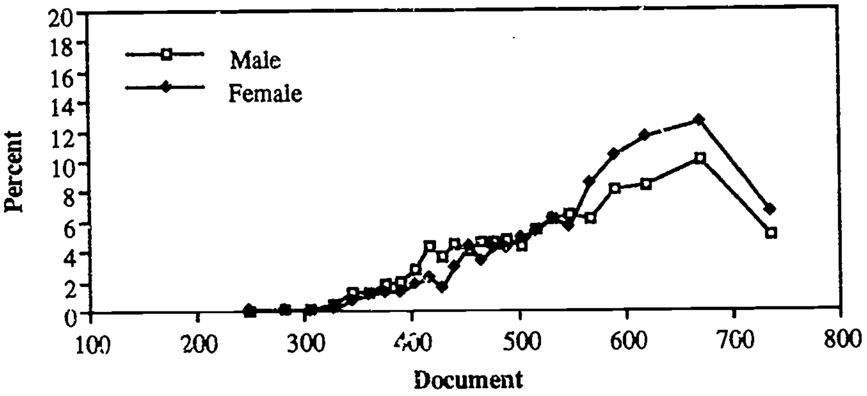
Document



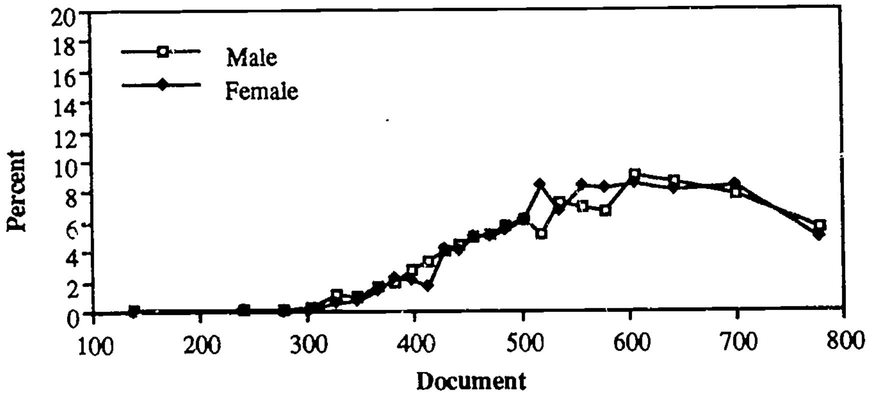
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-6. Distributions of scaled scores by gender: Grade 9

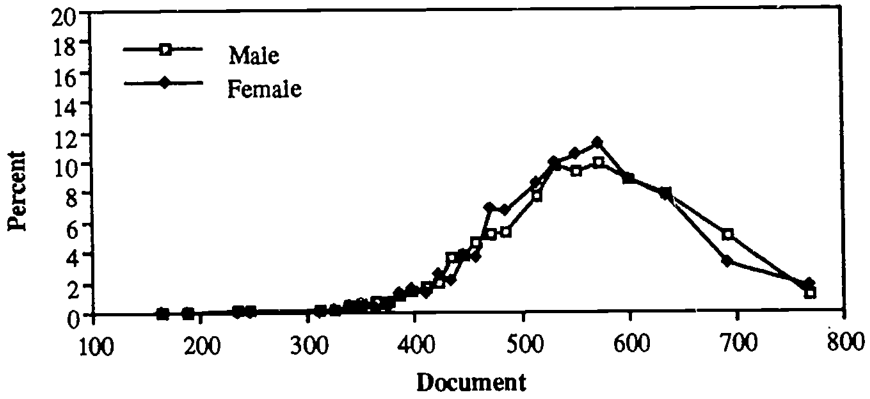
Narrative



Expository



Document



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As shown in Table 9-5, the 1992 NAEP reading assessment shows substantially higher means for females than males at grade 4 and especially at grade 8. Results are reported separately for the three NAEP subscales: reading for literacy experience, reading for information, and reading to perform a task (the last of these was not constructed at grade 4). These three scales correspond somewhat to the Reading Literacy Study scales of narrative, expository, and document, respectively.

Table 9-5. Comparison of gender means - IEA Reading Literacy Study and NAEP

Subgroup	IEA Reading Literacy Study			1992 NAEP		
	Narrative	Expository	Document	Literary experience	Information	Perform a task
Grade 4						
Overall mean	555.2	539.4	550.9	220.3	214.9	-
Standard deviation (S.D.) .	95.7	70.9	81.0	37.3	38.0	-
Male	546.1	535.1	552.1	215.8	211.6	-
Female	564.4	543.9	549.6	225.0	218.4	-
(Female-Male)/S.D.	+0.19	+0.12	-0.03	+0.25	+0.18	-
Grade 8 (NAEP)/ Grade 9 (IEA)						
Overall mean	541.9	543.5	530.4	259.0	261.1	261.2
Standard deviation (S.D.) .	97.5	105.8	82.0	37.9	36.3	39.3
Male	530.3	541.3	530.0	252.5	255.2	254.3
Female	557.8	545.8	530.7	265.6	267.1	268.2
(Female-Male)/S.D.	+0.24	+0.04	-0.01	+0.35	+0.33	+0.35

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991; 1992 NAEP Reading Assessment Data Almanac.

Table 9-5 also shows male-female differences expressed as standard deviation units, both for the three IEA scales and for the NAEP reading scales. The table shows that gender differences are probably less marked for the Reading Literacy Study than for NAEP. For grade 9, the largest difference for the Reading Literacy Study is on the narrative scale, where females have a mean 0.24 standard deviations higher than males. For NAEP the difference on the reading for literacy experience scale is 0.35 standard deviation units. Even though the gender difference varies considerably across scales in the Reading Literacy Study, it tends to be somewhat less marked than in NAEP. For the Reading Literacy Study document scale there is essentially no gender difference at either grade. This is in marked difference to the NAEP eighth grade result for the perform a task scale, where the mean for females is 0.35 standard deviation units higher than that for males.

9.3.2. Race/Ethnicity

As is shown in Table 9-6 and Figure 9-7, at both grades 4 and 9, white and Asian/Pacific Islander students had mean values above the international mean of 500 on all three domains, with statistical significance in each case. For black (non-Hispanic) students the mean scores were very close to 500 for all three domains at grade 4, and below 500 in each case at grade 9. In all six cases, however, the mean for black students was not statistically significantly different from 500. Hispanic students had

Table 9-6. Mean reading proficiency by race/ethnicity: Grades 4 and 9

Race/ethnicity	Grade 4			Grade 9		
	Domain			Domain		
	Narrative	Expository	Document	Narrative	Expository	Document
White	570.0 (2.3)	552.7 (2.6)	565.9 (2.1)	560.1 (4.5)	562.7 (5.4)	546.5 (3.7)
Black	505.4 (4.3)	499.5 (4.8)	504.2 (3.1)	481.5 (11.6)	478.2 (13.1)	472.5 (9.5)
Hispanic	527.8 (3.9)	509.1 (5.1)	520.7 (5.3)	499.5 (11.3)	506.1 (10.6)	501.2 (8.0)
Asian/Pacific Islander	567.4 (8.4)	542.3 (7.2)	551.1 (7.1)	548.9 (12.1)	562.1 (12.6)	538.7 (9.2)
American Indian/ Alaskan Native .	548.5 (11.7)	533.5 (7.4)	543.7 (8.6)	477.2 (17.6)	456.0 (21.2)	469.1 (12.3)

Domain	Race/ ethnicity	Significances of mean differences (α = .05 for each grade by domain)									
		Grade 4					Grade 9				
		WH	BL	HI	AP	AI	WH	BL	HI	AP	AI
Narrative	WH	-	h	h	ns	ns	-	h	h	ns	h
	BL	l	-	l	l	l	l	-	ns	l	ns
	HI	l	h	-	l	ns	l	ns	-	ns	ns
	AP	ns	h	h	-	ns	ns	h	h	-	ns
	AI	ns	h	ns	ns	-	l	ns	ns	ns	-
Expository	WH	-	h	h	ns	ns	-	h	h	ns	h
	BL	l	-	ns	l	l	l	-	ns	l	ns
	HI	l	ns	-	l	ns	l	ns	-	ns	ns
	AP	ns	h	h	-	ns	ns	h	ns	-	h
	AI	ns	h	ns	ns	-	l	h	ns	l	-
Document	WH	-	h	h	ns	ns	-	h	h	ns	h
	BL	l	-	ns	l	l	l	-	ns	l	ns
	HI	l	ns	-	ns	ns	l	ns	-	ns	ns
	AP	ns	h	ns	-	ns	ns	h	ns	-	h
	AI	ns	h	ns	ns	-	l	ns	ns	l	-

ns = no significant differences; l = row group lower than column group; h = row group higher than column group. For a discussion of how to use this table, see pages 134 and 135.

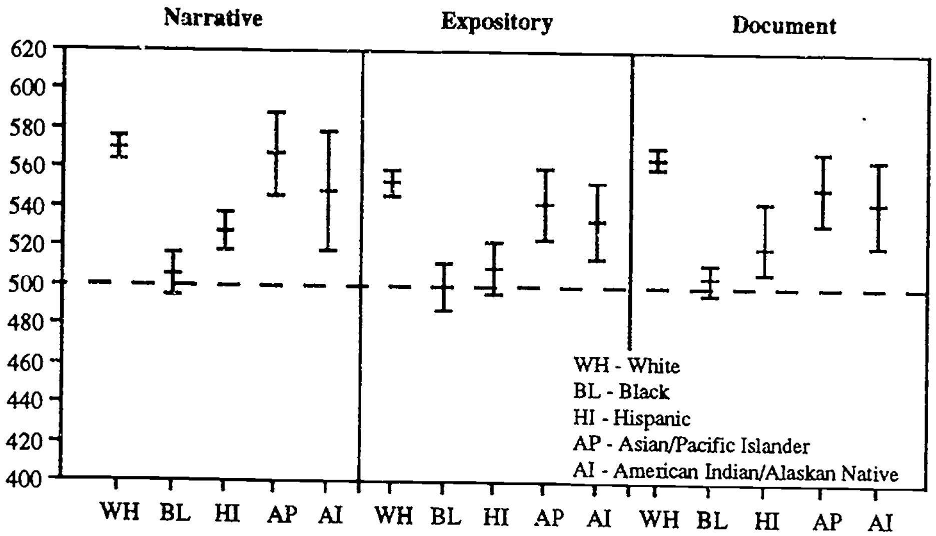
Key to race/ethnicity subpopulations: WH = White (non-Hispanic); BL = Black (non-Hispanic); HI = Hispanic (of any race); AP = Asian and Pacific Islander; AI = American Indian and Alaskan Native

NOTE: Standard errors, which appear in parentheses, were estimated using a jackknife replication method.

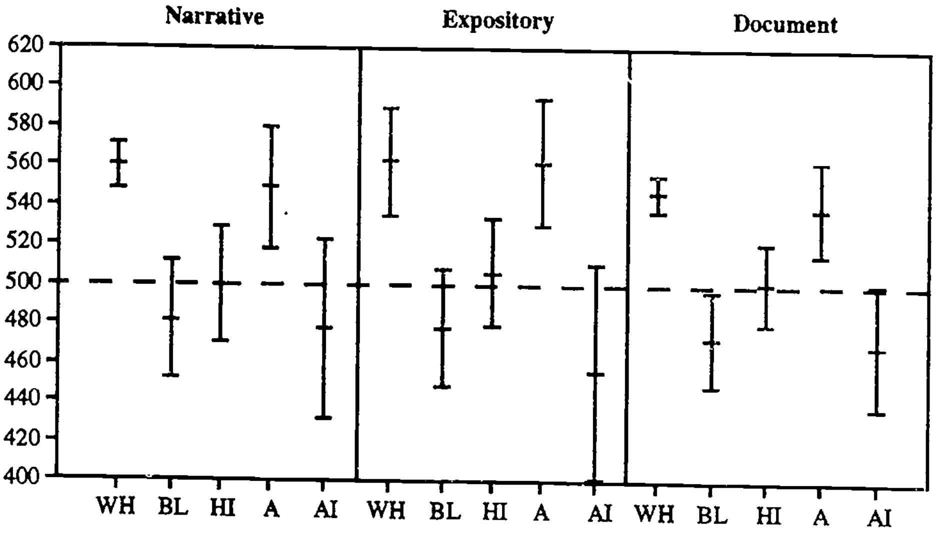
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-7. Simultaneous 95 percent confidence intervals for mean reading proficiencies, by reading literacy domain and race/ethnicity: Grades 4 and 9

Grade 4



Grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

mean scores above 500 for all three domains at both grades, but this excess was statistically significant in only two cases: the narrative and document domains at grade 4. For those students reporting themselves as American Indian (or Alaskan Native), at grade 4 the mean score for each domain was significantly above 500. At grade 9 all three means were below 500; in no case, however, was this statistically significant.

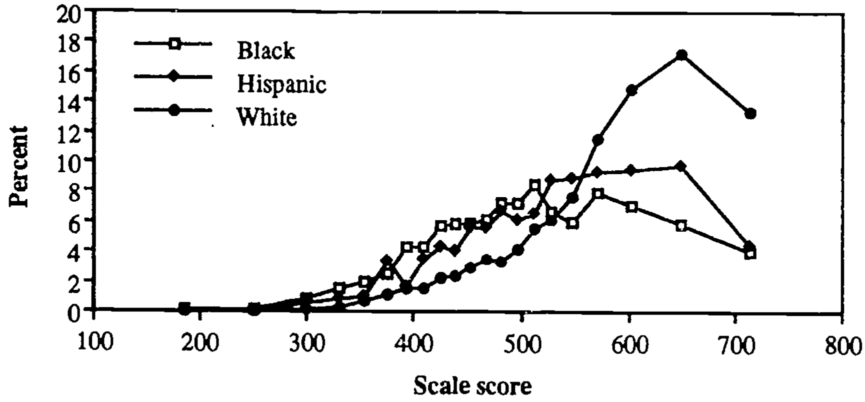
Leaving aside the comparison with the international mean, we turn to differences in mean performance between racial/ethnic groups within the U.S. Table 9-6 provides information as to the statistical significance of these differences, considering each grade and domain one at a time. As discussed earlier, a Bonferroni procedure was used for these comparisons to compensate for the fact that, with five racial/ethnic groups, there are 10 pairwise comparisons that can be made. Using an overall alpha value of 0.05 for each grade and domain, this means that an alpha of 0.005 was used for each pairwise comparison of racial/ethnic groups means. The mean scores for white students significantly exceeded those for black and Hispanic students for each domain at each grade level. The mean scores for Asian/Pacific Islander students significantly exceeded those of black and Hispanic students for the narrative and expository domains at grade 4.

Figures 9-8 and 9-9 show the distribution of scaled scores by race/ethnicity. Since the sample sizes for Asians/Pacific Islanders and American Indians were low, the distributions for these two subpopulations are not included. These figures show that students reporting themselves as white had a distribution shifted to the higher end of the scale compared to students reporting themselves as either black or Hispanic. The inequality between whites and the other two racial/ethnic groups becomes more evident when one considers the proportion of students scoring at the extreme upper ends of the distributions. For example, for the grade 4 narrative scale an estimated 6 percent of black students scored above the 75th percentile for whites, while the corresponding figure for Hispanics was 8 percent.

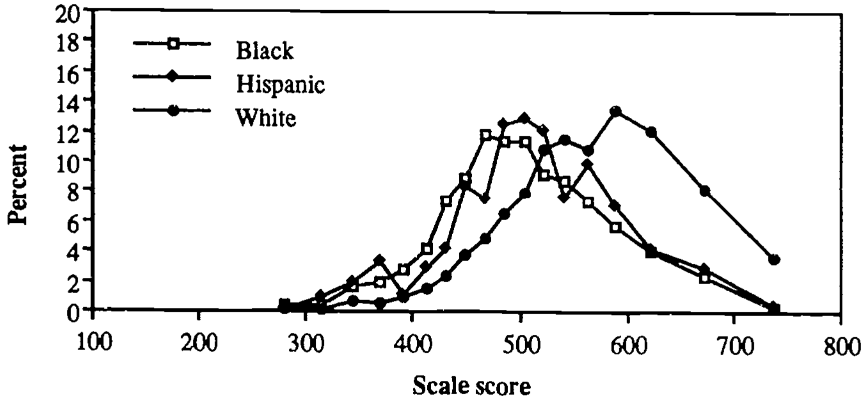
These figures show that ceiling effects are relatively more pronounced for white than for minority students. For example, for the grade 4 narrative scale, about 14 percent of white students received the maximum score, in comparison to only about 4 percent of the minority students. One implication of this finding is that mean differences between white and minority students may have been underestimated because of ceiling effects. This is because the scores for a significant proportion of the highest ability white students may be substantially understated because of the relatively easy nature of the assessment.

Figure 9-8. Distributions of scaled scores by race/ethnicity: Selected groups, grade 4

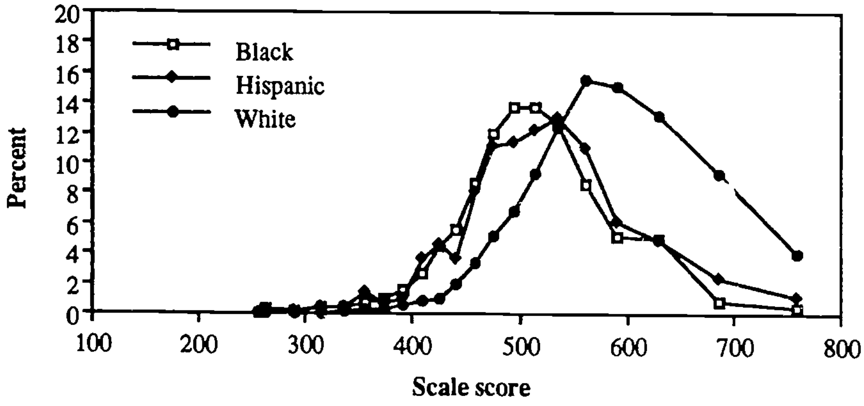
Narrative



Expository



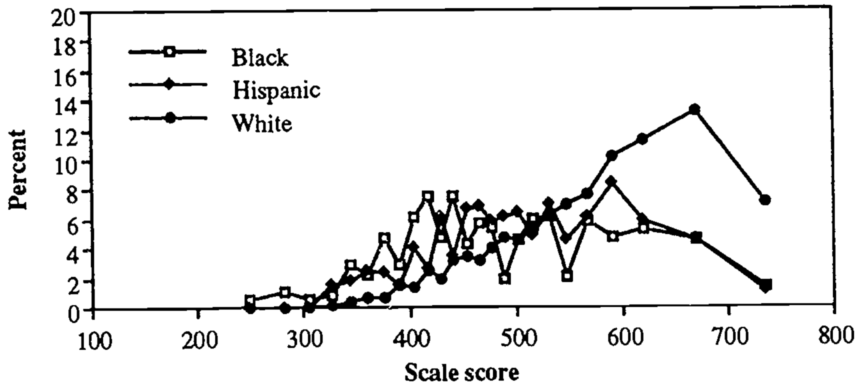
Document



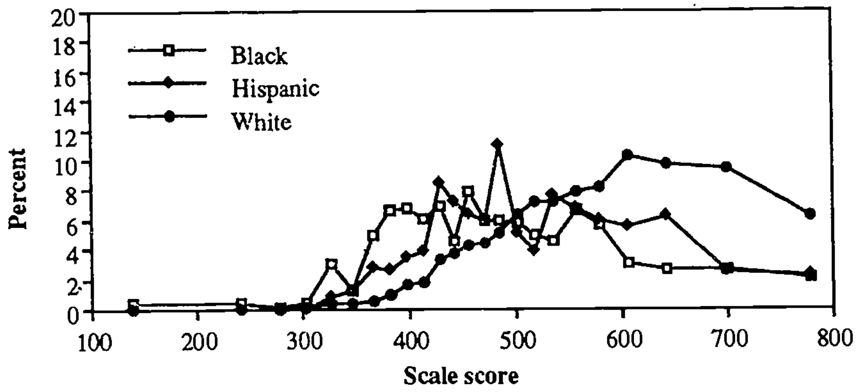
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-9. Distributions of scaled scores by race/ethnicity: Selected groups, grade 9

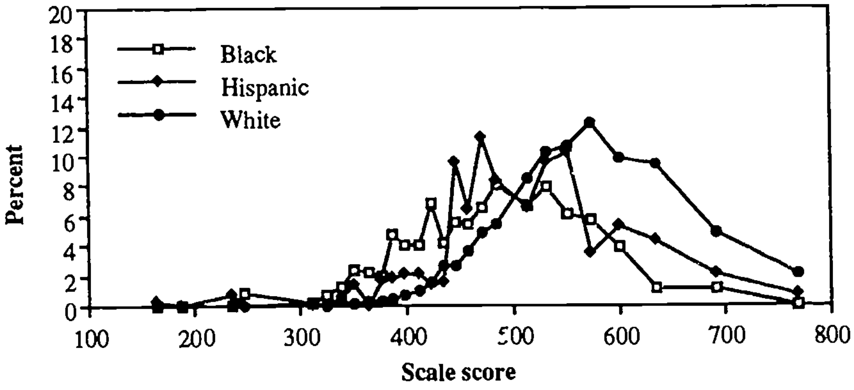
Narrative



Expository



Document



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 9-7 shows that differences between white students, on the one hand, and black and Hispanic students, on the other, are very similar across grades and scales, and are in strong concordance with NAEP results. The difference is sizable in each case. For instance, for the narrative scale, grade 4, the estimated difference between white and black student means is 0.67 standard deviations, a substantial difference.

Table 9-7. Comparison of racial/ethnic group means - IEA Reading Literacy Study and NAEP

Subgroup	IEA Reading Literacy Study			1992 NAEP		
	Narrative	Expository	Document	Literary experience	Information	Perform a task
Grade 4						
Overall Mean	555.2	539.4	550.9	220.3	214.9	-
Standard Deviation (S.D.) .	95.7	70.9	81.0	37.3	38.0	-
White	570.0	552.7	565.9	228.0	223.3	-
Black	505.4	499.5	504.2	195.7	190.3	-
Hispanic	527.8	509.1	520.7	206.7	196.2	-
(White-Black)/S.D.	+0.67	+0.75	+0.76	+0.87	+0.87	-
(White-Hispanic)/S.D. ...	+0.44	+0.61	+0.56	+0.57	+0.71	-
(Black-Hispanic)/S.D.	-0.23	-0.14	-0.20	-0.29	-0.16	-
Grade 8 (NAEP)/ Grade 9 (IEA)						
Overall mean	541.9	543.5	530.4	259.0	261.1	261.2
Standard deviation (S.D.) .	97.5	105.8	82.0	37.9	36.3	39.3
White	560.1	562.7	546.5	265.8	268.4	269.6
Black	481.5	478.2	472.5	237.9	239.2	235.9
Hispanic	499.5	506.1	501.2	241.7	242.0	240.4
(White-Black)/S.D.	+0.81	+0.80	+0.90	+0.74	+0.80	+0.86
(White-Hispanic)/S.D. ...	+0.62	+0.53	+0.55	+0.64	+0.73	+0.74
(Black-Hispanic)/S.D.	-0.19	-0.27	-0.35	-0.10	-0.08	-0.11

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991; 1992 NAEP Reading Assessment Almanac.

Thus, the Reading Literacy Study and NAEP results are very similar. That is, the differences between white and minority students, for both grades, are very similar between the two studies. This finding provides some further evidence of the validity of the IEA assessment for use in comparing population subgroup achievement means.

9.3.3. Region

Table 9-8 shows means and standard errors by region for each grade and domain. These results are presented graphically in Figure 9-10, which shows the mean relative to the international mean of 500 in each case. Simultaneous confidence intervals for the regional means have been calculated for each grade. An overall alpha of 0.05 was used so that each confidence interval is based on an alpha level of 0.05/12. Thus, the confidence intervals are constructed as the mean plus and minus 2.86 times the standard error in each case. The results show that, across grades and domains, the regional means were

Table 9-8. Mean reading proficiency by region: Grades 4 and 9

Region	Grade 4			Grade 9		
	Domain			Domain		
	Narrative	Expository	Document	Narrative	Expository	Document
Northeast	565.8 (9.8)	544.6 (9.6)	556.5 (9.7)	558.8 (15.7)	557.4 (17.7)	541.9 (10.5)
Southeast	533.5 (6.5)	521.9 (5.7)	532.1 (4.2)	525.8 (7.7)	524.6 (10.0)	512.5 (7.4)
Central	558.0 (5.7)	548.7 (5.4)	558.9 (4.4)	559.3 (8.2)	564.7 (10.0)	545.7 (7.4)
West	562.2 (3.5)	541.2 (3.4)	554.4 (2.7)	529.5 (8.6)	532.5 (9.3)	524.7 (6.1)

Domain	Region	Significance of mean differences (α = .05 for each grade and domain)							
		Grade 4				Grade 9			
		NE	SE	C	W	NE	SE	C	W
Narrative	NE	-	ns	ns	ns	-	ns	ns	ns
	SE	ns	-	ns	l	ns	-	ns	ns
	C	ns	ns	-	ns	ns	ns	-	ns
	W	ns	h	ns	-	ns	ns	ns	-
Expository	NE	-	ns	ns	ns	-	ns	ns	ns
	SE	ns	-	ns	ns	ns	-	ns	ns
	C	ns	ns	-	ns	ns	ns	-	ns
	W	ns	ns	ns	-	ns	ns	ns	-
Document	NE	-	ns	ns	ns	-	ns	ns	ns
	SE	ns	-	l	l	ns	-	ns	ns
	C	ns	h	-	ns	ns	ns	-	ns
	W	ns	h	ns	-	ns	ns	ns	-

ns = no significant differences; l = row group lower than column group; h = row group higher than column group. For a discussion of how to use this table, see pages 134 and 135.

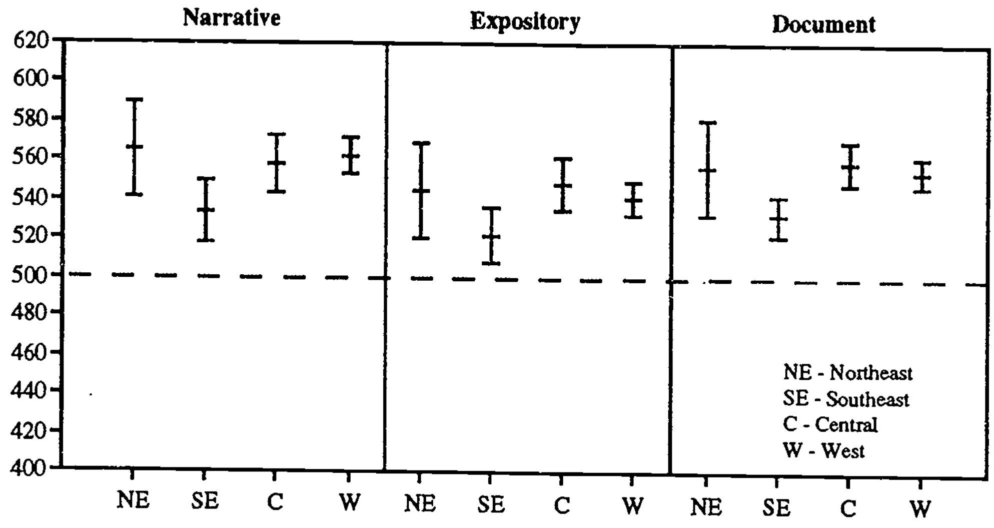
Key to regional subpopulations: NE = Northeast; SE = Southeast; C = Central; W = West

NOTE: Standard errors, which appear in parentheses, were estimated using a jackknife replication method.

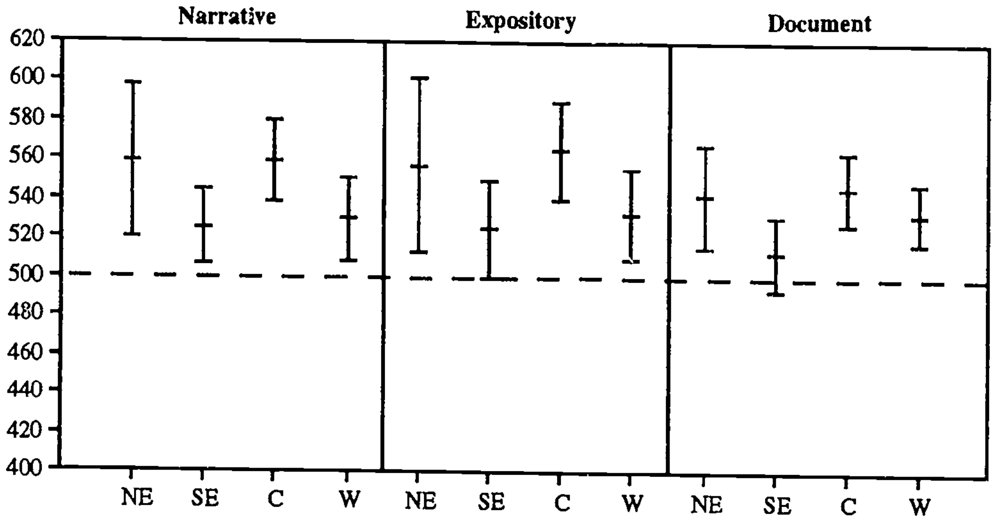
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-10. Simultaneous 95 percent confidence intervals for mean reading proficiencies, by reading literacy domain and region: Grades 4 and 9

Grade 4



Grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

consistently above the international means. These excesses were statistically significant with one exception: the Southeast region for the document domain for grade 9.

Table 9-8 summarizes the statistical significances of differences between regional means. In this case each pairwise comparison was conducted using an alpha level of 0.05/6, since there are six possible comparisons that can be made among the four regions for each domain and grade.

Although the Southeast region has estimated mean scores consistently below those of the other three regions, these differences were significant only in the case of the grade 4 assessment for the West region for the narrative and document domains, and for the Central region for the document domain. The Central region had the highest mean for five of the six domain/grade combinations but was not significantly different from the other regions except in the case noted above. The West region appeared to be similar to the Central and Northeast regions at grade 4 but substantially below these two regions and similar to the Southeast at grade 9. Due to the lack of statistical significance of the differences between regional means, especially at grade 9, this finding is very speculative.

Figures 9-11 and 9-12 show the comparisons of the distributions for each scale for each grade. It can be seen that for grade 4 the distributions are very similar in each case for the Northeast, Central, and West regions. The Southeast region shows substantially fewer students in the upper part of the scale in each case. For grade 9, there is some variation among the regions in the proportions of students scoring in the upper part of each scale.

Table 9-9 shows the differences between regions for the Reading Literacy Study and for NAEP, expressed in terms of population standard deviations. Each of the other regions is compared with the Northeast. The results show a degree of consistency across grades across the three Reading Literacy Study scales, and NAEP. For example, for the Reading Literacy Study the difference between the Northeast and Southeast is estimated to be a little over 0.3 standard deviations in each case, a substantial difference. For NAEP the estimated differences are in the range of 0.2 to 0.3 standard deviations. This similarity in regional differences between the two studies is further evidence of the validity of the IEA Reading Literacy Study. In particular, it is suggestive that when ceiling effects are minimal or similar in magnitude for subgroups being compared (as is generally the case for regions, as shown in Figures 9-11 and 9-12), they have little negative impact on subgroup comparisons in the IEA Reading Literacy.

9.3.4. Community Size

Although there is some variation in the mean scores for the five community size groups presented in Table 9-10, all are above the international mean of 500 for each domain and grade (Figure 9-13). Furthermore, at grade 4 all of these are statistically significantly above 500, using a multiple comparison procedure with alpha level of 0.05/15 (so that the confidence interval in each case is given by the mean plus and minus 2.93 times the standard error). At grade 9, all community sizes are significantly above 500 for each domain, with the exception of large city. For these, there was no significant difference from 500 for any domain.

In comparing community sizes, within domains and grades, there are no differences with any statistical significance. This is based on an alpha level of 0.05/10 since there are 10 possible pairwise comparisons for each domain and grade combination. Thus, although a consistent pattern emerges, with the medium city estimated mean being the highest in all six cases and the large city estimated mean being the lowest, this cannot be interpreted in any strong fashion. Not only are the differences not statistically

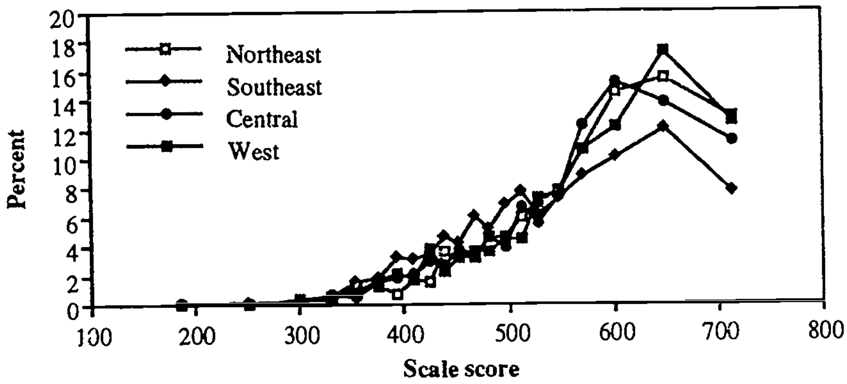
Table 9-9. Comparison of region means - IEA Reading Literacy Study and NAEP

Subgroup	IEA Reading Literacy Study			1992 NAEP		
	Narrative	Expository	Document	Literary experience	Information	Perform a task
Grade 4						
Overall mean	555.2	539.4	550.9	220.3	214.9	-
Standard deviation (S.D.) .	95.7	70.9	81.0	37.3	38.0	-
Northeast	565.8	544.6	556.5	225.3	219.8	-
Southeast	535.5	521.9	532.1	215.5	211.6	-
Central	558.0	548.7	558.9	222.0	219.0	-
West	562.2	541.2	554.4	218.9	210.1	-
(Northeast-Southeast)/S.D.	+0.33	+0.32	+0.30	+0.25	+0.22	-
(Northeast-Central)/S.D. . .	+0.08	-0.06	-0.03	+0.08	+0.02	-
(Northeast-West)/S.D.	+0.04	+0.05	+0.03	+0.16	+0.26	-
Grade 8 (NAEP)/ Grade 9 (IEA)						
Overall mean	541.9	543.5	530.4	259.0	261.1	261.2
Standard deviation (S.D.) .	97.5	105.8	82.0	37.9	36.3	39.3
Northeast	558.8	557.4	541.9	261.6	264.8	264.3
Southeast	525.8	524.6	512.2	253.0	255.2	253.8
Central	559.3	564.7	545.7	261.3	265.7	267.6
West	529.5	532.5	524.7	260.1	259.2	259.4
(Northeast-Southeast)/S.D.	+0.34	+0.31	+0.36	+0.23	+0.26	+0.27
(Northeast-Central)/S.D. . .	-0.01	-0.07	-0.05	+0.01	-0.02	-0.08
(Northeast-West)/S.D.	+0.30	+0.24	+0.21	+0.04	+0.15	+0.12

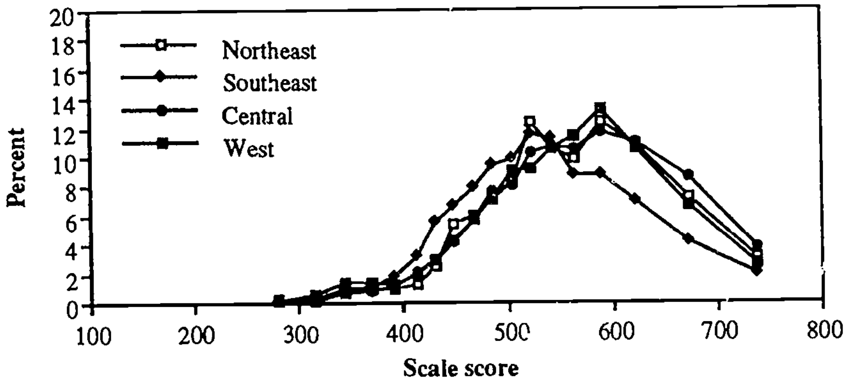
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991; 1992 NAEP Reading Assessment Almanac.

Figure 9-11. Distributions of scaled scores by region: Grade 4

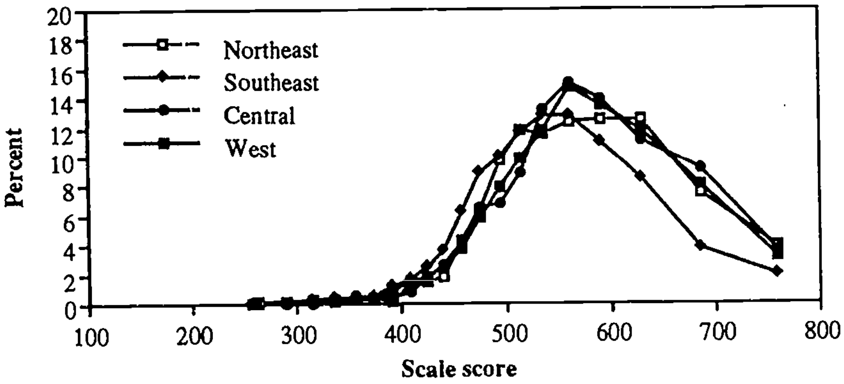
Narrative



Expository



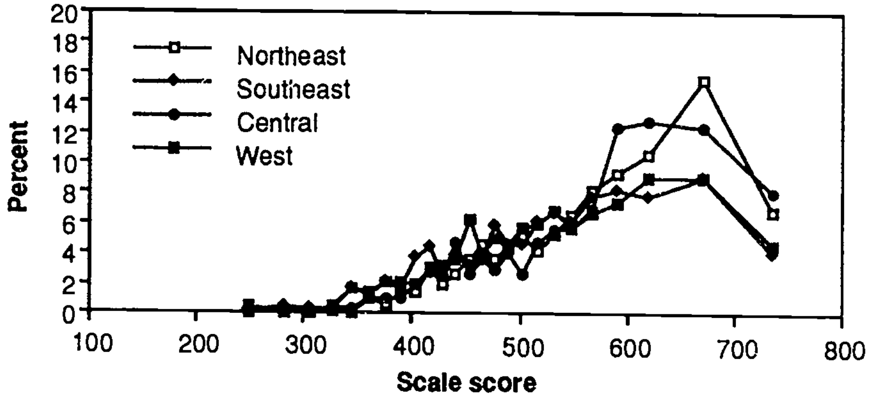
Document



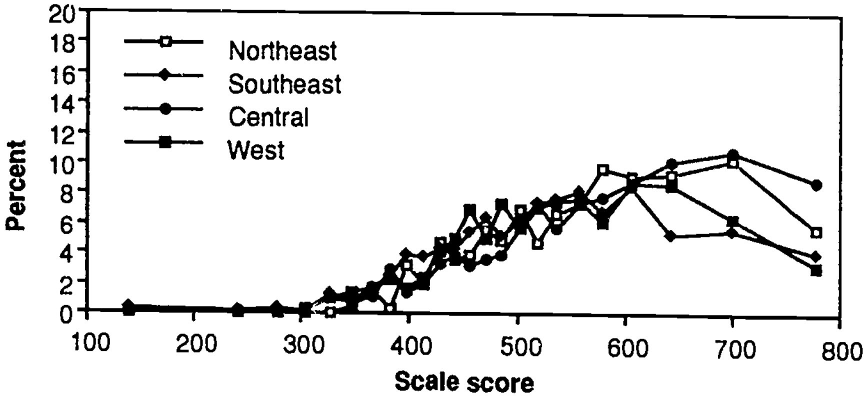
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-12. Distributions of scaled scores by region: Grade 9

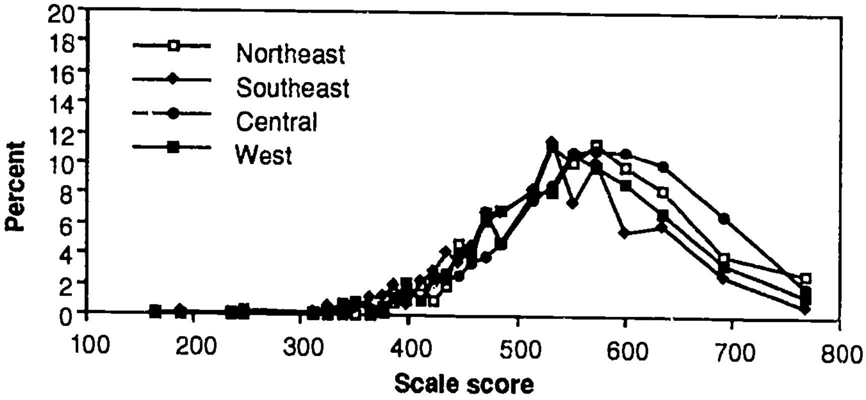
Narrative



Expository



Document



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 9-10. Mean reading proficiency by community size: Grades 4 and 9

Size of community	Grade 4			Grade 9		
	Domain			Domain		
	Narrative	Expository	Document	Narrative	Expository	Document
Rural	554.7 (7.6)	543.8 (5.9)	556.6 (6.4)	534.3 (7.5)	536.0 (10.1)	525.1 (7.0)
Small city	557.8 (6.2)	539.0 (6.3)	551.6 (6.1)	552.1 (7.4)	551.6 (7.9)	536.8 (6.2)
Medium city	564.1 (6.2)	547.5 (5.6)	556.9 (6.1)	564.3 (18.4)	567.8 (19.0)	541.6 (11.7)
Large city	546.5 (7.6)	533.0 (7.0)	542.1 (6.2)	514.6 (14.9)	516.7 (16.9)	511.0 (10.7)
Very large city	553.5 (8.7)	534.5 (8.3)	548.3 (6.9)	542.7 (13.9)	545.4 (17.8)	535.5 (11.7)

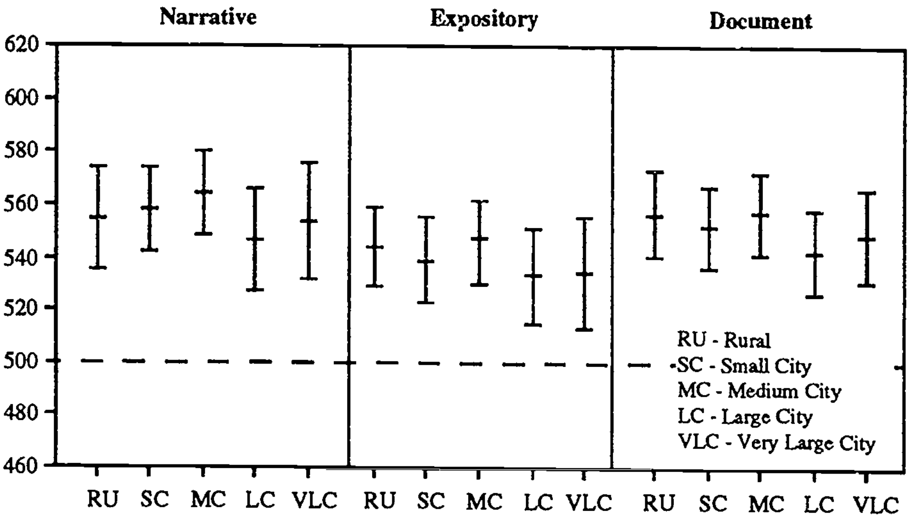
Key to community size type: Rural = rural or farming community; Small city = fewer than 50,000 people; Medium city = 50,000-100,000 people; Large city = 100,000-500,000 people; Very large city = more than 500,000 people.

NOTE: Standard errors, which appear in parentheses, were estimated using a jackknife replication method.

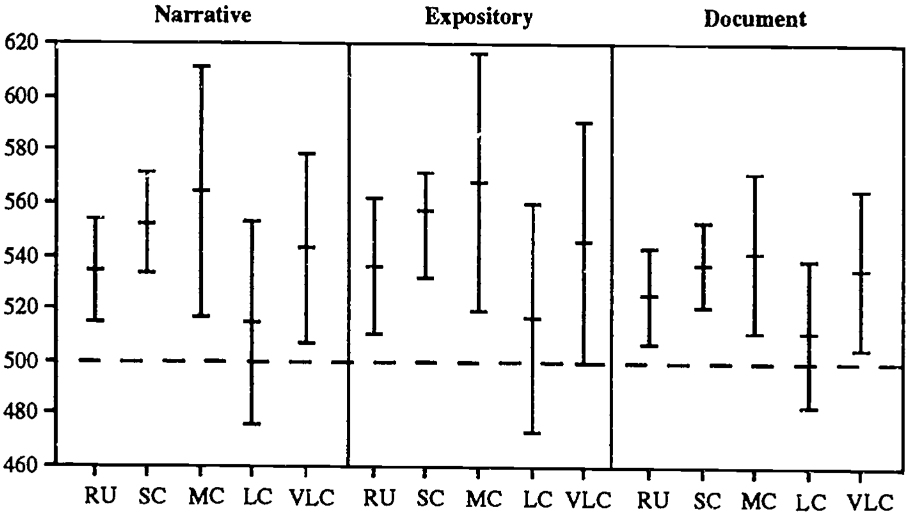
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-13. Simultaneous 95 percent confidence intervals for mean reading proficiencies, by reading literacy domain and community size: Grades 4 and 9

Grade 4



Grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

significant, but it will be noted that the estimated means for very large cities consistently fall between these two so that there is no evidence of a monotone effect of city size.

Figures 9-14 and 9-15 indicate that differences in reading proficiencies of students by size of community were generally small. In most cases, these distributions overlap to the extent that it was difficult to differentiate among them.

The community size classifications used for the IEA Reading Literacy Study are not comparable to those used to report NAEP data.

9.3.5. Summary of Subgroup Comparisons

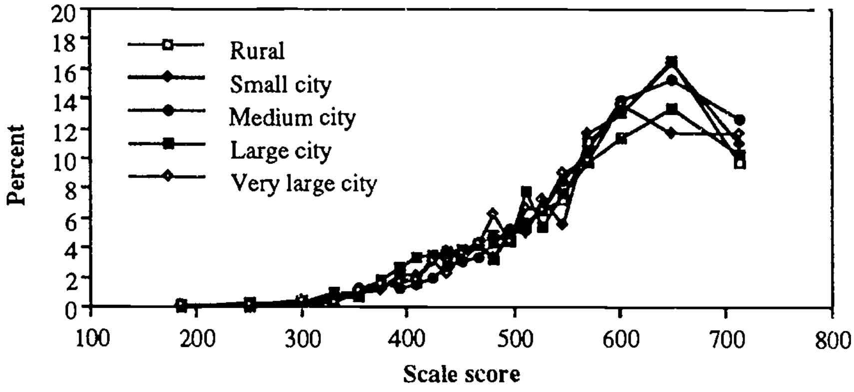
In summary, many demographic subgroups in the U.S. population have means above the international mean for all domains and both grades, and those that are not above the international mean are not significantly below it. The patterns among subgroups are very similar to those found in the 1990 NAEP reading assessment. The greatest differences are seen among racial/ethnic subgroups, with black and Hispanic students having means well below those of the rest of the nation.

The following comparisons based on the distributions are noted:

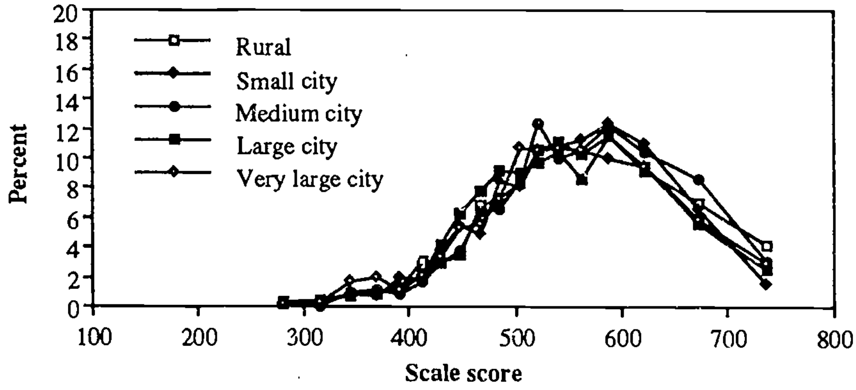
- Race/ethnicity differences were categorized by a general shift of the distribution to higher scores for white students compared to black and Hispanic students.
- Ceiling effects may have masked true differences between the male and female students. Ceiling effects appear likely to have had little impact on the comparison of mean proficiencies for other subgroups, with the exception of some subgroups for the narrative scale at grade 4.
- For other demographic characteristics (region, community size, gender), the distributions showed similar patterns across subgroups, to the extent consistent with the variations in means.

Figure 9-14. Distributions of scaled scores by community size: Grade 4

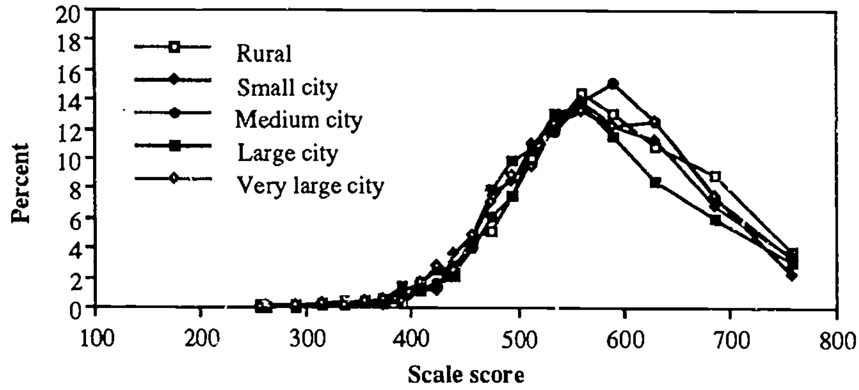
Narrative



Expository



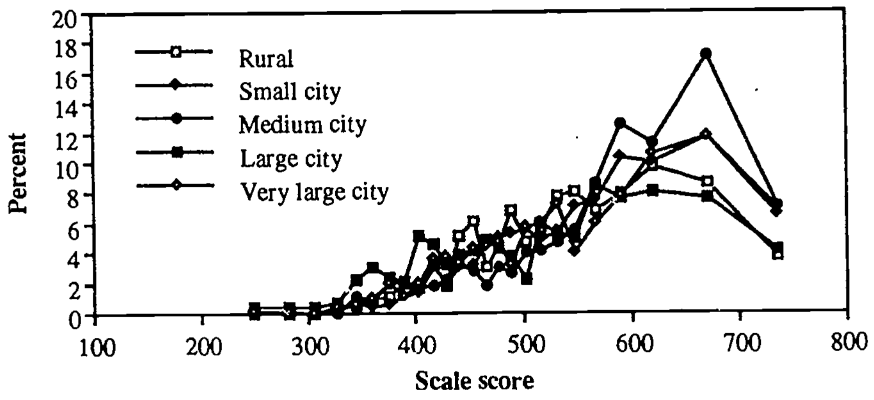
Document



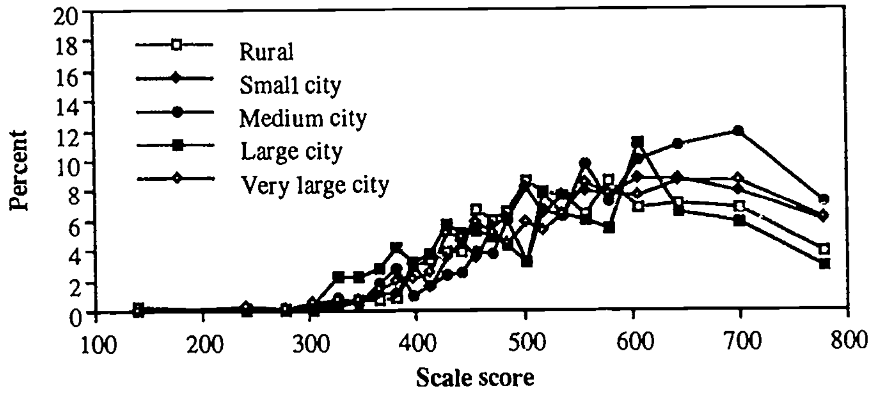
SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 9-15. Distributions of scaled scores by community size: Grade 9

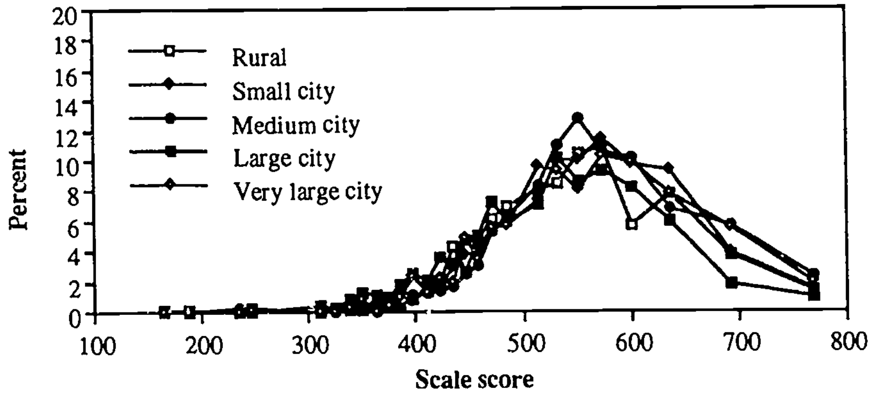
Narrative



Expository



Document



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

10. EVIDENCE SUPPORTING THE VALIDITY OF THE IEA READING LITERACY COGNITIVE INSTRUMENTS

10.1. Introduction

In the three preceding chapters we have described the reading literacy assessments and the scaling procedures used and have taken a first look at the scores, as well as examining subpopulation differences in achievement. To understand what those data indicate, it is important to establish exactly what the test measured and how it compares to instruments we are familiar with, and then to derive our interpretation of the results.

The purpose of this chapter is to establish a valid interpretation of the IEA Reading Literacy Study cognitive instruments. These tests were developed for the specific purpose of drawing inferences about population and subpopulation differences in reading literacy (investigated in Chapter 9), and factors affecting reading literacy, which will be discussed in Chapters 13 and 14. In this chapter we attempt to define clearly what the cognitive instruments measured. We do so by focusing on the available evidence of validity so that users of these data may be informed about the limitations of the instruments and consequently the range of valid uses for the resulting scores. The chapter is organized into the following sections:

- Defining validity;
- Criterion-related evidence of validity;
- Content-related evidence of validity; and
- Construct-related evidence of validity.

10.2. Defining Validity

As defined in the *Standards for Educational and Psychological Testing* (American Educational Research Association et al. 1985, 9), validity refers to

> *...the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores...A variety of inferences may be made from scores produced by a given test, and there are many ways of accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself.*

Consistent with this definition, Cronbach (1989) argues that, "validation of an instrument calls for an integration of many types of evidence. The varieties of investigation are not alternatives any one of which would be adequate." Therefore, study staff have gathered as much evidence as possible, given the lack of a specifically designed validity study so that they might establish what kinds of interpretations are valid based on the available scores.

The organization of this evidence for presentation follows the traditional means of accumulating validity evidence. As described in the *Standards for Educational and Psychological Testing*, this evidence can be grouped into three categories -- criterion-related, content-related, and construct-related evidence of validity.

Criterion-related evidence of validity comes from systematically relating the test scores to one or more outcome criteria. The underlying purpose is to establish how accurately the desired performance can be predicted from the scores on the particular test in question. Criterion-related evidence may be divided into two subcategories -- predictive and concurrent. Predictive criterion-related evidence indicates the extent to which an individual's future level on the criterion is predicted from prior test performance. Concurrent criterion-related evidence indicates the extent to which the test scores estimate an individual's present standing on the criterion.

Content-related evidence of validity "demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content. ...[Therefore,] the first task for test developers is to specify adequately the universe of content that a test is intended to represent, given the proposed uses of the test" (American Educational Research Association et al. 1985, 10). According to Cronbach (1989), "content validity is established by showing that the test items are a sample of a universe in which the investigator is interested. Content validity is ordinarily to be established deductively, by defining a universe of items and sampling systematically within this universe to establish the test. Logically, content validation is established only in test construction, by specifying a domain of tasks and sampling rigorously. The inference back to the domain can then be purely deductive."

Construct-related evidence of validity, in contrast, is evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain explanatory concepts or constructs account for performance on the test. And in the instance of construct-related evidence of validity, "...the trait or quality underlying the test is of central importance rather than either the test behavior or the scores on the criteria" (Cronbach 1989).

In the sections that follow, evidence supporting the intended uses of the IEA Reading Literacy Study findings will be presented. In considering criterion-related evidence of validity, three estimates will be provided. These are based on available data, collected at the time of testing, regarding teachers' ratings of students' literacy levels, students' self-ratings, and scores on constructed-response items. Content-related evidence for validity will be described through a discussion of internal test consistency, consensus among experts on the definition of the framework, test specifications, and by a comparison with another major American reading test, the National Assessment of Educational Progress (NAEP). Construct-related evidence for validity will be studied by examining the IRT scaling procedure results, the unidimensionality of the scale scores, the correlation among the scale scores, and relationships between scale scores and other constructs.

Consistent with the description put forth in the *Standards for Educational and Psychological Testing*, the use of the category labels and the placement of certain pieces of evidence within particular categories does not imply that there are distinct types of validity or that specific pieces of evidence only relate to that category. In fact, in many instances, the same piece of evidence might have an important bearing on establishing multiple arguments for validity. However, for the purposes of parsimony, each particular piece of evidence has been assigned to a single category.

10.3. Criterion-Related Evidence for Validity

Criterion-related evidence for validity, as defined earlier, is typically established by correlating test scores with external variables considered to be more direct measures of the characteristic under consideration. Predictive criterion-related evidence for validity cannot be evaluated in the context of the IEA Reading Literacy Study since information concerning future performance of sampled students is not and will not be available. Concurrent criterion-related evidence for validity can be evaluated in the context of the IEA Reading Literacy Study by examining the relationship between test scores and teachers' ratings of students' reading literacy levels and students' rating of their own levels. Additionally, it can be evaluated by comparing students' performance on related constructed response items and their Reading Literacy Test scores.

As part of the international design of the IEA Reading Literacy Study, a number of countries, including the U.S., collected both teachers' ratings and students' self-rating of their own reading literacy levels. Both these measures were collected at the time of testing, March 1991 (the eighth month of the school year), and were on the same scale, asking similar questions.

While we recognize the inherent problems in using judgmental ratings, these glimpses into the correlation between other sources of information regarding students' current performance in the global domain of reading proficiency and their scaled scores on the three scales can provide some evidence of the degree to which American perception of reading literacy proficiency is reflected in this test.

10.3.1. Teachers' Rating of Students' Reading Literacy Proficiency

As part of the administrator's manual, teachers were given definitions of the various levels of reading proficiency (Table 10-1) and were asked to provide a rating for each of their students based on these definitions. These ratings were recorded on the student attendance sheet. The definitions of reading proficiency levels were provided in an effort to establish a common standard across teachers. However, there is no available evidence to support any measure of cross-rater reliabilities.

Table 10-1. Teachers' rating scale

1 = very poor reader	Consistently demonstrates little understanding of what has been read. Interpretation is very literal. Often cites wrong information in response to a specific question.
2 = poor reader	Generalizes based on only one dimension. Often overlooks relevant information that may be in surrounding text.
3 = average reader	Tends to take a number of dimensions into consideration. Can develop some generalizations based on combining information from source materials but often does not account for all inconsistencies or alternative interpretations.
4 = good reader	Uses all relevant information from texts, discriminating between relevant and irrelevant information. Forms generalizations which account for a variety of possibilities. Draws from personal experiences to elaborate conclusions.
5 = very good reader	Forms generalizations based on information from the text and his experience, accounting for alternative interpretations. Tests his generalizations in new situations and applies his knowledge in new contexts.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The teacher ratings were provided by those teachers most knowledgeable about students' reading performance. For grade 9, these ratings were provided by the students' English/language arts teachers, and for grade 4, the students' reading teachers. Because these ratings were collected in March, one can argue that teachers had ample time and opportunity to formulate accurate assessments of their students' reading ability during the school year.

Table 10-2 presents the mean scale scores for each of the IEA Reading Literacy Study domains by categories of the teachers' rating scale. As shown, the mean scale scores for students with high teacher ratings are substantially higher than the mean scale scores for students with low teacher ratings. Furthermore, the pattern of increase in mean scale scores are similar for fourth and ninth grade students).

Table 10-2. Mean scale scores by categories of teacher ratings of students' reading literacy: Grades 4 and 9

Teacher rating category	Number of students	Narrative		Expository		Document	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Grade 4							
Very poor	434	440.6 (5.7)	73.7	453.0 (4.1)	61.1	468.8 (5.1)	64.1
Poor	1,118	487.3 (3.9)	73.9	483.3 (3.0)	62.1	503.2 (2.8)	62.2
Average	2,010	550.9 (3.2)	83.1	533.9 (3.2)	64.9	542.7 (2.3)	65.9
Good	1,493	591.7 (4.7)	81.1	565.6 (3.3)	66.6	576.6 (4.1)	74.1
Very good	1,019	631.7 (3.3)	69.8	610.1 (3.4)	71.1	617.1 (3.6)	75.2
Grade 9							
Very poor	95	420.2 (7.0)	58.3	415.8 (9.4)	58.7	418.7 (8.5)	66.3
Poor	515	467.4 (7.8)	80.3	463.5 (7.8)	83.0	475.5 (5.7)	69.5
Average	1,179	524.8 (5.8)	82.8	523.2 (5.8)	88.2	519.5 (5.0)	75.6
Good	919	573.5 (4.6)	88.9	577.4 (5.4)	94.0	553.2 (3.9)	71.7
Very good	598	617.0 (4.6)	79.9	628.8 (5.9)	95.5	585.5 (3.7)	71.5

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figures 10-1 to 10-6 represent "box and whisker" plots of the reading literacy scale scores by categories of teacher ratings for both grades. The figures reveal that although mean scale scores increase as teacher ratings of reading literacy increases, there is some overlap in the distribution of scale scores for categories of teachers' ratings of students' reading literacy. As expected, the overlap between adjacent groups is higher than the overlap between nonadjacent groups. If we visually superimpose the distribution of scale scores for two adjacent groups (e.g., categories 1 and 2), about half of the distribution of scores overlap. It should also be noted that the overlap between two nonadjacent groups (e.g., literacy levels 1 and 3) is quite small.

Figure 10-1. Distribution (weighted) of reading scale scores by teacher ratings, narrative domain, grade 4

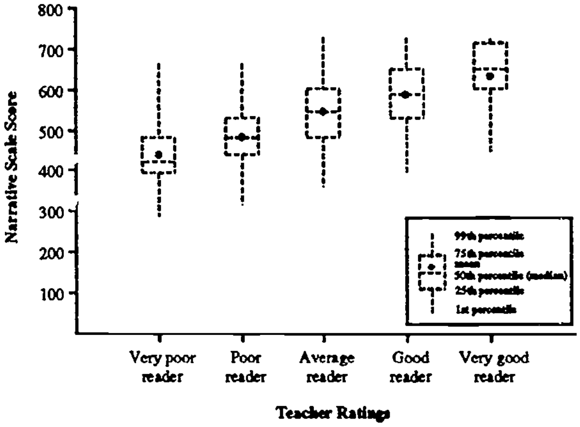


Figure 10-2. Distribution (weighted) of reading scale scores by teacher ratings, expository domain, grade 4

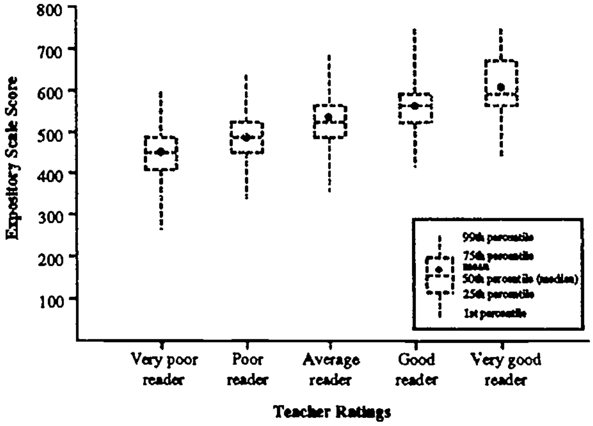
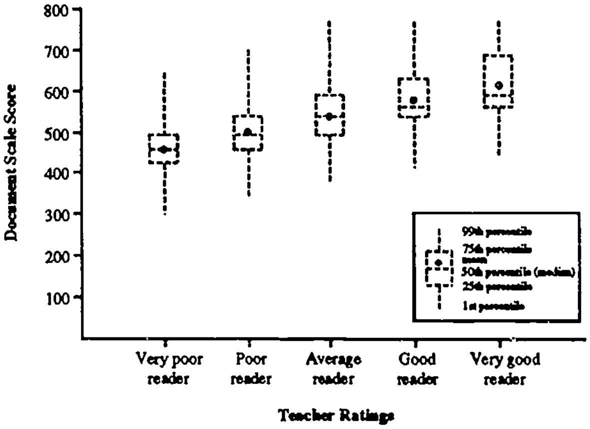


Figure 10-3. Distribution (weighted) of reading scale scores by teacher ratings, document domain, grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 10-4. Distribution (weighted) of reading scale scores by teacher ratings, narrative domain, grade 9

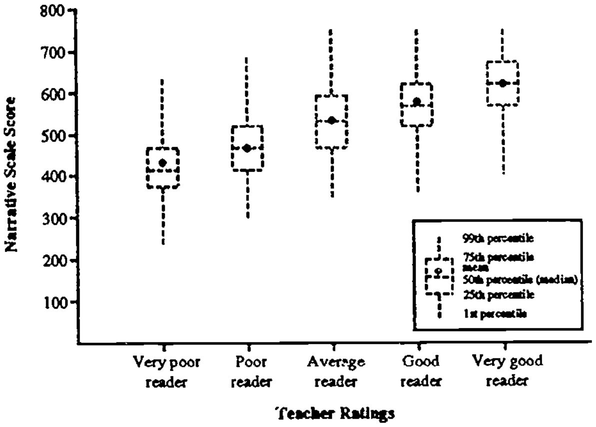


Figure 10-5. Distribution (weighted) of reading scale scores by teacher ratings, expository domain, grade 9

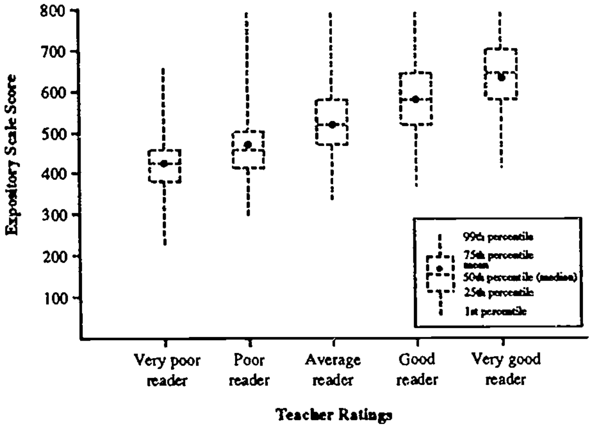
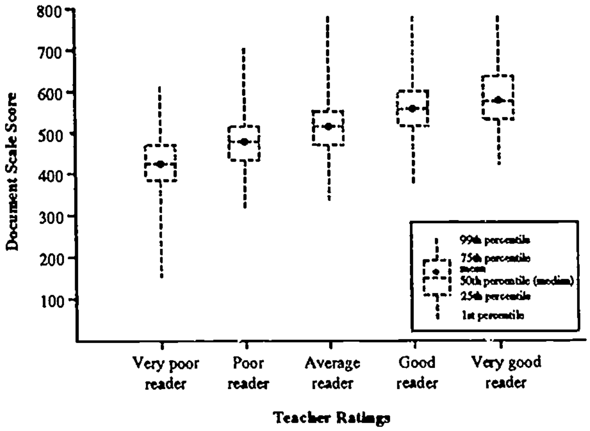


Figure 10-6. Distribution (weighted) of reading scale scores by teacher ratings, document domain, grade 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

To summarize the strength of the relationship between reading literacy scale scores and teachers' ratings of students' reading literacy levels, the coefficient of determination (R^2) was computed (Table 10-3) utilizing the fact that the criterion variable (i.e., teachers' ratings of students' reading literacy levels) was categorical.

Table 10-3. Coefficient of determination (R^2) between reading literacy scale scores and teachers' ratings of students' reading literacy levels: Grades 4 and 9

Domain	Grade 4	Grade 9
Narrative	0.329	0.283
Expository	0.324	0.291
Document	0.272	0.227

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

R^2 indicates the proportion of variance in each reading literacy scale score explained by teachers' ratings of students' literacy levels. Thus, for grade 4, teachers' rating of students' literacy levels accounted for about 33 percent of the variation in the narrative and expository scale scores, whereas the corresponding figure for document scales scores was about 27 percent. For grade 9, the corresponding figures are 28 percent, 29 percent, and 23 percent. The proportion of variance in narrative and expository scales accounted for by teachers' ratings are about the same for grades 4 and 9. Further, for both grades, the proportion of variance in the document domain accounted for by the criterion variable is lower than the corresponding numbers for the other two domains. Across all domains, the proportion of variance accounted for is larger for grade 4 than grade 9.

On the basis of the data presented above, there seems to be a significant relationship between teachers' ratings of students' reading literacy levels and scale scores based on students' responses to IEA Reading Literacy Study tests.

10.3.2. Students' Self-Ratings of Reading Literacy Proficiency

A question asking how students rate their own abilities as readers was included in the student surveys for both populations. While the wording of the question varied slightly across the two grades, the ratings were consistent across the grades and matched the rating scale the teachers were asked to use.

Student Self-Rating Questions	
(Grade 4) How good are you at reading? (Grade 9) How would you rate yourself as a reader?	
very poor	1
poor	2
average	3
good	4
very good	5

We did not expect every student to report his/her reading ability accurately. Nevertheless, in general students who have rated themselves as good or excellent readers should have higher scale scores than students who have rated themselves as poor readers. Table 10-4 presents mean scale scores for each of the IEA Reading Literacy Study domains by categories of student's own rating of his/her reading literacy level.

Table 10-4. Mean scale scores by categories of student's self-rating of reading literacy level: Grades 4 and 9

Self-rating score	Number of students	Narrative		Expository		Document	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Grade 4							
Very poor	0	- -	-	- -	-	- -	-
Poor	0	- -	-	- -	-	- -	-
Average	1,251	513.6 (3.3)	89.7	503.5 (3.0)	74.8	528.1 (3.1)	77.5
Good	2,108	541.6 (3.2)	91.3	528.0 (2.8)	73.4	544.7 (2.7)	77.2
Very good	2,811	582.4 (3.6)	94.0	563.1 (3.1)	80.1	565.2 (3.3)	83.2
Grade 9							
Very poor	46	460.3 (13.3)	77.6	453.7 (14.9)	88.1	453.1 (12.3)	73.6
Poor	101	480.4 (8.4)	76.3	482.5 (7.8)	75.3	505.1 (6.5)	67.8
Average	1,061	500.0 (5.7)	84.5	503.3 (6.1)	91.2	508.2 (4.4)	75.6
Good	1,233	547.5 (5.8)	96.9	546.6 (6.4)	104.3	532.3 (4.2)	81.1
Very good	905	593.0 (5.3)	89.3	597.0 (6.8)	103.0	559.8 (4.8)	83.0

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As can be seen, the mean scale scores for students with high self-ratings tend to be higher than the mean scale scores for students with low self-ratings. For example, for grade 4, the difference in average scale scores between students rating themselves as "very good" and "average" are 69, 60, and 37 for narrative, expository, and document, respectively. For grade 9, the corresponding figures are 93, 94, and 52.

To summarize the strength of the relationship between reading literacy scale scores and students' ratings of their reading literacy levels, the coefficient of determination (R^2) was computed (Table 10-5).

R^2 indicates the proportion of variance in each reading literacy scale score explained by students' ratings of their literacy levels. Thus, for grade 4, students' rating of their literacy levels accounted for about 8 to 9 percent of the variation in the narrative and expository scale scores, whereas the corresponding figure for document scales scores was about 3 percent. For grade 9, the corresponding figures are 15 percent, 13 percent, and 7 percent. The proportion of variance in narrative and expository scales accounted for by students' ratings is about the same (within grade) for both grades 4 and 9. Further, for both grades the proportion of variance in the document domain accounted for by the criterion variable is lower than the corresponding number for the other two domains. Across all domains, the proportion of variance accounted for is substantially larger for grade 9 than grade 4.

Table 10-5. Coefficient of determination (R^2) between reading literacy scale scores and students' ratings of their reading literacy levels: Grades 4 and 9

Domain	Grade 4	Grade 9
Narrative	0.082	0.152
Expository	0.088	0.132
Document	0.032	0.071

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Are these patterns of relationships meaningful? One interpretation of the differences across populations is that ninth grade students seem to have formulated a more realistic assessment of their reading literacy levels than fourth grade students. This is hardly surprising because older students not only have more evidence upon which to base their evaluations, but they also are likely to be more self-critical. The difference between the document and the narrative and expository domains can also be explained. It appears that students' self-ratings are more closely linked to their ability to comprehend narrative and expository texts than documents--hardly a surprising finding considering the relative emphasis in reading instruction given to the narrative and expository types of texts that pervade all their instruction and the rather limited exposure, if any, to documents within the school environment. Further, the difference may also be attributed to the qualities of the items. Students will have had a great deal of exposure to narrative and expository test items. In contrast, experience with the document test items, as well as documents in the form in which they appear in this test, is likely to be very limited.

10.3.3. Constructed-Response Items

As will be discussed in Section 10.4. on content validity, American reading experts regard reading as a process of "constructing meaning" from text, not simply "getting the meaning." In the U.S. this philosophical stance has resulted in an interest in having test items more closely model the desired underlying behavior. To accomplish this goal, major U.S. reading literacy tests are turning to performance-based items that require students to produce something--most frequently a written response. Although the IEA Reading Literacy Study scale scores are based on dichotomously scored item responses, predominantly of the multiple-choice variety, for each population two open-ended items were also included. These open-ended items, however, were not included in constructing the scale scores used in reporting results. A comparison of responses to the constructed response items and the scale scores can provide further criterion-related evidence of the validity of the reading literacy assessments because these measures can be viewed as another measure of reading literacy proficiency.

Table 10-6 presents mean scale scores for each IEA Reading Literacy Study domain by ratings of constructed responses. As these tables indicate, the mean scale scores for students with high scores on the constructed-response items are substantially higher than the mean scale scores for students with low scores on the constructed-response items. The pattern of increase in mean scale scores are similar across both constructed-response items for each population. Furthermore, the patterns of increase in mean scale scores are similar across the two populations (i.e., fourth and ninth grade students).

Table 10-6. Mean scale scores and standard deviations by ratings of constructed response items: Grades 4 and 9

Test passage	Item rating	Number of students	Narrative		Expository		Document	
			Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Grade 4								
Walrus	2 ...	435	466.1	85.7	474.2	61.7	489.2	68.3
	3 ...	1,266	510.7	85.6	505.1	66.6	521.5	69.3
	4 ...	2,139	566.7	85.7	548.7	68.8	559.8	74.7
	5 ...	2,032	603.7	78.9	581.5	68.9	583.7	76.8
Grandpa	2 ...	2,220	502.8	82.6	509.1	66.3	521.0	71.2
	3 ...	1,265	574.7	76.9	553.5	69.0	561.1	77.7
	4 ...	1,334	615.1	69.5	581.3	66.1	585.3	72.9
	5 ...	746	634.2	66.4	597.5	69.3	598.0	73
Grade 9								
Literacy	2	10	415.5	46.2	404.8	59.3	400.9	64.5
	3	120	448.5	74.7	445.7	90.2	447.9	73.8
	4	968	523.7	87.7	524.6	94.0	519.1	75.1
	5	1,847	576.6	89.9	579.9	99.7	553.2	77.5
Shark	1	156	452.5	94.7	449.0	94.6	468.6	81.0
	2	573	508.1	87.2	504.2	91.3	507.5	75.1
	3	1,089	549.3	91.6	547.9	101.7	532.8	78.0
	4	1,387	569.4	92.6	576.4	101.5	552.5	77.9

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

To summarize the strength of the relationship between reading literacy scale scores and students' responses to the open-ended items, the coefficient of determination (R^2) was computed (Table 10-7).

Table 10-7. Coefficient of determination (R^2) between reading literacy scale scores and students' ratings of constructed responses to two open-ended items: Grades 4 and 9

Domain	Grade 4		Grade 9	
	Walrus	Grandpa	Literacy	Shark
Narrative	0.211	0.330	0.128	0.104
Expository	0.206	0.214	0.118	0.109
Document	0.136	0.151	0.102	0.077

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As shown, the proportion of variance in reading literacy scale scores explained by students' responses to the open-ended items ranges from 14 percent to 33 percent for grade 4, and from 8 percent to 13 percent for grade 9. With the exception of the *Grandpa* passage for grade 4, the proportion of variance in narrative and expository scales accounted for by students' scores on open-ended items are about the same both for grades 4 and 9. Further, for both grades the proportion of variance in document scales accounted for by the criterion variable is lower than the corresponding number for the other two domains. Across all domains, the proportion of variance accounted for by the open-ended items is larger for grade 4 than grade 9, although it should be kept in mind that the passages were not the same for both grades.

While there may be some concern that constructed-response items are more a measure of writing ability than of reading ability, two cautions were taken in evaluating/constructing this comparison. First, the scoring guides for these constructed-response items stressed the content of the response in terms of evidence directly cited in the passage, rather than the quality or length of the written response. Second, further empirical tests were conducted of the relationship between length of response and qualities of writing with item scores.[1] Those findings indicate that these two attributes of writing are not strongly confounded with these scores, suggesting that they are operating as a measure of reading proficiency rather than writing proficiency.

The foregoing evidence has not challenged the assumption of validity of this test for the specific purpose of drawing inferences about subpopulation differences in reading literacy achievement and factors affecting reading literacy. Although the R^2's were moderate to low, the pattern of relationships was consistent across the variables and grades considered. These R^2's were moderate partly due to the presence of measurement error in the criterion scores. Although the evidence provided did not challenge the assumption of validity of this test for these purposes, nevertheless this evidence alone is not sufficient (Cronbach 1989). Therefore, we turn to other modes of collecting validation evidence.

10.4. Content-Related Evidence of Validity

According to Cronbach, content validity rests on the definition of the domain of interest to be measured and interpreted. In Chapter 7, we presented the IEA framework for the Reading Literacy Study, which included a definition of the domain and the specifications for the types of items to be included in the test. To explore the content validity, we have looked at three aspects of the IEA Reading Literacy Test: the test items as they relate to the test specifications, the test specifications as they compare to specifications of another important national test within the United States (NAEP), and the IEA test items as they relate to the NAEP test specifications. In the first instance, we wish to establish that the test has, in fact, tested what it purports to test, as perceived in the U.S. In the second and third instances we are interested in determining if the domain as defined would be so defined in the U.S.

In summary, these explorations represent an effort to answer three questions:

1. Would American reading specialists agree that these items tap the defined processes of reading comprehension?

---

[1] For further information, refer to the paper by Kapinus et al. in the separate *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study.*

2. Would American reading specialists agree that the IEA definition of reading literacy is consistent with the behavioral domain they call reading literacy?

3. Given the full range of content that American reading specialists would include in the domain of reading literacy, how representative of the domain are these items?

10.4.1. Would American Reading Specialists Agree that These Items Tap the Defined Processes of Reading Comprehension?

A subcommittee of members of the U.S. National Steering Committee (NSC) were asked to review the provided definitions of domains and skills assessed, to read the passages and the associated items, and to classify both passages and items according to the provided definitions. This review was done according to the specifications laid out by Warwick Elley, chairman of the International Steering Committee (ISC). Although having experts classify items may not necessarily prove their validity, this exercise produces an initial hypothesis regarding whether the test actually measures certain aspects of reading comprehension.

The definitions for the classification of items provided were very similar to those used to define the framework for the IEA Reading Literacy Test. The NSC subcommittee found that they could clearly discriminate among the definitions, that the definitions did capture some of the differences in reading processes, but that the definitions were somewhat arbitrary and broad as compared to those represented in the U.S. research literature. To accommodate these differences, as elaborated below, the NSC subcommittee tried to refine the definitions so that more refined distinctions within the defined processes could be captured in the classification system.

The NSC subcommittee proposed no change in the definitions of the three domains -- narrative, expository, and document. While these definitions do not integrate the dimension of purpose for reading, the differentiation of the three text types is straightforward and does coincide reasonably well with U.S. research literature (Meyer and Rice 1984). They also proposed no changes to three of the five categories of skills assessed, leaving them as written. Thus, **verbatim** items were said to require the student merely to match the words of the item with those of the text, **paraphrase** items were said to require the student to choose or compose an answer that is explicitly stated in the text but expressed in words different from that of the item, and **follow direction** items were said to require the student to follow the directions contained in a structured document.

With regard to **main theme** items, the NSC subcommittee thought that a major difference in complexity of processing would occur if the main idea had been explicitly stated or if a student had to create the generalization on his/her own. Therefore, instead of having just one category as defined in the specifications that characterize main theme items as those requiring the student to identify the underlying message of the text or some specified part of it, the NSC subcommittee created two categories, differentiating between instances where the main theme was explicitly stated and those where students were required to derive the main idea.

With regard to **inference** items, the NSC subcommittee thought that a major difference in complexity would occur if all the needed information appeared in the passage or if the respondent had to use additional information or apply something in a new context. Therefore, instead of having just one category as defined in the specifications, which stated that inference items require the student to draw a generalization from the text using information that is not explicitly stated in the text, the NSC subcommittee created two subcategories -- the first requiring no additional information, the second requiring something additional.

With regard to the category of **locating information in a document**, the provided definition requiring the student to search and find some specified information contained in a structured document was considered too global by the NSC subcommittee. Therefore, they divided the category into three subcategories -- the information is literally there, the information is there but the reader must in some way process it, and the reader must generate new information and then act based on the new information.

The theme underlying these changes relates mostly to conceptions of reading as an interactive process. It is a statement of the degree of input provided by the reader in constructing meaning. The skills assessed by the test, as defined by the ISC, seem to emphasize the text-embedded nature of reading, while the NSC definitions emphasize the reader input somewhat more.

Having reached consensus among the NSC subcommittee members regarding classification definitions, the next step was to have each subcommittee member classify items and to determine how consistent the categorization of skills would be across the raters. Table 10-8 summarizes the level of agreement among the U.S. raters and between the U.S. raters and international (ISC) ratings.

Using the U.S. classification scheme, agreement among all raters was very high for grade 4 test items (91 percent) and moderately high (82 percent) for grade 9 items. In view of the fact that a restrictive definition of agreement was used (i.e., consistent classification by all raters), inter-rater agreement using the U.S. classification scheme is impressive. The inter-rater agreement goes up even higher when a less restrictive definition of agreement is adopted. The average agreement between pairs of raters (averaged across all items and raters) is 97 percent and 94 percent for grades 4 and 9, respectively.

Table 10-8. Level of agreement between U.S. raters and international raters, by item classifications: Grades 4 and 9

Inter-rater agreement	Grade 4 (66 items)	Grade 9 (89 items)
Among U.S. raters		
All three raters agree	91%	82%
At least two raters agree	97%	94%
Between U.S. raters and ISC raters		
All three U.S. raters agree with ISC rating	80%	64%
At least two U.S. raters agree with ISC rating	86%	66%
At least one U.S. rater agrees with ISC rating	89%	74%
No U.S. rater agrees with ISC rating	12%	26%
Item type breakdown-at least two U.S. raters agree with ISC rating		
Verbatim	91% (11 items)	86% (7 items)
Paraphrase	75% (16 items)	53% (19 items)
Main theme	25% (4 items)	30% (10 items)
Inference	100% (12 items)	84% (19 items)
Locate information	96% (23 items)	100% (21 items)
Follow directions	(0 items)	23% (13 items)

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

On grade 4 items, the raters only disagreed on five items, splitting between paraphrase and verbatim on two, paraphrase and inference on one, verbatim and inference on one, and inference and main theme on the last. On grade 9 items, the raters disagreed on 16 items, splitting between paraphrase and verbatim 7 times, paraphrase and inference 6 times, level of inference on 2 items, and level of locating information on 1 item. In considering whether an item should be either verbatim or paraphrase, the argument centered on whether the words were explicitly there and close enough in the text to constitute a verbatim response. In considering whether an item should be either a paraphrase or an inference item, the question was how closely the text and item wordings were related. With regard to differences between main theme or inference items, the question was whether a main theme could be limited to just one or two sentences or whether it should be a measure of understanding a larger piece of text.

If the American raters had limited themselves to the international definitions, how close would they have come in classifying items to the ISC classifications? To this end, the NSC subcommittee rating categories were collapsed into the ISC categories. Using the ISC categories, the U.S. raters reached 85 percent and 68 percent agreement with the ISC raters on grades 4 and 9, respectively. The major difference centered on whether the nine items (three from grade 4 and six from grade 9) the U.S. raters classified as inferences should be main theme items as defined by the ISC. Here, as before, the difference rests on how much text one must process to arrive at a response. A similar level of disagreement surrounded the classification of nine items (from grade 9) as either locating information or following directions. The U.S. raters preferred to classify items as locating information if the information was already there, reserving the classification of following directions to items that required the respondent to produce a product. The U.S. raters disagreed with the ISC classifications of 13 paraphrase items, rating 7 of them as inference and 6 as verbatim.

A conclusion one might draw from these two exercises is that within the set of U.S. experts there is greater consensus, even when the level of discrimination to be made is finer, than there is between the U.S. experts and the International Steering Committee. Evidence to support this hypothesis could also be drawn from the comparison between the specifications of this test and NAEP. It would appear that perhaps the U.S. raters value something different or broader within the domain of reading literacy and that the degree of consensus among them represents a common understanding that may not be shared with their counterparts on the International Steering Committee. To test these hypotheses we turn attention to the NAEP 1992 reading specifications.

10.4.2. Would American Reading Specialists Agree that the IEA Definition of Reading Literacy Is Consistent with the Behavioral Domain They Would Call Reading Literacy?

Answering this question requires comparing the assessment framework of the IEA Reading Literacy Study to the NAEP 1992 Reading Assessment Framework and specifications. NAEP has been chosen as the focus of this comparison because of its prominent position in the U.S., as well as the consensus process that was used to develop this framework.[2]

The comparison rests on the definition of the domain and the specification of tasks which represent the depth and breadth of the domain. Therefore, we begin with a comparison of definitions.

---

[2] For a discussion of the NAEP consensus process, see "Reading Objectives: 1990 Assessment" (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, April 1989).

As previously cited, the IEA defined reading literacy as

> *...the ability to understand and use those written language forms required by society and/or valued by the individual. Literacy occurs in a variety of language contexts (e.g., school, home, work, and religious or civic institutions). Reading literacy involves both a range of competencies and a set of habits or practices, arrayed along various dimensions such as reading "stop" signs to being aware of nuances in complex philosophic texts; from reading only what is necessary to pursuing one's further learning and recreation through books and journals.*

In contrast, the Council of Chief State School Officers' (CCSSO) NAEP document (1990, 5-6) defines reading in the following way:

> *Reading involves an interaction among the reader, the text, and the context of the reading situation. The reader brings certain prior knowledge about the topic and task as well as a degree of understanding and skill in reading. Texts have different organizations and features that have an effect on how the reader uses them. The context of a reading situation includes the purposes for reading and cues that the reader might use in building a meaning of the text. Sometimes readers are comfortable and successful reading stories but are nonplussed by tax forms. Readers may have learned how to read and learn from textbooks but are less able to approach and appreciate a poem. Reading is a deep specific interaction between the reader, text, and situation. This interaction affects the way readers understand and respond to what is read, and readers are more or less proficient in using different types of texts and purposes for reading. The reader is oriented to a text very differently depending on the text itself and the purpose for reading.*

There is a degree of overlap between the two definitions of the domain. Both IEA and NAEP stress the interaction between text and reader to develop an understanding of the text and to use real information. Both recognize the range of contexts for reading. Both recognize a range of competencies, although IEA stresses processing while NAEP talks about coordination of prior topical and strategic knowledge. Both recognize a variety of purposes, but NAEP talks about purpose as a function of context. IEA is alone in stressing reading literacy as related to societal and individual values. NAEP emphasizes text features as being significant in the reading performance, as well as strategic control over reading.

Given the international nature of the IEA study, it is more than reasonable to expect that its definition of reading would account for differences across nations, societies, or cultures. Alternatively, the U.S. emphasis on text features is a relatively new research phenomenon concurrent with the growth in interest in text linguistics. The U.S. reading research literature begins to reflect this interest in text structure, register, genre, and text type in the early 1970's (Meyer and Rice 1984, 319-47). However, the first time it was reflected in a reading assessment was in the Young Adult Literacy Assessment in which document reading was specifically separated from prose reading. Similarly, the U.S. interest in metacognition as it relates to reading (strategically controlling one's reading processes) can be traced to Flavell's work in the early 1970's (Baker and Brown 1984, 353-94). The first measures of metacognition in reading appeared in state reading assessments in Illinois and Michigan during the 1980's. Consequently, it is logical that there be differences in the scope of the domain given the contexts in which the instruments were to be developed.

These differences in how the IEA and the U.S. NAEP consensus group defined reading literacy clearly have implications for specifying the tasks that are representative of the domain. An important consideration is how each group would define what types of reading matter to include in the domain. For IEA, in some respects this is an issue not only of which written forms are valued in the society, but also of how broadly the topics covered in the sum total of the curricula are to be represented in the assessment. Given the extremes of the continuum of texts that could represent all the written forms valued by the society, there are two possible ways of defining the range of sample texts -- inclusion of the broadest range of possible texts or inclusion of only those that are common to all societies.

One might reasonably argue that the IEA test should encompass an extremely wide range of reading materials to accommodate the wide variation in what might be valued across the 31 participating countries. This would be tantamount to a call for perhaps the fullest possible representation of what might appear any place in each of the participating nations. This would imply the inclusion of texts that might never be included in an American public school instructional program (e.g., the reading of the Bible, Torah, or Koran), thus going far beyond school-test curricula. If that were the case, then, out of necessity, there would also have to be some measure of relative importance within each country of that text in terms of topic, text type, and desired reader performance.

Alternatively, one could see a more narrowly defined range of texts that represent only the common core of materials that students in all nations are likely to encounter. The underlying logic of this approach rests on two principles. First, literacy would be viewed in a single unified international framework. Second, the test instruments would then appear to have equal validity across all nations. Given the objectives of the study as specified at the outset, this was the approach taken.

Within the U.S. the range of texts for inclusion is limited in different ways. NAEP specifications state that texts

> *...will be drawn from authentic texts occurring naturally in the environments of students at each grade level. Authentic texts are those that are actually found and used by readers in real, everyday reading* (CCSSO 1990, 19-20).

Two dimensions are stressed: texts must be intact, authentic pieces of writing, and the texts should be clearly related to the students' environment. But the differences in how the domain is defined become more apparent as one moves down a level from the global definition to the definition of the tasks that represent the domain. IEA outlines three subdomains, narrative, expository, and document, citing attributes of these text genres as the distinguishing feature (see page 7-2 for definitions). In contrast, NAEP defines each of the three text types more explicitly:

> ***Literary Text Specifications:*** Reading for literacy experience *involves reading types of literary text such as novels, short stories, poems, drama, etc., where the reader can explore the human condition and consider the interplay in the selection among emotions, events, and possibilities. ...In addition, the passages selected must be intact and must meet the following criteria for structure and cohesiveness, as indicated by the mapping of the passage or similar text analysis: well-developed plot, thematic focus, multi-dimensional characters. The passages must contain appropriate features of the genre (e.g., an unstated moral in a fable, problem/conflict/resolution in realistic fiction) and must have literary merit as evidenced in style, imagery, and development of theme.*

> ***Information Text Specifications.*** Reading to be informed *involves reading informative passages such as magazine and newspaper articles, textbooks, advice and editorial columns, encyclopedias, catalogues, and books on hobbies or interests in order to obtain some general or specific information. ...Information passages must be intact or "stand alone" segments, e.g., a section of a textbook or trade book chapter that does not require knowledge of preceding or following parts of the chapter. They should include graphs and tables where appropriate. Information passages must meet certain criteria for structure and cohesiveness as determined by concept mapping or similar text analysis. These criteria include clearly defined concepts; "considerate" text, i.e., not confusing in style or organization; clearly identifiable central purpose; ideas presented at different levels with effective subordination; and ideas that are clearly linked together. The passages must have structure consistent with the domain, e.g., history has causally-linked events (causal relationships) and chronological order; science has theory and evidence.*
>
> ***Document Text Specifications.*** Reading to perform a task *involves reading materials such as transportation schedules, directions, forms, recipes, voter registration materials, maps, referendums, advertisements, consumer warranties, and office memos for the purpose of applying the information in the document. Authentic text of this type must be complete documents or intact portions of document text that can be used without knowledge of the surrounding text. These should be genuine documents with a real function and not be simply generated for this assessment. Where a document involves more than one part, the major parts that are to be used together will be presented; e.g., a bus route map and accompanying time table, or directions and a diagram. Document text must have the stimulus set in a context that defines the reader's task and must be related to tasks that are valid for the grade level and experiences of the students being assessed* (CCSSO 1990, 3-7).

The differences in the definitions of the subdomains between the IEA and NAEP lead to differences in the potential passages chosen for inclusion in the instruments. While both sets of specifications call for three types of text to be included (narrative or literary, expository or information, and document), the sets of passages will, however, have some different characteristics. The NAEP passages will reflect materials actually read by students in the format that they actually appear, while IEA passages may be developed specifically for a testing situation. Given the potential artificiality of the IEA texts, one might wonder if some features of genre or register might not be lost.

An additional difference is the incorporation within the NAEP definition of these text types of differences in intended purpose. Here the semantic function of a text genre is specifically stated. Hence, the NAEP passages are used in the same manner that they actually would be used in reality. For example, the questions the readers are asked would be ones that readers are likely to have posed themselves. Alternatively, the questions might derive directly from the functional meaning of the text. By merging the text structure differentiations into types of actual reading materials, (e.g., newspapers, magazines, textbooks, novels, short stories, bus schedules, etc.), the NAEP documentation implies that assessment and instruction should be carried out on actual materials that serve a communicative function. This would lead to differences in how the materials looked on the page as well. The NAEP passages would be surrounded by appropriate accompanying artwork, title, author designations, and the like, where appropriate. The IEA passages would look like testing materials.

So far, we have established that there are differences in the types of texts to be sampled. The IEA sample would include texts that generally follow the specifications of three genre types, while the NAEP texts would be authentic pieces and would have to be used in a manner consistent with their intended purpose. Greater differences between the definitions of the domain become apparent when we look at the definitions of the reading processes to be sampled.

As described in Chapter 7, the IEA definitions call for six categories of processing. These categories relate closely to features of the text and rarely have the reader using what has been read in a new context. At its most simplistic level it is a test of recognition of what appears in the text itself. In contrast, the CCSSO's NAEP document (1990, 10-15) defines this dimension much differently:

> *Readers use a range of cognitive abilities that should be assessed within each of the reading-purpose scale areas....In understanding, they form an initial concept or image from the text, and they develop that understanding and fill it in. In extending and elaborating their understanding, they respond to the text personally and critically, in various ways and for various purposes. They take a critical stance on the text, judging its quality, or its logical plausibility.*
>
> *...These cognitive aspects of reading are not to be conceived of as a sequence or hierarchy (students might respond to a part of a text critically without necessarily developing their overall understanding). Further, while these abilities are related and somewhat interdependent, some situations do not require students to engage in all these responses to reading.*

It continues by addressing the aspects of building an understanding, elaborating, and critically responding to text:

> *Readers handle texts in a variety of ways as they use their background knowledge and information from the text in order to build an understanding, extend the meaning, and respond critically to the text. These specific behaviors, the various interactions between readers and texts, do not fall along a continuum or in a hierarchy. They should be in the repertoire of readers who are at every developmental level in reading...*
>
> <u>*Constructing meaning*</u> *is understanding what is read in a general manner. This concept is based on the recognition that reading is a process of constructing understanding on the part of the reader. It includes at least two kinds of general ability: forming an initial understanding and developing an interpretation...*
>
> *Forming an initial understanding requires the reader to provide an initial impression or global understanding of what was read. It involves considering the text as a whole or in a broad perspective...*
>
> *Developing an interpretation requires the reader to go beyond the initial impression to develop a more complete understanding of what was read. It involves linking information across parts of a text as well as focusing on specific information...*
>
> <u>*Elaborating and responding critically*</u> *involves shifting, consciously or unconsciously, to analytic reading. Here, readers try to extend or examine the meaning of the text. It involves applying and judging the information or ideas from the text. There are two broad categories of tasks in this type of reading: personal reflection and response and taking a critical stance.*

> *Personal reflection and response require the reader to connect knowledge from the text with his or her own personal background knowledge. The focus here is on how the text relates to personal knowledge...*
>
> *Demonstrating a critical stance requires the reader to stand apart from the text and consider it objectively. It involves a range of tasks including such behaviors as critical evaluation, comparing and contrasting, application to practical tasks, and understanding the impact of such text features as irony, humor, and organization.*

This interaction between text type/purpose and range of cognitive abilities is summarized in the following matrix, taken from the Reading Framework report.

The NAEP Matrix of Reading

	Literary Experience	To Be Informed	To Perform a Task
Constructing Meaning Initial Understanding Developing an Interpretation			
Elaborating/Extending Personal Response Critical Stance			

The strength of the NAEP framework and test specifications within the United States rests on the consistent logic that is inherent throughout the document and that represents a consistent theory or model of reading as a process and as an inclusive instructional model. As the consensus group noted,

> *The field of reading education is characterized by some of the deepest and longest-running schisms in education. ...In reading, this rift pertains both to one's concept of the nature of reading and to one's view of how it should be taught. A framework or set of objectives for the reading assessment had to be developed which mediated and transcended philosophical and theoretical differences about the nature of reading, and the framework had to provide instructionally useful information that somehow addressed or transcended differences over how reading should be taught.*
>
> *Our goal instead was to produce a framework which was built on and consistent with sound, contemporary research, no matter what its topic or orientation....The framework recommended for the reading assessment was inclusive, and it was deliberately based on a wide range of sound research bases.*
>
> *The implication of this is that assessment planning can transcend the presence of competing research orientations or different models of the subject being assessed. This can be accomplished by being inclusive and insisting on the principle that valid research be accounted for, no matter what its orientation* (CCSSO 1990, 6-8).

The essential element of this operating principle is a well-developed inclusive model of reading that focuses on skilled outcomes. To accommodate competing research orientations, the assessment contains measures that address the underlying processes necessary for reading comprehension--addressing those agreed-upon common threads--but at the same time, the assessment is respectful of the principles inherent in the competing research orientations. For example, while all researchers would recognize the need for fluency, an attempt was made to develop a measure that did not violate the principles of those researchers and practitioners who would object to word recognition in isolation. At each juncture, the underlying principle was identified and respected.

While the IEA framework has an implicit model underlying its design, given the IEA organizational structure and mechanisms for gaining consensus across nations, this model is never fully specified. To a degree this lack of full specificity during the test development phase is a function of the IEA process that allows for modifications of the instruments by virtue of the input of National Research Coordinators (NRCs). Each NRC supports a particular national perspective and helps to develop that line of thought. Additionally, within the Reading Literacy Study, with 32 participating countries, the test specifications must accommodate an even broader range of variation in theories than would be accommodated within a single country, even one as diverse as the United States.

Given the complexity of reaching agreement on a model of reading theory that encompasses the differences across more than 30 countries, one must ask whether the IEA model adequately reflects the U.S. national model, as reflected in NAEP. Considering the overlap between the dimensions of the NAEP and the IEA specifications, the IEA model, while not as inclusive as the NAEP model, does contain many of the same elements. There are separate measures of reading of three text types, there is recognition of the context of reading as a determiner of purpose, and there is also a measure of fluency. However, as demonstrated in the foregoing discussion of reading processes as defined by IEA and NAEP, there are other aspects of the specifications that would not reflect current practice within the U.S.

Based on the comparison between the IEA and NAEP specifications, American researchers (and practitioners) are likely to argue that the domain of reading literacy is much larger than that defined by the IEA. Therefore, the use of the identical term, "reading literacy," could lead to false interpretations of the test scores. This brings up another question. If the IEA Reading Literacy Assessment is not as inclusive as the NAEP, how much of the NAEP domain does the IEA test tap?

10.4.3. Given the Full Range of Content that American Reading Specialists Would Include in the Domain of Reading Literacy, How Representative of the U.S. Defined Domain Are the IEA Reading Literacy Items?

Members of the NAEP item writing team were asked to review the passages and items of the IEA Reading Literacy Test and to classify the items according to the NAEP specifications. This group of people were chosen as expert raters because they were fully knowledgeable about the NAEP specifications and consequently would require little training to reach a high level of agreement on the attributes of items. (In fact, the inter-rater agreement on this exercise was 97 percent for grade 4 and 95 percent for grade 9.)

This committee received the items and directions 4 days prior to meeting. They were asked to classify the items independently, and they met to review their responses and to discuss the differences that arose.

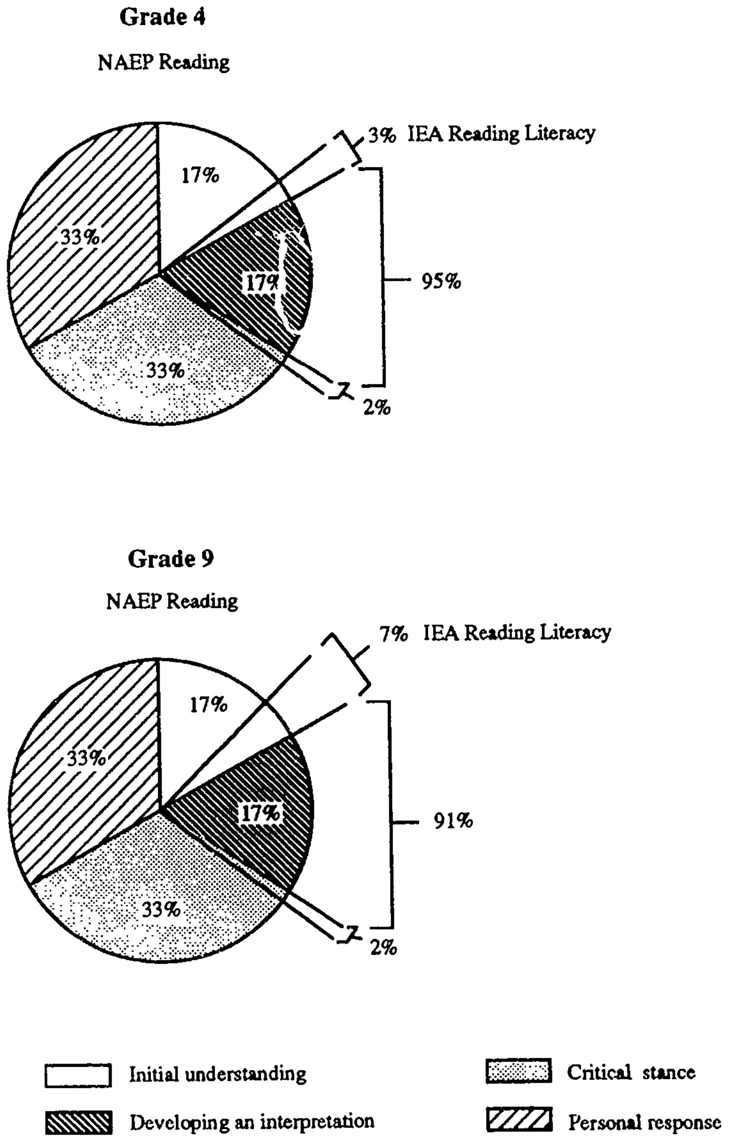
As shown in Table 10-9, when the IEA items are classified according to the NAEP categories, there is an imbalance in sampling across the domain. The IEA items are very heavily grouped in just one of the NAEP categories--developing an interpretation. The imbalance is also obvious when comparing the IEA Reading Literacy item distributions to the NAEP item distributions (Figure 10-7).

Table 10-9. IEA items classified by NAEP categories: Grades 4 and 9

NAEP category	Grade 4		Grade 9	
	Number of items	Percentage of total	Number of items	Percentage of total
Initial understanding	2	3%	6	7%
Developing an interpretation	63	95%	81	91%
Personal response	0	0%	0	0%
Critical stance	1	2%	2	2%

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 10-7. Comparison of IEA and NAEP item distributions: Grades 4 and 9



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The raters were then asked to consider which of the passages and item prompts were likely to be included in a NAEP assessment. In grade 4, they eliminated 9 of the 15 texts, with two others rated as suitable but probably too short. All six of the document texts were eliminated because they were not set in a functional context portraying their appropriate use. Of the five texts classified under reading for information, three were eliminated (*walrus, postcard*, and *quicksand*), a fourth was considered passable but probably too short (*marmots*), and the last (*trees*) was considered for possible inclusion. Those four passages classified as reading for a literary experience fared much better. All of them were considered possible for inclusion, although one (*bird and elephant*) was probably too short.

In grade 9, as in grade 4, all nine of the document texts were eliminated, not only because they were not contextually situated, but also because the questions were not those one would ask in using the documents. One document, *aspirol*, might have been considered appropriate for the grade 4 test. Under the classification of reading to be informed, three of the five texts were eliminated: one, *marmots*, because it was too simple for the population; the other two because of the length and quality of writing. The literary texts fared better here, too. Three of the five texts would not have been included because they were considered to be too simple for grade 9. However, the three would have been considered appropriate for grade 4.

When asked to consider which of the items associated with the acceptable passages were likely to be considered for a NAEP assessment, the raters' first global statement was that very few of the prompts would appear with the same distractors and that many would have appeared as open-ended items instead. For grade 4, only 12 items would have been retained out of 66, 11 of them developing an interpretation and 1 a critical stance. For grade 9, only 11 items would have been retained out of 89, 8 of them developing an interpretation, 2 initial understanding, and 1 a critical response.

In terms of comparing interpretations of the IEA Reading Literacy assessment to NAEP and of the generally accepted U.S. definition of the domain of reading literacy, one must keep in mind that this test does not represent the entire domain of reading literacy as generally defined in the U.S. It is a test of a limited range of the tasks within the U.S. definition of the domain, representing less than one-third of the overall domain defined in NAEP. But even among those tasks, only a small subset would meet the specifications of NAEP. Therefore, it is essential that the two tests not be interpreted as if they were measuring the same thing. While they both test aspects of reading literacy, it is clear that the IEA Reading Literacy assessment taps a narrow range of what is included in the NAEP assessment.

10.5. Construct-Related Evidence for Validity

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association et al. 1985, 9-10), "the evidence classed in the construct-related category focuses primarily on the test score as a measure of the psychological characteristic of interest....the construct of interest for a particular test should be embedded in a conceptual framework, no matter how imperfect that framework may be. The conceptual framework specifies the meaning of the construct, distinguishes it from other constructs, and indicates how measures of the construct should relate to other variables."

The IEA Reading Literacy Study assumed three constructs of reading literacy. These were associated with the differences in the three specified text types, narrative, expository and document. The NAEP 1992 consensus moved one step beyond by differentiating the three text types in association with three different purposes for reading (CCSSO 1990). The statement of purpose more tightly frames the types of questions to be asked in relation to a particular type of text. Therefore, the intersection of text type and purpose may more accurately represent the differences among the genre and may be more likely to yield three separate constructs than if processes were classified either by text type or purpose alone. If the IEA operationalized their framework so that the questions associated with particular passages reflected the differences associated with purpose for reading, three separate constructs would likely be seen.

As attention is turned toward establishing construct-related evidence for the validity of score interpretation in the U.S., it is desirable to establish whether three separate constructs were measured and whether these measures are reliable. To evaluate these questions, the inquiry is organized into the following four tasks:

- Establishing whether there is sufficient internal consistency across the items within each construct as defined by the IEA, so that each could be recognized as a single trait;
- Establishing whether each construct is unique and separate from the other two;
- Examining how well the distribution of item difficulty levels matches examinees' ability levels and thereby establish some estimate of reliability; and
- Examining the relationship between these constructs and other constructs to determine whether the observed pattern of relationships are consistent with existing theoretical frameworks.

10.5.1. Evidence Relating to Internal Consistency of Tests

In order to establish whether there is sufficient internal consistency across the items within each construct as defined by the IEA so that each can be recognized as a single trait, one must focus on evidence for the unidimensionality of the tests. This perspective will not only provide measures of consistency across items but will help to determine whether one common factor accounts for the observed covariation among the items within each construct. From the variety of methods available for assessing the unidimensionality of responses to a set of items, four types of evidence were evaluated: reliability, principal components, factor analysis, and latent trait models.

Reliability. Coefficient alpha, the internal consistency index, has been widely used to assess unidimensionality (Hattie 1985). Cronbach (1951) has shown that coefficient alpha is a lower bound to the proportion of test variance attributable to common factors among the test items. Therefore, we have calculated the coefficient alpha for each of the domains/scales. Table 10-10 presents the coefficient alpha and number of items for each domain.

Table 10-10. Coefficient alpha and number of items for each domain: Grades 4 and 9

Domain	Grade 4			Grade 9		
	Number of items	Alpha		Number of items	Alpha	
		Observed	Adjusted		Observed	Adjusted
Narrative	20	0.857	0.906	26	0.875	0.896
Expository	19	0.766	0.820	24	0.846	0.875
Document	21	0.733	0.807	32	0.791	0.791

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Based on coefficient alpha, it appears that for both grades narrative items are more homogeneous than expository items, and, in turn, expository items are more homogeneous than document items. Since the number of test items varies by domain, one cannot readily compare the reliability coefficients across domains or grades. To facilitate such comparisons, using the Spearman-Brown prophecy formula, we have estimated coefficent alpha for a test with 32 items--the largest number of items for a given domain. These adjusted estimated coefficients also indicate that for both grades narrative items are the most homogeneous, followed by expository items, while the document items are the least homogeneous.

What can we say about the observed coefficient alpha? If we adopt the rule of thumb that coefficient alpha greater than 0.80 is "high," between 0.75 and 0.80 is "moderate," and lower than 0.75 is "low," it can be concluded that (a) coefficient alpha for narrative items (both grades) and expository items (grade 9) is high; (b) coefficient alpha for expository items (grade 4) and document items (grade 9) is moderate; and (c) coefficient alpha for document items (grade 4) is low.

Green et al. (1977) have noted that a high coefficient alpha does not necessarily mean that a general factor is present, since high alpha can be obtained even though a general factor does not exist. As noted above, coefficient alpha is also dependent on the number of items. To overcome these limitations of coefficient alpha, Armore (1974) has suggested that item intercorrelations should be examined. Patterns of low or negative correlations among items can provide additional information regarding unidimensionality. For grade 4, the intercorrelations among narrative items were typically around 0.20, with the lowest correlation at 0.121; for expository items the intercorrelations were typically in the teens, with the lowest at 0.019; and for document items the intercorrelations were typically around 0.10, with some correlations as low as 0.04. For grade 9, the correlations among both narrative and expository items were typically in the teens, with the lowest correlation at 0.03; for document items the correlations were typically around 0.10, with some correlations as low as 0.01. This difference in pattern of correlations among test items is further evidence that document items are less homogeneous than narrative items.

Correlation of test items with total test score provides additional information concerning the unidimensionality of the IEA Reading Literacy Test data. The point-biserial correlations between the total test score (continuous variable) and a dichotomous item scores are presented for each domain in Tables 10-11 and 10-12.

These tables indicate that the correlation between test items and domain scores are generally high. With the exception of three items for grade 4 and eight items for grade 9 (seven of which are from the document domain), all items have correlations that are higher than 0.30, which is generally considered acceptable. In each of the exceptions, some type of ambiguity may account for this low correlation with the domain scores. For example, the two expository items in grade 4 that had correlations less than 0.30 were both associated with the same passage. Reading specialists in the U.S. had difficulty in determining whether this passage should be classified as a document or as an expository text. In the case of the two specific items in question, there was some question as to whether the reader had to process text or understand the format of a postcard and correctly identify the answer based on its position in the address. In the other two cases, one might attribute the low correlation to problems in the test item construction. In the case of *island*, the question is somewhat vague. In the case of *fox*, one of the distractors could easily be considered a correct answer.

In summary, the indices based on reliability of the domains seem to support the assumption of unidimensionality of the IEA Reading Literacy Test scores. There is, however, some question about the tenability of this assumption for the document domain for grade 9.

Table 10-11. Grade 4 item correlations with domain total score

Item	Point-biserial correlation	Item	Point-biserial correlation
NARRATIVE		EXPOSITORY (continued)	
Bird1	0.579	Marmot1	0.493
Bird2	0.424	Marmot2	0.427
Bird3	0.514	Marmot3	0.420
Bird4	0.569	Marmot4	0.454
Bird5	0.428	Trees1	0.570
Dog2	0.505	Trees2	0.447
Dog3	0.501	Trees3	0.516
Dog4	0.543	Trees4	0.387
Dog5	0.537	Trees5	0.476
Dog6	0.546	DOCUMENT	
Shark1	0.522		
Shark2	0.464	Island1	0.311
Shark3	0.514	Island2	0.246
Shark4	0.539	Island4	0.325
Shark5	0.564	Maria1	0.335
Grandpa1	0.529	Maria2	0.428
Grandpa3	0.585	Maria3	0.421
Grandpa4	0.529	Bottle1	0.318
Grandpa5	0.599	Bottle2	0.515
Grandpa6	0.456	Bottle3	0.325
		Bottle4	0.376
EXPOSITORY		Bus1	0.443
Card1	0.281	Bus2	0.549
Card2	0.187	Bus3	0.464
Walrus1	0.368	Bus4	0.472
Walrus2	0.400	Content1	0.302
Walrus3	0.524	Content3	0.379
Walrus4	0.508	Temp1	0.415
Walrus5	0.460	Temp2	0.414
Walrus6	0.494	Temp3	0.439
Sand2	0.489	Temp4	0.430
Sand3	0.408	Temp5	0.465

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 10-12. Grade 9 item correlations with domain total score

Item	Point-biserial correlation
NARRATIVE	
Fox2	0.364
Fox3	0.431
Fox4	0.308
Fox5	0.190
Mute1	0.496
Mute2	0.550
Mute3	0.574
Mute4	0.370
Mute5	0.537
Shark2	0.336
Shark3	0.401
Shark4	0.372
Shark5	0.412
Reveng1	0.624
Reveng2	0.461
Reveng3	0.561
Reveng4	0.557
Reveng5	0.525
Reveng6	0.497
Reveng7	0.560
Angel1	0.572
Angel2	0.717
Angel3	0.592
Angel5	0.583
Angel6	0.602
Angel7	0.554
EXPOSITORY	
Marmot1	0 409
Marmot2	0.434
Marmot3	0.343
Marmot4	0.429
Laser1	0.350
Laser2	0.522
Laser3	0.441
Laser4	0.490
Laser5	0.474
Laser6	0.554
Liter1	0.504
Liter3	0.479
Liter4	0.563
Parac1	0.436
Parac2	0.332
Parac3	0.351
Parac5	0.382
EXPOSITORY (continued)	
Parac6	0.459
Smoke1	0.488
Smoke2	0.507
Smoke3	0.563
Smoke4	0.488
Smoke5	0.565
Smoke6	0.503
DOCUMENT	
Card1	0.215
Card3	0.165
Card4	0.167
Card5	0.323
Card6	0.330
Card7	0.176
Resourc1	0.343
Resourc2	0.483
Resourc3	0.464
Job1	0.377
Job2	0.344
Lynx1	0.254
Lynx2	0.401
Lynx3	0.363
Bus1	0.359
Bus2	0.428
Bus3	0.432
Direct1	0.438
Direct2	0.505
Direct3	0.465
Weather1	0.407
Weather2	0.329
Weather3	0.436
Weather4	0.412
Temp1	0.281
Temp2	0.422
Temp3	0.366
Temp4	0.357
Temp5	0.290
Aspirol1	0.384
Aspirol2	0.396
Aspirol3	0.502

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Principal Components. Principal component analysis (PCA) and factor analysis (FA) have been traditionally used to investigate the dimensionality of responses to a set of items. Principal component analysis is a method of transforming a given set of variables into a new set of composite variables (principal components) such that the composite(s) extracts maximum variance from the original set of variables. The first principal component may be viewed as the single best summary of linear relationships exhibited in the data. Since the first principal component explains the maximum variance, then this variance, expressed as the percentage of total variance, has been used as an index of unidimensionality. "The implication is that the larger the amount of variance explained by the first component the closer the set of items is to being unidimensional" (Hattie 1985, 146).

The eigenvalues and percent variance explained by the first three principal components for the IEA Reading Literacy Test data are shown in Table 10-13. The eigenvalues represent the amount of total variance in the data any given factor explains. Thus, the total variance accounted for by the first principal component is 5.523 for grade 4 narrative items. The percent variance represents the proportion of the total variance explained by a given factor, and the variance explained by the first principal component is highest for the narrative domain (i.e., 28 percent) and lowest for the document domain (i.e., 17 percent).

Table 10-13. Eigenvalue and percent variance explained by the first three principal components (PC) for each reading literacy domain: Grades 4 and 9

Domain	Grade 4		Grade 9	
	Eigenvalue	Percent variance	Eigenvalue	Percent variance
Narrative				
1st PC	5.523	27.6	6.141	24.6
2nd PC	1.211	6.1	1.843	7.4
3rd PC	0.992	4.9	1.456	5.8
Expository				
1st PC	3.865	20.3	5.244	21.9
2nd PC	1.520	8.0	1.298	5.4
3rd PC	1.233	6.5	1.179	4.9
Document				
1st PC	3.567	17.0	4.504	14.5
2nd PC	1.455	6.9	1.439	4.6
3rd PC	1.170	5.6	1.393	4.5

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The question may be raised, "How high should the variance explained by the first principal component be to indicate unidimensionality?" Reckase (1979) recommended that the first component should account for at least 20 percent of the variance. Thus, the narrative and especially the expository items just meet the "rule-of-thumb" suggested by Reckase, whereas the document items fall a little short of the 20 percent criterion.

The sum of squared residual correlations, after removing the first component, has also been used as an index of unidimensionality. If the one-component model fits the data well, the residual correlations (i.e., the difference between observed correlations and correlations implied by the model) would be small. The root mean squares of the residual correlations are 0.0514, 0.0681, and 0.0586 for grade 4 narrative, expository, and document items, respectively. For grade 4 this index shows that the assumption of unidimensionality does not seem to be violated for any of the three domains. For grade 9, the root mean squares of the residual correlations are 0.0679, 0.0488, and 0.0468 for narrative, expository, and document, respectively. Based on this index, the assumption of unidimensionality does not seem to be violated for any of the three domains for either grade.

Factor Analysis. It was stated earlier that factor analysis (FA) has been traditionally used to investigate the dimensionality of responses to a set of items. Linear factor analysis of dichotomously scored items in general does not produce satisfactory results (see Carrol 1945; Drasgow and Lissak 1983). "In applying a linear factor analysis model, we are hypothesizing that dichotomous variables are linear combinations of continuous latent variables with infinite range, a mathematical impossibility" (Zwick 1987, 246-47).

Two promising alternatives to the conventional factor analysis are factor analysis of item parcels (Cook and Eignor 1984) and full-information factor analysis (Bock and Aitkin 1981; Bock, Gibbons, and Muraki 1985). Factor analysis of item parcels was achieved by grouping items relating to the same passage in one subtest and then applying conventional factor analysis to the subtest scores. Table 10-14 presents the results of the factor analysis on parcels for grades 4 and 9.

Table 10-14. First factor statistics based on parcels: Grades 4 and 9

Domain	Number of parcels	Eigenvalue	Percent variance	RMS residual correlations
Grade 4				
Narrative	4	1.945	48.6	0.032
Expository	5	1.304	26.1	0.044
Document	6	1.496	24.9	0.027
Grade 9				
Narrative	5	2.589	51.8	0.126
Expository	5	2.603	52.1	0.121
Document	7	2.605	37.2	0.104

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

These results support the unidimensionality assumption because the percentage of variance attributed to the first factor exceeds the 20 percent criterion; root mean squares of the residual correlations are generally small; and with the exception of *card* (grade 4 expository), all parcels showed high loadings on the first factor.

Bock and Aitkin (1981) developed a method of factor analysis, based directly on item response theory, that does not require estimation of inter-item correlation coefficients. "Because the Bock-Aitkin approach uses as data the frequencies of all distinct item response vectors, it is called 'full-information' item factor analysis" (Bock, Gibbons, and Muraki 1988, 262).

The full-information item factor analysis was implemented using the TESTFACT computer program (Wilson, Wood, and Gibbons 1991). The program requires as input the fixed values of the c-parameter in the three-parameter IRT model. For analyses conducted on the IEA Reading Literacy Study data, by fixing the c-parameter to zero, the two-parameter IRT model was the underlying procedure.

Additionally, items that examinees omitted at the end of the test were scored as "incorrect." Table 10-15 presents the results of the TESTFACT program for each of the reading literacy domains for grades 4 and 9.

Table 10-15. Full information item factor analysis (two-factor solution): Grades 4 and 9

Domain	Factor number	Percent variance	Chi-square change	Degrees of freedom
Grade 4				
Narrative	1	36.1	-	-
	2	3.9	207.9	19
Expository	1	28.6	-	-
	2	6.1	1,018.9	18
Document	1	25.8	-	-
	2	4.6	267.4	20
Grade 9				
Narrative	1	32.5		
	2	7.6	1,819.0	25
Expository	1	28.4		
	2	4.4	157.5	23
Document	1	21.1		
	2	5.2	181.5	31

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

For narrative items, the first factor extracted about 36 percent of the total variance for grade 4 and 33 percent for grade 9. In the unrotated two-factor solution, the second factor accounted for only 4 percent of the total variance for grade 4 and 8 percent of the total variance for grade 9. The chi-square change (for the improvement in fit by adding a second factor) was 208 (d.f. = 19) for grade 4 and 1,819 (d.f.= 25) for grade 9. These values of chi-square need to be evaluated keeping in mind the large sample sizes (in excess of 6,000 for grade 4 and 3,000 for grade 9) and the design effect of around 6 and 8 for grades 4 and 9, respectively. This means that modest variance percentages will appear to be highly significant, and the significance levels are in fact somewhat overstated. Thus, upon close examination of the results, the conclusion that a dominant factor runs through the items within each domain can be supported.

Item Response Theory. The IEA Reading Literacy Test data, which consisted of dichotomously scored item responses, were scaled using the Rasch model (one-parameter IRT). "One of the major advantages of the Rasch model often cited is that there are many indices of how adequately the data 'fit' the model" (Hattie 1985, 152). Wright and Panchapakesan (1969, 25) asserted that "if a given set of items fit the (Rasch) model this is evidence that they refer to a unidimensional ability, that they form a conformable set." Thus, one of the most useful tests of the unidimensionality assumption in the context of the Rasch model is the test of fit to the model that is part of the calibration process. Specifically, item-fit statistics provided as part of the calibration can be useful in this regard.

The BIGSCALE computer program, which was used to calibrate item difficulties and student abilities for the IEA Reading Literacy Test data, provides two types of item-fit statistics: INFIT and OUTFIT statistics (see Tables 8-1 and 8-2).

In summary, the variety of evidence examining the internal consistency of the Reading Literacy Tests indicate that the items measuring each domain hang together. However, more information is needed to evaluate the construct validity of the Reading Literacy Tests.

10.5.2. Evidence Relating to the Match Between Student Abilities and Item Difficulties

One advantage of the Rasch model is that item difficulty and examinees' abilities are calibrated on a common scale. Based on pilot test results (or previous information), educators can enhance test validity by selecting items that match the abilities of the examinees for whom the test is designed. This increase in test validity is achieved by matching item difficulties and student abilities. If this can be accomplished the information derived from the test, given that other factors (e.g., number of test items) are constant, is increased. From another perspective, by matching student abilities and item difficulties, the standard error of measurement is generally decreased, thereby increasing the reliability of the test. Since reliability establishes the upper bound of validity, by matching item difficulty levels to examinees' ability levels, test validity is generally increased.

For the IEA Reading Literacy Tests, items were selected taking into consideration students in all participating countries (i.e., the international group). Thus, while the IEA Reading Literacy Tests may be optimal for the international group, they may not necessarily be optimal for the U.S. students. For example, if the abilities of the U.S. students are significantly different from the abilities of the international group, one can argue that if the test is optimal for the international group, it would not be optimal for the U.S. students. Consequently, to the degree that the abilities of U.S. students are different from those of the students from other countries, to that same degree the test would be less optimal for the U.S. students.

Figures 10-8 to 10-13 provide the Rasch output indicating the distribution of item difficulties (calibrated on the international group) and abilities of U.S. students and international students placed in juxtaposition for each domain for grades 4 and 9. The distributions of student scores on the Rasch logit scale are shown both for the U.S. students for a given grade and scale and for the international data set (in which each participating country is equally represented). The item difficulties for the assessment items are also presented on the logit scale. Presenting these data together enables us to consider the difficulty of the assessment instrument for U.S. students, for each grade and scale, and to compare this with the difficulty in the aggregate of the participating countries. An examination of these figures reveals the following observations:

- For the U.S., the difficulties of the items do not optimally match the abilities of the students. In particular, items that would provide useful information for examinees of high abilities have not been included in the tests.
- For the international group, similar to the U.S. group, there also seems to be a mismatch between examinees' abilities and item difficulties. However, the degree of mismatch (between student abilities and item difficulties) is less for the international group than for the U.S. students.
- The abilities of students at the low to middle range of the reading literacy scales will generally be estimated more reliably than the ability of students at the higher levels of reading literacy scales, both internationally and especially for the U.S.

Figure 10-8. Map of student proficiencies and item difficulties for Rasch analysis: Grade 4 narrative domain

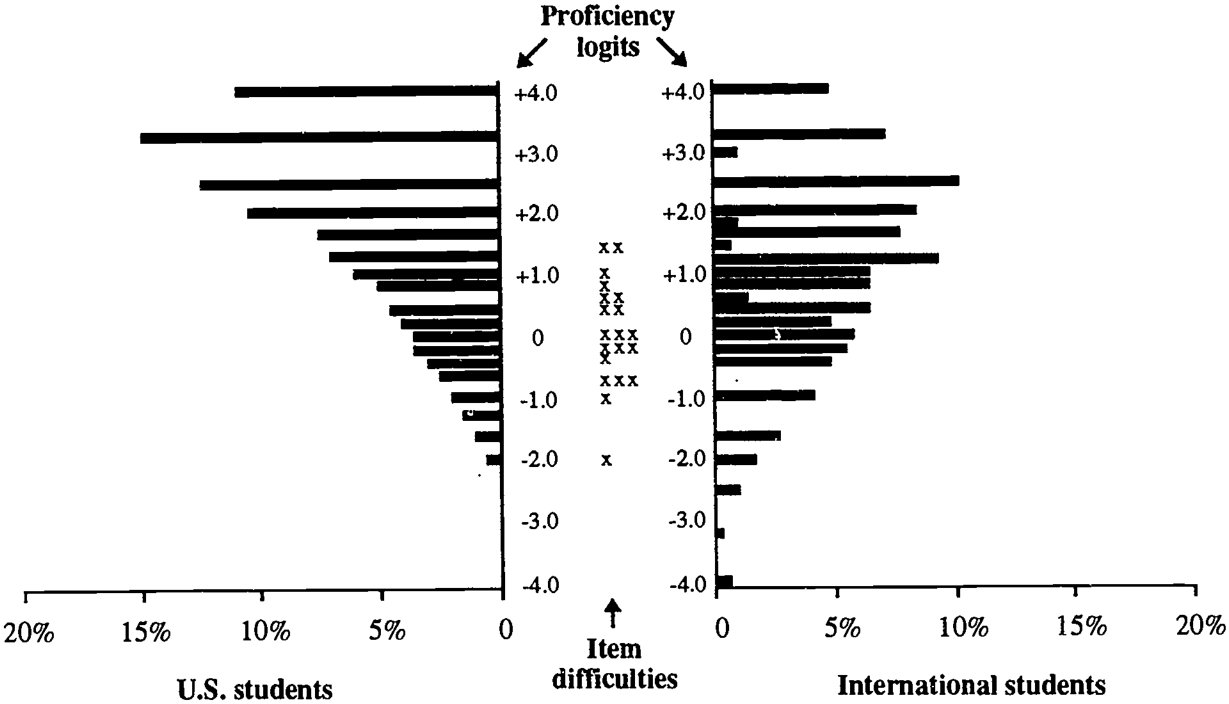
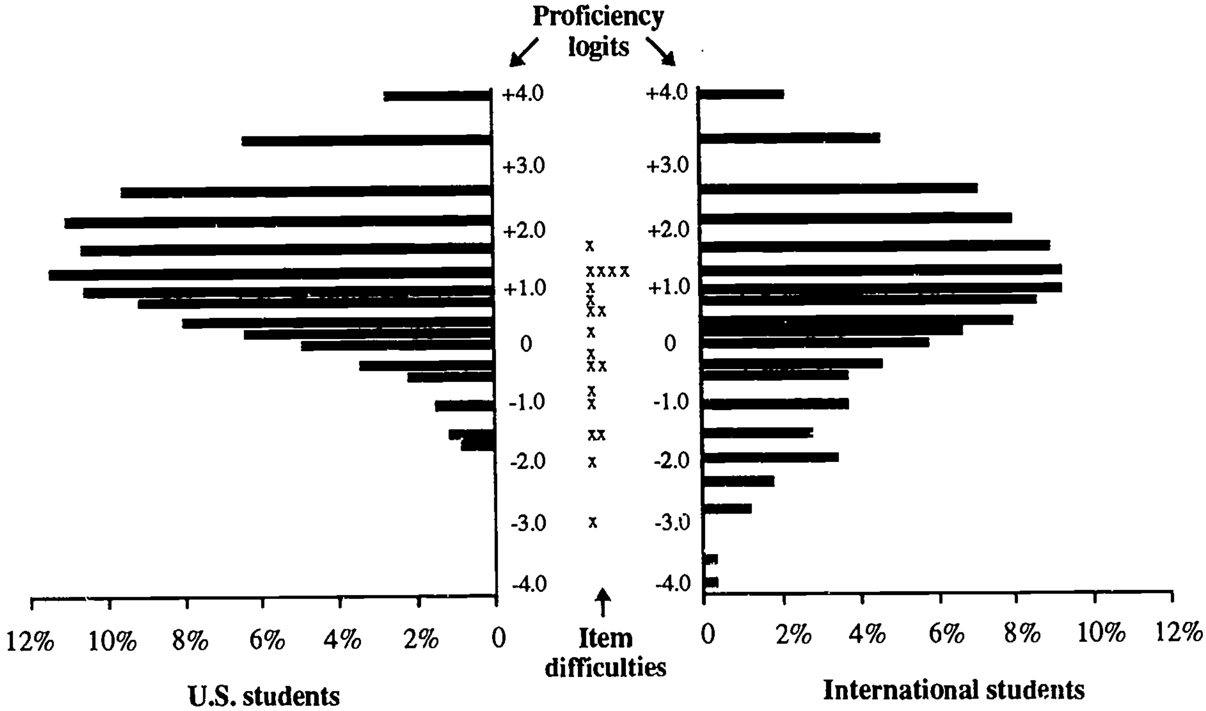


Figure 10-9. Map of student proficiencies and item difficulties for Rasch analysis: Grade 4 expository domain



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 10-10. Map of student proficiencies and item difficulties for Rasch analysis: Grade 4 document domain

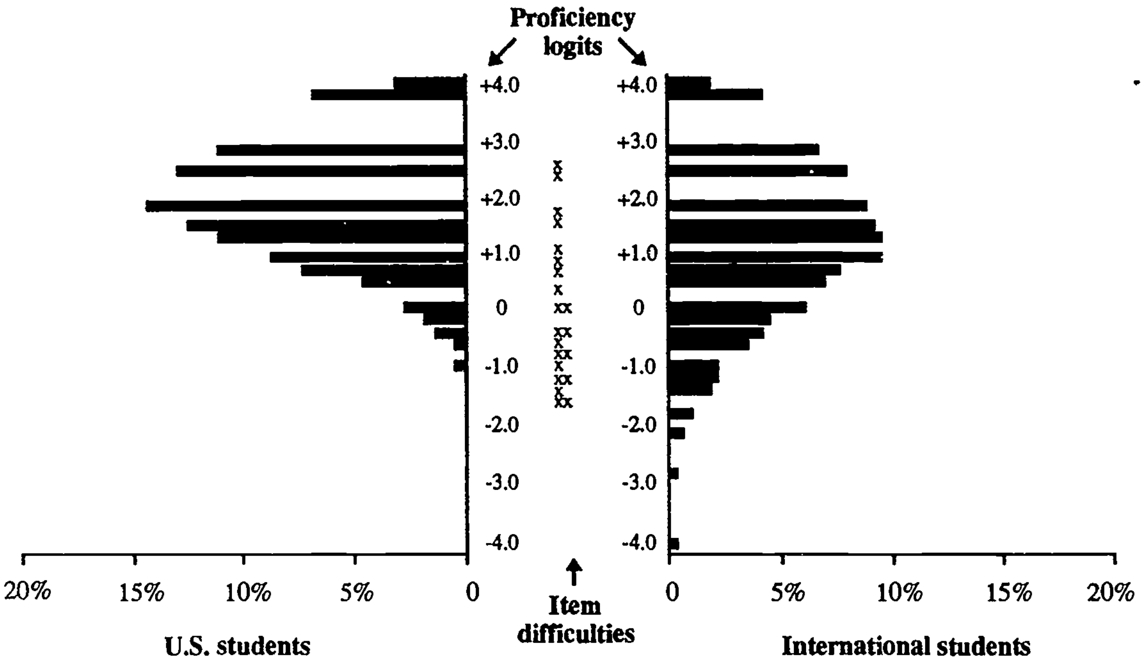
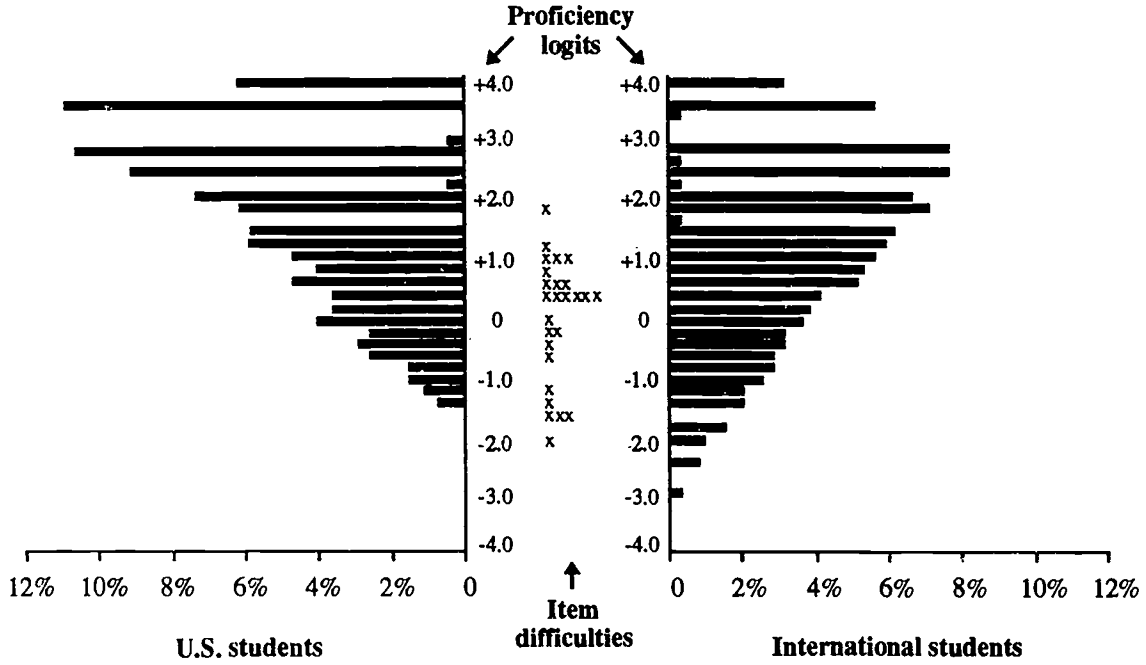


Figure 10-11. Map of student proficiencies and item difficulties for Rasch analysis: Grade 9 narrative domain



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 10-12. Map of student proficiencies and item difficulties for Rasch analysis: Grade 9 expository domain

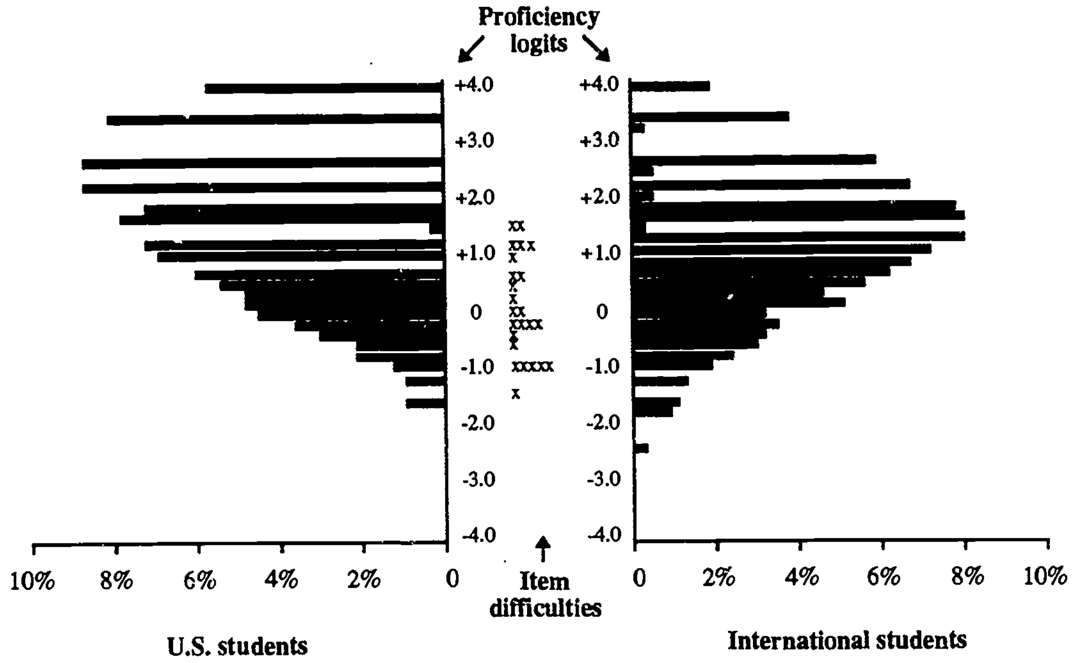
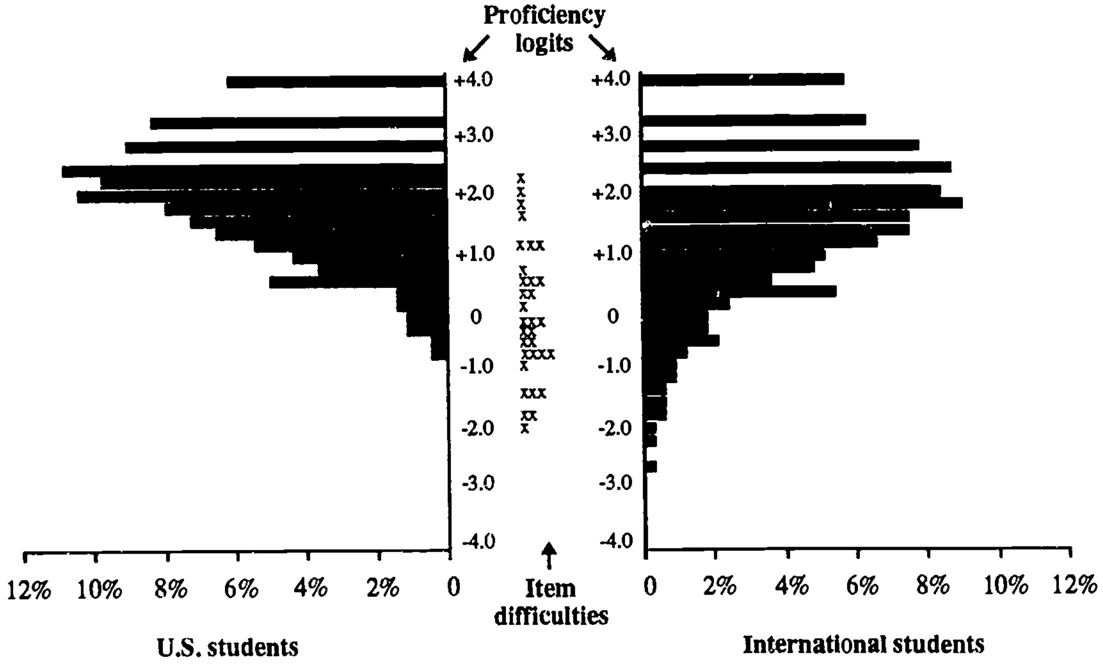


Figure 10-13. Map of student proficiencies and item difficulties for Rasch analysis: Grade 9 document domain



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

10.5.3. Correlation Among Scale Scores

The correlation among scale scores is another source of evidence relating to the construct validity of the IEA Reading Literacy Tests. Theoretically, we expect these correlations to be moderately high (i.e., between .50 and .75) but not very high. The reason for expecting such moderately high correlations relates to the definition of reading literacy. Although reading literacy is defined to be contextual (i.e., may vary from one context to another), nevertheless some degree of consistency across various contexts is expected. Experience has shown that the correlation between mathematics performance and reading may vary between low to moderately high. Thus, it is reasonable to expect a higher correlation among different aspects of reading literacy. On the other hand, if the distinctiveness of the domains is to be supported, the correlations among reading literacy domains should not be very high (i.e., .90 or higher). On the basis of these considerations, the expectation was that the correlations among the domains would be moderately high.

Table 10-16 presents the correlations among the reading literacy domains for both grades. The upper-right triangle relates to grade 4 and the lower-left triangle relates to grade 9. The observed correlations are as expected. The highest correlation is 0.75 between narrative and expository for grade 9. The lowest correlation (.57) is between expository and document for grade 4. As expected, for both grades the correlation between narrative and expository domains was higher than the correlation between the narrative and document or between the expository and document.

Table 10-16. Correlation among scale scores (grade 4 above diagonal; grade 9 below diagonal)

Domain	Narrative	Expository	Document
Narrative	1.000	0.702	0.591
Expository	0.750	1.000	0.569
Document	0.612	0.642	1.000

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The proportion of shared variance among the domains can be obtained by squaring the correlations presented above. For example, for grade 4 the shared variance between the narrative and expository domains is 56 percent. Thus, while there is a strong relationship among the reading literacy domains, there is support for the distinctiveness of these domains--the common (shared) variance between two domains is typically less than half of the total variance.

10.5.4. Relationship with Other Constructs

By examining the relationship of the Reading Literacy Test scores and other constructs (gender, parental education) and determining whether the observed pattern of relationships is consistent with educational theory, the construct validity of IEA Reading Literacy Test scores can also be investigated. Table 10-17 presents mean scale scores for each of the IEA Reading Literacy domains by categories of selected demographic variables. The demographic variables, gender and parental education, were selected for two reasons: theoretical frameworks were available to evaluate the observed pattern of relationships, and comparisons to NAEP results could be made.

Table 10-17. Comparison of gender and parental education means - IEA Reading Literacy Study and NAEP

Subgroup	IEA Reading Literacy Study			1992 NAEP		
	Narrative	Expository	Document	Literary experience	Information	Perform a task
Grade 4						
Overall mean	555.2	539.4	550.9	220.3	214.9	-
Standard deviation (S.D.)	95.7	70.9	81.0	37.3	38.0	-
Male	546.1	535.1	552.1	215.8	211.6	-
Female	564.4	543.9	549.6	225.0	218.4	-
(Female-Male)/S.D.	+0.19	+0.12	-0.03	+0.25	+0.18	-
Parent(s) graduated from high school	557.1	540.6	552.9	216.3	209.6	-
Parent(s) did not graduate from high school	514.8	505.8	509.9	202.9	193.8	-
(Graduated-did not graduate)/S.D	+0.44	+0.49	+0.53	+0.36	+0.42	-
Grade 8 (NAEP)/ Grade 9 (IEA)						
Overall mean	541.9	543.5	530.4	259.0	261.1	261.2
Standard deviation (S.D.)	97.5	105.8	82.0	37.9	36.3	39.3
Male	530.3	541.3	530.0	252.5	255.2	254.3
Female	553.7	545.8	530.7	265.6	267.1	268.2
(Female-Male)/S.D.	+0.24	+0.04	-0.01	+0.35	+0.33	+0.35
Parent(s) graduated from high school	545.9	547.6	533.3	249.1	253.0	250.5
Parent(s) did not graduate from high school	480.2	480.1	484.4	242.8	243.0	241.9
(Graduated-did not graduate)/S.D.	+0.67	+0.64	+0.60	+0.17	+0.28	+0.22

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991; 1992 NAEP Reading Assessment Data Almanac.

Gender. Research has shown that the verbal skills of females are superior to those of males (Beaton 1987). To determine whether this pattern was consistent across the IEA and other studies, we converted the NAEP's reported difference between performance of males and females into effect sizes. Based on NAEP effect sizes at grade 4, females outperformed males by +0.25 on the literacy experience scale and by +0.18 on the information scale. At grade 8, females outperformed males by +0.35 on the literary experience scale, +0.33 on the information scale, and +0.35 on the perform task scale. For the Reading Literacy Study, for fourth grade students, the effect sizes of the difference between performance of females and males are 0.19, 0.12, and -0.03 for narrative, expository, and document, respectively. For ninth grade students, the corresponding effect sizes are 0.24, 0.04, and -0.01. These figures show that females clearly outperform males in narrative text. For expository texts, females appear to outperform males, although as shown in Chapter 9, the difference is not statistically significant in either grade. For documents, there is not a difference between the performances of females and males. It is notable that for the NAEP scale comparable to document (perform a task), at grade 8 there is still a substantial gender effect in favor of females.

Parental Education. Previous research has shown a positive correlation between student achievement and parents' level of education (Beaton 1987). The IEA Reading Literacy results provide additional confirmation of these findings. For grade 4, the effect size based on the difference in performance between students whose parents did not graduate from high school and those students whose parents did graduate from high school was 0.44, 0.49, and 0.53 for narrative, expository, and document domains, respectively. For grade 9, the corresponding effect sizes were 0.67, 0.64, and 0.60. If we convert into effect size NAEP's reported difference between reading scores of students whose parents did not graduate from high school and those students whose parents did graduate from high school, a value of +0.36 and +0.42 would be obtained for fourth grades from the literary experience and information scales, respectively. At grade 8, the effect sizes are +0.17, +0.28, and +0.22 for the three scales. At grade 4, these effect sizes are comparable to the effect sizes obtained from the IEA Reading Literacy Study, while at grades 8 and 9, NAEP shows smaller differences than the IEA Reading Literacy Study. Thus, the IEA Reading Literacy Test results are generally consistent with NAEP findings with respect to differences in reading performance by parents' educational level.

In summary, based on internal consistency of the domain scores, the distinctiveness of the reading literacy domains, and the observed relationship with other constructs, one can conclude that these domains represent specific sets of reading tasks that differ from one another.

10.6. Summary Comments

At the beginning of this chapter, validity was defined as the degree to which evidence supports the inferences made from test scores. The set of inferences we wished to make were also defined. They were of two types: drawing inferences about population and subpopulation differences in reading literacy and drawing inferences about factors affecting reading literacy. Throughout Chapters 9 and 10 we have carefully examined the data from the perspective of drawing inferences about population and subpopulation differences in reading literacy. We have compared the data to other extant data (i.e., NAEP), and we have examined the internal consistency of the data itself. In this concluding section, we wish to establish the degree to which the evidence provided will support inferences regarding

- The reading literacy proficiency of the U.S. fourth and ninth grade students;
- Comparisons between distributions of the reading proficiences of U.S. students and international means and standard deviations;
- Comparisons of reading proficiences among U.S. subpopulations of interest; and
- Bivariate correlations between Reading Literacy Test scores and explanatory variables of interest.

10.6.1. To What Degree Does the Evidence Provided Support Inferences Regarding the Reading Proficiencies of the U.S. Fourth and Ninth Grade Students?

Our strongest concern arises from the definition of reading literacy. To American reading specialists, the definition, as put forth by the IEA International Steering Committee, implies more than what is measured in this test. As demonstrated in Section 10.4, within the U.S., this test instrument would be considered a measure of less than a third of the domain we defined as reading literacy. Over 90 percent of the instruments for both grades correspond to the most literal understanding of text (defined in NAEP as "developing an interpretation"). Therefore, we would wish to limit the inferences drawn to discussion of only literal understanding of text.

Many might argue that, because this is at the base of all reading comprehension and is so highly correlated with the other reading processes, one should be able to make further inferences and generalizations. In the current environment where we are interested in establishing rigorous "world class" standards, however, this might erroneously create a false sense of complacency. As is consistently demonstrated by NAEP reports, very few of our students demonstrate the high levels of reading proficiency that U.S. policymakers, the consensus groups who framed the NAEP specifications, the National Assessment Governing Board, and the standards groups have deemed necessary.

Assuming the more limited definition of reading literacy as restricted to literal understanding of the text, then the inferences regarding the reading proficiences of the U.S. fourth and ninth grade students involve two interrelated sets of inferences -- inferences from a sample of test items to the population of test items and inferences from a sample of students to the population of students.

Estimates of test reliability and standard errors of measurement (Sections 10.5.1. and 9.2.1, respectively) were provided as indices for ascertaining the adequacy of the first type of inference. There are two conclusions to be drawn based on the indices provided. First, are each of the tests sufficiently reliable? Based on the evidence provided in Section 10.5.1 we conclude that

- The reliability of the narrative tests was adequate for both grades;
- The reliability of the expository test was adequate for the ninth grade but only marginally adequate for the fourth grade; and
- The reliability of the document tests was low for both grades.

Second, are estimates of reading proficiency equally reliable across the entire scale? Findings indicate that estimates of reading proficiences in the lower or middle ranges of the scales were measured more precisely as compared to the estimates of the reading proficiences of students in the upper range of the scales. This difference is due to the lack of sufficient numbers of items at the upper ranges of the scales. While this would be a major concern if decisions regarding individual students' performance were being made, it is much less of an issue for the types of inferences we wish to make.

The standard errors of sampling provide an index for determining the adequacy of this instrument for making inferences from a sample of students to the population of students. Evidence presented in Chapter 9 indicated that the standard errors of the mean for both populations were reasonably small, thus indicating that the inferences over the population of examinees were adequate.

10.6.2. To What Degree Does the Evidence Provided Support the Inferences Regarding the Reading Proficiencies of the U.S. Subpopulations of Interest?

Again, the inferences regarding the reading proficiencies of the U.S. subpopulations of interest involve two interrelated inferences -- inferences from a sample of test items to the population of test items and inferences from a sample of students to the population of students. To a large degree, the evidence provided support the inferences regarding the reading proficiencies of the U.S. subpopulations of interest.

Because estimates of reading proficiencies of students in the lower or middle ranges of the scales were more precise (as compared to the estimates of the reading proficiencies of students in the upper ranges of the scales), estimates of reading proficiencies for subpopulations whose mean reading proficiencies are at the lower ranges of the scales (e.g., black students) would be more precise than estimates for reading proficiencies for subpopulations whose mean reading proficiencies are at the upper ranges of the scales (e.g., white students).

The standard errors of sampling indicated that for most subpopulations of interest the sampling errors were reasonably small. However, for certain subpopulations (e.g., Asians, American Indians) the sampling errors were large. Therefore, in further analyses we will not make inferences about these subpopulations.

Ceiling effects were found to interact with race/ethnicity (i.e., ceiling effects for white students were more pronounced than for black or Hispanic students). Thus, ceiling effects may have masked true differences between white and minority students.

10.6.3. To What Degree Does the Evidence Provided Support the Inferences Regarding the Comparisons Between Distributions of the Reading Proficiencies of the U.S. Students with International Means and Standard Deviations?

Two types of evidence are required to ascertain the adequacy of inferences regarding the differences between the distributions of the reading proficiencies of the U.S. students and the international means and standard deviations. First, evidence supporting the estimates of reading proficiencies of the U.S. populations and subpopulations need to be examined. Second, evidence supporting the estimated means and standard deviations of different countries need to be examined. In Sections 10.5.1 and 10.5.2 we summarized the adequacy of inferences regarding the reading proficiencies of the U.S. populations and subpopulations.

The evidence in support of the estimates of national means and standard deviations in different countries appears in Elley (1992). For the purpose of the discussion, we summarize some of the important evidence currently available.

- Because test items were scrutinized by National Research Coordinators for all countries participating in the study, to some extent, test items and passages that were thought to be problematic in one or more countries were revised or eliminated.
- Because extensive pilot testing was performed in most countries participating in the study, on the basis of which test items were finalized, it is reasonable to expect that test reliability and validity would be enhanced.
- Because items not fitting the Rasch model for the international calibration sample were dropped, it is reasonable to expect greater test validity for estimating the international distributions.
- Because items not fitting the Rasch model within a country were not used in estimating student abilities for that country, it is reasonable to expect greater test validity for estimating within country distributions.

- Although measures of reliabilities varied across countries, in general these variations were not very large, indicating that items were operating consistently across countries.

It should be pointed out, however, that the evidence summarized above in support of the international scaling procedures is far more limited than the evidence provided in support of the intended uses of these test scores for the U.S. populations. For example, we do not have a clear picture of the shapes of international distributions (i.e., distribution of scaled scores for all students participating in the study) to ascertain the adequacy of the mean and standard deviation as appropriate measures of the central tendencies and dispersions, respectively. If these distributions are highly skewed, or multimodal, then the mean and standard deviation would not be appropriate statistics for characterizing the central tendencies and dispersions of these distributions.

10.6.4. To What Degree Does the Evidence Provided Support the Inferences Regarding the Bivariate Correlations Between Reading Literacy Test Scores and Explanatory Variables of Interest?

A major intended use of the IEA Reading Literacy test scores within the U.S. is to identify important correlates of those scores. Complex statistical methods (such as hierarchical linear modeling), which are generally based on bivariate correlations, will be used. Therefore, we will focus on the degree to which the evidence provided support the inferences regarding the bivariate correlations between the Reading Literacy Test scores and explanatory variables of interest.

It is well known (e.g., Nunnally 1967) that errors in measurement tend to attenuate the observed correlation between two variables. Thus, if two variables are highly correlated but the variables are measured with a large degree of error, the observed correlation between the two variables may not be high. The implication for this study is that the bivariate correlations between the narrative scaled scores and explanatory variables of interest would not be as attenuated as similar correlations for the expository and document scaled scores.

It should be pointed out that measurement errors associated with both the Reading Literacy Test scores and the explanatory variables of interest would attenuate the observed correlations. In general, the measurement errors associated with the explanatory variables are larger because great care is exercised in constructing tests for measuring the dependent variables while relatively lax standards are applied in measuring the explanatory variables of interest.

Ceiling effects also tend to attenuate these bivariate correlations. Ceiling effects operate as if the scaled scores were truncated at the upper ranges of the scale. Truncated scores tend to exhibit lower correlations as compared to distributions that are not truncated. For the following reasons, we do not expect the attenuation due to ceiling effects on the bivariate correlations between the reading literacy scaled scores and explanatory variables of interest to be as large as the attenuation due to measurement errors:

- With the exception of the fourth grade narrative scaled scores, ceiling effects were generally small; and
- Relatively, measurement errors associated with the explanatory variables may be larger than the "truncation" associated with the reading literacy scaled scores.

Other characteristics of the distributions, for example, the shapes of the distributions, may also affect the observed correlations. If a distribution is skewed (as is the fourth grade narrative scaled scores distribution, for example), the bivariate correlations tend to be attenuated. Once again, we expect the attenuation due to the shapes of the distributions to be smaller than the attenuation due to measurement errors.

Thus, a number of factors may have operated to attenuate the bivariate correlations between the Reading Literacy Test scores and the explanatory variables of interest. Because the errors of measurement associated with the narrative scaled scores were smaller than the other two domains, it is reasonable to expect that the bivariate correlations between the narrative scaled scores and the explanatory variables of interest would be less attenuated than similar correlations for the expository and document domains.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Armore, D.J. (1974). Theta reliability and factor scaling. In H.L. Costner (ed.), *Sociological Methodology,* 17-50. San Francisco: Jossey-Bass.

Baker, L., and Brown, A.L. (1984). Metacognitive skills and reading. In P.D. Pearson (ed.), *Handbook of Reading Research.* New York: Longman.

Beaton, A.E. (1987). *The NAEP 1983-84 technical report.* Princeton, NJ: Educational Testing Service.

Bock, R.D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika,* 46, 443-459.

Bock, R.D., Gibbons, R.D., and Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement,* 12, 261-280.

Carroll, J.B. (1945). The effect of difficulty and chance success of correlation between items or between tests. *Psychometrika,* 10, 1-19.

Cook. L.L., and Eignor, D.R. (1984). Assessing the dimensionality of NAEP reading test items: Confirmatory factor analysis of item parcel data. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1984.

Council of Chief State School Officers. (1990). *Reading framework.* Washington, DC: CCSSO.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16, 297-334.

Cronbach, L.J. (1989). Construct validation after thirty years. In R.L. Linn (ed.), *Intelligence: Measurement, theory, and public policy -- Proceedings of a symposium in honor of Lloyd G. Humphreys,* 147-171. Chicago: University of Illinois Press.

Drasgow, F., and Lissak, R.I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology,* 68, 363-373.

Elley, W.B. (1992). *How in the world do students read? IEA study of reading literacy.* The Hague: The International Association for the Advancement of Educational Achievement.

Green, S.B., Lissitz, R.W., and Mulaik, S.A. (1977). Limitation of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement,* 37, 827-838.

Hattie, J. (1985). Methodology review: Assessing undimensionality of tests and items. *Applied Psychological Measurement,* 9, 139-164.

Meyer, B.J.F., and Rice, G.E. (1984). The structure of text. In P.D. Pearson (ed.), *Handbook of Reading Research*. New York: Longman.

Nunnally, J.C. (1967). *Psychometic theory*. New York: McGraw-Hill.

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Wilson, D.T., Wood, R., and Gibbons, R. (1991). TESTFACT. Chicago, IL: Scientific Software.

Wright, B.D., and Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23-48.

Zwick, R. (1987). Assessment of dimensionality of year 15 reading data. In A.E. Beaton (ed.), *The NAEP 1983-84 technical report*, 245-284. Princeton, N.J.: Educational Testing Service.

PART III. THE VARIABLES THAT AFFECT READING LITERACY

As a major part of IEA studies, Student, Teacher, and School Questionnaires are used to provide the basis for explanatory analyses. The prime purpose of these analyses is to account for the observed between-student variation in achievement in items of parallel variation in characteristics of the students, their families, their teachers and the instruction they provide, and their schools. These analyses are designed to explain why some students, and some nations, do better than others.

This section describes the data available from the IEA Reading Literacy Study to explain differences in reading literacy achievement among students within the United States. Some of the variation in performance detected in this data set can be traced to pre-existing conditions that students have when they enter a particular class, some can be attributed to the conditions of the school and the class as the context in which learning takes place, and some can be attributed to the interactions that occur within the learning environment. By systematically examining these variables and the interactions among them, we may be able to test certain theories and myths about how reading literacy proficiency is acquired.

In an effort to relate those theories and myths to the study data, we have organized this part of the report into four chapters.

- **Chapter 11: The Survey Instruments.** *Marilyn Binkley*. A description of the survey instruments used in the IEA Reading Literacy Study.
- **Chapter 12: Imputation.** *Marianne Winglee, Marilyn Binkley, Graham Kalton, Keith Rust*. An analysis of the quality of the responses to the surveys and the uses of imputation to improve the quality of the data.
- **Chapter 13: Constructs and Data.** *Marilyn Binkley, Trevor Williams, Jacqueline Haynes*. A description of the available constructs that affect reading literacy.
- **Chapter 14: Modeling the Reading Literacy of Fourth and Ninth Graders.** *Trevor Williams*. A description of the model developed to explain why students differ in their ability to comprehend written text.

11. THE SURVEY INSTRUMENTS

11.1. Introduction

As with most IEA studies, Reading Literacy Study data were collected through survey instruments. Three sources of information were tapped: the students themselves, their teachers, and the principal of their school. (An additional instrument, in the form of a national questionnaire, was also used to collect data about the national system. However, for the purposes of the U.S. technical report, nationally aggregated information had no bearing on the analyses.) The international instruments were composed of items requesting data on standard demographic characteristics, available resources, and practices related to reading achievement.

Survey questions were written by committees working independently, reviewed and revised by the International Steering Committee, and then vetted by the National Research Coordinators, who were given an opportunity to modify and add questions. The survey instruments were pilot tested and revised for the main study. In the U.S., additional items were added to the instruments at that time.

11.2. School Questionnaire

The international School Questionnaires, designed to be answered by the school principal or headmaster, were identical for both populations. The 24 general questions were intended to collect information that would provide a backdrop for understanding the context of instruction.

Questions concerned principal's years of experience, size of the school and grade being tested, type of school, type of community, and community characteristics such as availability of public libraries, bookstores, secondary and tertiary level schools, degree of parent cooperation, and resources in the school, followed by items closely related to instruction. These questions concerned the size and accessibility of the school library, the number and gender of teachers and whether they were classroom or specialty teachers, the amount of instruction per week, the number of weeks per year, and student attendance. This was followed by general questions about reading instruction -- the types of special programs and improvement programs available and the problems encountered related to providing reading instruction -- and those about frequency and type of teacher evaluation and the role of the principal in it.

Consistent with IEA policy that allows each National Center to add to its own questionnaires, the U.S. did include additional items related to issues of importance to the U.S. Those who designed the questions hoped to obtain data relevant to U.S. practice, policy, and current interests. The added questions were of three types: refinements of demographic information, which relate directly to American practice; information about student assessment and evaluation procedures; and extended information about the principal, such as age, experience, training, and reading habits.

The additional demographic information was sought because we wished to know more about the ethnicity of both the student body and the faculty. We used the common U.S. classifications for community type, which are more refined than the international categories, in an effort to get a better measure of urban and rural differences. We asked more detailed questions about the types of special teachers available within a school to determine the effect of differing amounts of available resources.

Based on the effective schools literature, and the current move to shift greater curricular control and decision making to the school site, the role of the principal as an instructional leader was also of interest. The set of questions focused on the qualifications of the principals themselves, especially their preparation for this leadership role as it pertains to reading, by asking about principals' training in education administration, teaching experience and training in reading education, and personal reading habits. It was hoped that these questions might give us a better understanding of how well prepared principals were to have an impact on instructional practice.

11.3. Teacher Questionnaires

The international Teacher Questionnaires were designed to be answered by the teacher of the class being tested. Unlike the School Questionnaire, they were different at the two population levels, with the grade 4 questionnaire being somewhat more extensive. This difference was based on two rationales. First, fourth grade students are still likely to be receiving explicit reading instruction, while ninth grade students usually receive such instruction only if they are in remedial programs. Second, because there were likely to be too many antecedent conditions to be controlled for in grade 9, information about instruction could not be tied directly to student achievement. Given these differences, the two surveys will be described separately.

Grade 4. The international Teacher Questionnaire for grade 4 consisted of 46 separate questions. It included queries about teacher characteristics such as gender, mother-tongue, education and training, inservice training, reading habits, and years of experience. Additional questions tapped characteristics of the class being tested, including the number of years the teacher taught the class as a group, size of the class, number of students in the class whose mother-tongue was not the language of instruction, degree of parent-teacher interaction, need for and availability of remedial instruction, and amount of instruction provided in general and in reading particularly.

Attention then shifted to how the teacher organized instruction. These questions focused on reading activities, textbook availability, aims of instruction, instructional strategies, grouping, relative emphasis of text genre, teachers' attitudes toward issues in reading instruction, assessment methods, and assignment of homework. These were followed by questions related to the availability and use of classroom and school libraries, and those about school organization, focusing on evaluation, the principal's role, and staff meetings.

The U.S. made additions to the grade 4 Teacher Questionnaire that increased its length to 63 questions. We sought additional demographic information with questions on age and ethnicity, and further specified attributes of certification and training. We also asked for more information on the types of inservice available. The major addition to the survey included a set of statements used to discern the frequency of teaching practices involving selection of materials and use of particular strategies.

Grade 9. The international Teacher Questionnaire for grade 9 was limited to 28 questions. It followed the same general pattern as the international teacher questionnaire for grade 4. However, it did not contain questions on further training in reading through course work or inservice, whether the class was multigrade, how frequently teachers met with parents of students in the class, grouping for instruction, attitude toward issues in reading instruction, assessment strategies, homework, and classroom libraries. The questions on teaching strategies and student reading strategies were less extensive than those on the grade 4 questionnaire. We made the same additions to the grade 9 questionnaire as we had to the grade 4 questionnaire, increasing its length to 44 questions.

11.4. Student Questionnaires

Grade 4. The international Student Questionnaire for grade 4 contained 43 separate questions, divided into five groups. *Yourself and your home* included age, gender, language at home, meals, access to newspaper, TV viewing, books in home, a measure of wealth, and whether you are read to or read to someone whether at home or elsewhere. *Your reading* focused on whether books were borrowed from the library, a self-rating of reading ability, rating ways of becoming a good reader, frequency of reading at home, frequency parents ask about your reading. *Your reading homework and class work* focused on the frequency and length of time spent doing homework, whether teachers followed up on the reading done at home, help with reading homework, carry over of class work to homework, and frequency of written assignments related to reading work. *Reading for enjoyment* asked about kinds of reading materials you read for enjoyment, how frequently you read them, and reading aloud at home -- how frequently, to whom, and what was read. *Reading in school* focused on the frequency that textbooks, story books, and workbooks are used in reading instruction, the use of text materials in other subject areas, and the use of reference materials, such as encyclopedias, dictionaries, manuals and maps.

The U.S. expanded this questionnaire by adding 26 questions. We wished to know more about the family in terms of ethnicity, social status based on the education levels of parents, composition of the family, use of languages other than English in the home, and, where English was not the students' first language, we wished to get some sense of the degree of fluency. We expanded the query on how students read by asking about the strategies they use before, while, and after reading, and also asked what kinds of reading work was normally done for homework.

Grade 9. The international Student Questionnaire for grade 9 contained 72 separate questions. It followed the same format as the grade 4 questionnaire, but added to specific sections. Included in the category yourself and your home, students were asked about their parents' education, their own job or family responsibilities, and their expectations for further education. They were not asked about reading aloud activities. They were asked how often they are given homework in general and how much time they spend on all their homework. Questions about reading homework were significantly reduced; instead, these students were asked about the amount of time they spend reading silently in class, and how often they are asked to do written work related to what they have read. These students were asked indepth questions about the frequency of what they read, including types of materials used for school and homework, as well as types of books read for pleasure, topics they normally read about in magazines and newspapers, and types of documents they use. Finally, they were asked questions about the quality of school life.

The U.S. expanded this questionnaire with additional questions focused on family ethnicity, family composition, use of second language, fluency in English, fluency in the second language, reading strategies used, writing outside of school, and types of homework normally assigned.

11.5. Summary Description of the Questionnaires

Table 11-1 serves as an index to the available information. Items are categorized according to a general framework (as described in Chapter 13), grouping information into preliminary meaningful groups. The table indicates the survey instrument in which the item appears on and the originating source for the item, that is, the international instrument, a U.S. rework of an international instrument, a U.S. rework of an international item, an agreed-upon special national option, or a U.S. national option.

Table 11-1. Questionnaire items by general framework and specific category: Grades 4 and 9

Item	Grade 4	Grade 9	Type
1. Student Attributes			
Category			
a. Age	S4Q1 S4Q2	S9Q1 S9Q2	U.S./International U.S. National Option
b. Gender	S4Q3	S9Q3	International
c. Language of test	S4Q25-32	S9Q31-38	Special National Option International
d. Ethnicity	S4Q4 P4Q3	S9Q4 P9Q3	Special National Option
e. Bilingualism	S4Q15-24	S9Q21-30	Special National Option International
2. Family Attributes			
Category			
a. Father's education	S4Q5	S9Q5	U.S. National Option
b. Mother's education	S4Q6	S9Q6	U.S. National Option
c. Family wealth	S4Q11-12	S9Q17-18	U.S./International U.S. National Option
d. Ethnicity	S4Q4	S9Q4	Special National Option
e. Bilingualism	S4Q15-24	S9Q21-30	Special National Option International
f. Family size	S4Q13-14	S9Q19-20	U.S. National Option
3. School Attributes			
Category			
a. Principal's experience	P4Q35	P9Q35	U.S./International
b. Age	P4Q30	P9Q30	U.S. National Option
c. Gender	P4Q29	P9Q29	U.S. National Option
d. Ethnicity	P4Q31	P9Q31	U.S. National Option
e. Preservice educ. administration	P4Q32-33	P9Q32-33	U.S. National Option
f. Inservice educ. administration	P4Q34	P9Q34	U.S. National Option
g. Years of teaching experience	P4Q36	P9Q36	U.S. National Option
h. Professional training	P4Q37	P9Q37	U.S. National Option
i. Education in teaching reading	P4Q38	P9Q38	U.S. National Option
j. Inservice reading education	P4Q39	P9Q39	U.S. National Option
k. General reading interest	P4Q40	P9Q40	U.S. National Option
l. School enrollment	P4Q1	P9Q1	International
m. # students/grade/sex	P4Q2	P9Q2	International
n. Public/private	P4Q4	P9Q4	International
o. Urbanicity	P4Q5	P9Q5	U.S./International
p. Resources/activities	P4Q6 P4Q8	P9Q6 P9Q8	U.S./International International
q. # FTE teaching teachers	P4Q13	P9Q13	U.S./International
r. Total instructional time	P4Q14	P9Q14	U.S./International
s. Time school open; weeks/year	P4Q15	P9Q15	U.S./International
t. Days instruction lost	P4Q16	P9Q16	International
u. % students absent	P4Q17	P9Q17	International
v. Standardized tests	P4Q18-19	P9Q18-19	U.S. National Option
w. Rating students' progress	P4Q20	P9Q20	Special National Option

Table 11-1. Questionnaire items by general framework and specific category: Grades 4 and 9 (continued)

Item	Grade 4	Grade 9	Type
4. Teacher Attributes			
Category			
a. Gender	T4Q1	T9Q1	International
b. Language of test	T4Q3	T9Q3	U.S./International
c. General education	T4Q5	T9Q5	U.S./International
d. Preservice teacher education	T4Q6-8	T9Q6-8	U.S./International U.S. National Option
e. Postsecondary education	T4Q5	T9Q5	U.S./International
f. Further education in reading	T4Q11	T9Q11	U.S./International
g. Inservice reading education	T4Q12-14	T9Q12-14	International U.S. National Option
h. General reading interests	T4Q17	T9Q17	International
i. Years of teaching experience	T4Q15-16	T9Q15-16	International U.S. National Option
j. Aims of reading instruction	T4Q34	T9Q30	International
k. Attitudes to reading instruction	T4Q43	No items	International
l. Teacher rating students' literacy	T4Q46	T9Q24	International U.S./International
m. Ethnicity	T4Q4, P4Q12	T9Q4, P9Q12	Special National Option
n. Age	T4Q2	T9Q2	U.S. National Option
o. Teacher certification	T4Q9-10	T9Q9-10	Special National Option
5. Quality of school life			
No items	No items	S9Q68	
6. Family Environments			
Category			
a. Meals eaten each day	S4Q7	S9Q7	International
b. Responsibilities/job	No items	S9Q8-11	
c. Newspaper in home	S4Q8	S9Q13	International
d. Hours of TV watching	S4Q9	S9Q14	U.S./International
e. # of books in home	S4Q10	S9Q15	International
f. Wealth (possessions)	S4Q11-12	S9Q17-18	U.S./International U.S. National Option
g. Parent interest	S4Q42	S9Q16	International
h. Parents read (test language)	S4Q31	S9Q37	International
i. Others read (test language)	S4Q32	S9Q38	International
j. Parents read (other language)	S4Q23	S9Q29	International
k. Others read (other language)	S4Q24	S9Q30	International
l. Reads to parent (test language)	No items	No items	
m. Help with reading homework	S4Q50	No items	International
n. Read aloud at home	S4Q38-41	S9Q48-51	International
7. School Environment			
Category			
a. T's work evaluated by parent	T4Q63	T9Q40	International
b. P discusses with teacher	T4Q64	T4Q41	International
c. Staff meetings; frequency	T4Q65-66	T9Q42-43	U.S. National Option International
d. Items at staff meeting	T4Q67	T4Q44	International

Table 11-1. Questionnaire items by general framework and specific category: Grades 4 and 9 (continued)

Item	Grade 4	Grade 9	Type
e. Community resources	P4Q6 S4Q33	P9Q6 S9Q52	U.S./International U.S. National Option
f. Parent cooperation	P4Q7	P9Q7	International
g. Books in library	P4Q9	P9Q9	International
h. Books added	P4Q10	P9Q10	International
i. Students borrow books	P4Q11	P9Q11	International
j. Special reading programs	P4Q21 S4Q37	P9Q21	Special National Option U.S. National Option
k. Informal reading programs	P4Q22	P9Q22	International
l. Programs improve read. instruction	P4Q23	P9Q23	Special National Option
m. Reading problems at school	P4Q24-25	P9Q24-25	U.S. National Option U.S./International
n. Parent's activities	P4Q26	P9Q26	International
o. Parent evaluates teachers; frequency	P4Q27	P9Q27	International
p. Parent evaluates procedures	P4Q28	P9Q28	U.S./International
8. Classroom environment			
Category			
a. Teacher's years teaching class	T4Q18	No items	U.S./International
b. Multigrade class	T4Q19	No items	International
c. Class size	T4Q20	T9Q18	International
d. # non-test-language students	T4Q28	T9Q19	International
e. Meet with parents	T4Q23-24	No items	U.S./International U.S. National Option
f. # students need remedial	T4Q21	T9Q20	International
g. # students receive remedial	T4Q22	T9Q21	International
h. Instructional time; total	T4Q25	T9Q22	International
i. Instructional time; test language	T4Q26 T4Q29	T9Q23	International U.S. National Option
j. Instructional time; reading	T4Q27	T9Q23	International
k. Availability of reading texts	T4Q31-33	T9Q33-35	U.S. National Option U.S./International
l. Classroom library	T4Q54 T4Q58	No items	International U.S. National Option
m. Classroom library; # books	T4Q55	No items	U.S./International
n. Classroom library; mags., newsp.	T4Q56	No items	U.S./International
o. Students can borrow	T4Q57	No items	International
p. School library	T4Q59 T4Q62	T9Q36, T9Q39, S9Q53	International U.S. National Option
q. Class visits to school library	T4Q60	T9Q37	International
r. Students can borrow	T4Q61	T9Q38	International
s. Student reading ability	T4Q37	No items	U.S. National Option
9. Teacher instruction categories			
Category			
a. General homework	S4Q54	S9Q39-42, S9Q45, S9Q47	U.S. National Option
b. Teaching; T directed/S directed	T4Q53	T9Q32	Special National Option

Table 11-1. Questionnaire items by general framework and specific category: Grades 4 and 9 (continued)

Item	Grade 4	Grade 9	Type
10. Teacher instruction; content			
Category			
a. Time on narr./expos./doc.	T4Q41-42	T9Q25-26	International U.S. National Option
11. Teacher reading strategies			
Category			
a. Reading homework	S4Q46-53	S9Q43-44, S9Q46	U.S. National Option International
b. Time silent reading	No items	S9Q59	
c. Assigned written work on reading	S4Q52	S9Q60	International
d. Instruction strategies	T4Q35	No items	U.S./International
e. Instructional grouping	T4Q38	No items	International
f. Type of grouping	T4Q39	No items	U.S./International
g. # of groups	T4Q40	No items	International
h. Assign reading homework	T4Q48	No items	International
i. Assign reading homework; freq.	T4Q49, T4Q51	T9Q29, S9Q43	U.S./International
j. Assign reading homework; time	T4Q50, T4Q52	No items	U.S./International
k. Use of 28 reading activities	T4Q30	T9Q28	International
l. Encourage reading out of school	T4Q46	T4Q27	U.S./International
m. Increase S comprehension	T4Q36	T9Q31	Special National Option
n. Use of comprehensive strategies	No items	No items	
12. Teacher remedial strategies			
No items	No items	No items	
13. Teacher assignment strategies			
Category			
a. 10 methods assess student needs	T4Q45	No items	U.S./International
b. Assess reading; freq.	T4Q46	No items	International
c. 7 methods of assessment	T4Q47	T9Q24	U.S./International
14. Student reading strategies			
Category			
a. Student reading strategies	S4Q43-45	S9Q56-58	U.S. National Option
15. Student reading activities			
Category			
a. Books for school/homework	S4Q64-69	S9Q63	International
b. Leisure reading	S4Q55-58	S9Q64	International
c. Magazine reading	S4Q59-60	S9Q65	International
d. Newspaper reading	S4Q61-62	S9Q66	International
e. Document reading	S4Q63	S9Q67	International

Table 11-1. Questionnaire items by general framework and specific category: Grades 4 and 9 (continued)

Item	Grade 4	Grade 9	Type
16. Student reading attributes/behaviors			
Category			
a. Borrow library books; freq.	S4Q34	S9Q53	International
b. Self-concept ability; reading	S4Q35	S9Q54	U.S./International
c. 3 ways to become good reader	S4Q36	S9Q55	International
17. Student attitudes; education generally			
Category			
a. Educational expectations	No items	S9Q12	U.S./International International

KEY: T = teacher; P = principal; S = student; FTE = full-time equivalent

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

12. IMPUTATION

12.1. Introduction

The U.S. component of the IEA Reading Literacy Study involved national probability samples that exceeded 6,500 assessed fourth grade students and 3,200 assessed ninth grade students, with 167 schools participating at grade 4 and 165 at grade 9. Some 300 fourth grade and 160 ninth grade teachers provided data for the study.

Responses to the questionnaire items were not always complete, although item nonresponse would be considered low by most standards. With the view to improving the statistical properties of the data and facilitating the analyses designed to model family, teacher, and school influences on reading comprehension, the decision was made to impute values for the missing observations. In this chapter, we document the following aspects of the missing data:

- The nature, extent, and sources of item nonresponse for each of the three questionnaires in both samples for the U.S. component of the IEA Reading Literacy Study;
- The procedures used to impute replacement values for these missing data; and
- The effects of this imputation on univariate statistics, relational statistics, and sampling error.

12.2. The Nature of the Missing Data

We have compiled summary data on the extent of item nonresponse in each of the six data sets corresponding to the three questionnaires used with each of the two samples (Table 12-1). This summary is the basis for a more detailed discussion to follow. With the view to providing an overall picture of the extent of nonresponse in the data, we show the proportion of items missing in each data set in each of five categories: 5 percent or less, 6 to 10 percent, 11 to 15 percent, 16 to 20 percent, and more than 20 percent missing.

These data make clear that, overall, we do not have a serious problem with item nonresponse. School principals completed the questionnaires to the extent that close to 90 percent of items have no more than 5 percent missing, and the teachers surveyed were similarly conscientious--92 and 84 percent, respectively, of the items in the fourth and ninth grade Teacher Questionnaires had no more than 5 percent missing. Ninth grade students had similar levels of item nonresponse to the 241 items on their questionnaire. Nonresponse was somewhat problematic only in the predictable situation, that is, among fourth grade students. Even here, only 46 percent of the items had more than 5 percent missing data, and 80 percent were missing 10 percent or less.

In providing a more detailed look at the extent of the missing data item by item, we have assumed arbitrarily that items with 5 percent or less missing data do not require specific attention here. Thus, in the tables that follow we provide detail only on those items with more than 5 percent missing responses. (Information on the percent of missing data on each item is presented in the seven appendices to Chapter 12, available on request from Westat, Inc.)

Table 12-1. Percentage of questionnaire items with missing data: Grades 4 and 9

Amount of missing data for item	Percent of items	
	Grade 4	Grade 9
Student Questionnaires		
5 percent or less	54	87
6-10 percent	26	5
11-15 percent	14	5
16-20 percent	5	2
Over 20 percent	1	1
Total number of items	N = 134	N = 241
Teacher Questionnaires		
5 percent or less	92	84
6-10 percent	5	14
11-15 percent	2	2
16-20 percent	0	0
Over 20 percent	1	0
Total number of items	N = 250	N = 153
Principal Questionnaires		
5 percent or less	89	87
6-10 percent	4	12
11-15 percent	6	0
16-20 percent	1	0
Over 20 percent	0	1
Total number of items	N = 113	N = 117

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

There is a consistent pattern of nonresponse across the six questionnaires. Six kinds of questionnaire items appear to generate much of the nonresponse:

- **Not applicable items**--Items that do not apply to a respondent and do not allow a "not applicable" response are more likely to be skipped;
- **List checking items**--Multiple-part questions requiring a yes/no response for each part tend to be troublesome. Many respondents appear to answer by circling only those items where "yes" would be the answer;
- **Other category items**--Items having a residual "other" category, which respondents tend to ignore;
- **Information retrieval item**--Items requiring information that may not be readily available without investing effort in information retrieval;
- **Rank order items**--Those requiring respondents to rank order a whole or partial list; and
- **Definitional items**--Those items missing the necessary definitions for clarity.

12.2.1. Principal Questionnaires

In examining item-by-item information for both the fourth and ninth grade Principal Questionnaires, two aspects of these data are immediately apparent: first, the highest percentage missing on any item is 24 percent; and second, most of the nonresponse is located in multipart questions--question 3 at the fourth grade, question 19 at the ninth grade, questions 20 and 28 at both grades (Table 12-2).

Table 12-2. Percentage of items with over 5 percent missing data: Principal Questionnaires

Variable label	Percent of items	
	Grade 4	Grade 9
3 Race/Etn: Asian, Pac Island	11	
Am Indian, Alaskan	11	
Hispanic	11	
White (non-Hisp)	11	
Black (non-Hisp)	11	
9 Schl Lib/ N books contained	7	6
10 Schl Lib/ N books added last yr		7
14H Instruct time/week, hours		6
19 Used student's standardized test to evaluate:		
teachers		7
textbooks		7
20 Rate of satisfaction with:		
norm-ref test score	7	9
criterion-ref score	17	24
stud wrk sample		10
teacher judgment		8
grade report		10
28 Procedure used to gather information for evaluation:		
interviews	7	
self-report by teacher	10	
student ratings	11	8
other	13	8
35 N Yrs principal, career		6
37 Courses in Eng/Lang Arts/Rdg	7	7

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Not Applicable Items. An examination of question 20 may make it clearer why nonresponse is high for this question.

20. Rate your satisfaction with the following sources of evidence of your students' progress. (Circle one number on each line.)

	Highly satisfied			Highly dissatisfied	
a. Norm-referenced test scores	1	2	3	4	5
b. Criterion-referenced test scores	1	2	3	4	5
c. Student work samples	1	2	3	4	5
d. Teachers' judgment	1	2	3	4	5
e. Grade report	1	2	3	4	5

As written, this question does not provide for the possibility of a "not applicable" response. Considering the variation in nonresponse across the five categories in this question, one might conclude that this variation in nonresponse is associated with the extent to which these assessment methods are used by schools. For instance, the highest nonresponse rates are associated with criterion-referenced test scores. Since the use of standardized tests of this kind has declined over the past 10 years, the item probably requires a "not applicable" response for many schools. Further, it is also likely that principals do not refer to tests by their technical names and subsequently may not understand the question.

List-Checking Items. Question 28 from the fourth grade Principal Questionnaire and question 19 from the ninth grade Principal Questionnaire demonstrate another type of item nonresponse, one we have labeled "list-checking" items.

28. Which of the following procedures do you use to gather information for your evaluation? (Circle one per line.)

	Yes	No
a. interviews	1	2
b. written or oral self-report by teachers	1	2
c. observational data on teachers' classroom work	1	2
d. student ratings of teachers' performance	1	2
e. other forms of systematic evaluation.	1	2

19. Do you use the results of student's standardized tests to evaluate:

	Yes	No
a. student progress?	1	2
b. curriculum?	1	2
c. teachers?	1	2
d. textbooks/materials?	1	2
e. special programs?	1	2

In these questions nonresponse varies from less than 5 percent to 13 percent across the parts, and very few respondents totally skipped the question. Consequently, we believe that the item format is responsible in large part; respondents responded positively to those procedures they actually use and skipped over the others, rather than circling the "No" response as instructed.

Information Retrieval Items. Question 3 from the fourth grade Principal Questionnaire demonstrates another type of nonresponse pattern, that we have labeled "information retrieval" items.

3. How many of the full-time fourth grade students in your school are:

a. Asian or Pacific Islander ___________students
b. American Indian or Alaskan Native ___________students
c. Hispanic ___________students
d. White (non-Hispanic), or ___________students
e. Black (non-Hispanic)? ___________students

For question 3, there is a consistent 11 percent nonresponse rate across all of the subparts. Given this pattern of nonresponse, we suspect that while principals may know the overall number of students within the school belonging to each group, they may not know the breakdown by grade. Although they may be willing to invest the time into answering the questionnaire based on information they can immediately recall, they may not be willing to search for all the requested information. Additionally, we do know that some schools have policies that prohibit reporting data of this kind.

12.2.2. Teacher Questionnaires

Nonresponse, other than that generated by the question design problems noted previously, is not a problem of any size in the Teacher Questionnaires (Table 12-3). Only 5 of the 67 questions on the fourth grade questionnaire exceeded 10 percent nonresponse, and only 3 of the 44 questions on the ninth grade questionnaire were at this level.

Other Category Items. The highest level of nonresponse (32 percent and 41 percent) occurred in two questions, questions 39d and 44f, respectively, both from the fourth grade questionnaire. Both questions 39d and 44f are residual categories in multipart items requiring a response to "other," so it is reasonable to assume that nonresponse to this part of the question alone corresponds to an answer of "never" and "no," respectively.

39. How often do you use each of these types of groupings?

	Frequency			
	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
a. age groups	1	2	3	4
b. ability groups	1	2	3	4
c. interest groups	1	2	3	4
d. other (please specify) _______	1	2	3	4

44. Do you regularly (i.e., at least once a week) do the following activities to encourage your students to read outside school?

	Yes	No
a. suggest books (to students) to read	1	2
b. suggest newspaper articles to students to read	1	2
c. read stories to students	1	2
d. hold discussions about books	1	2
e. encourage students to borrow library books	1	2
f. other ________________	1	2

246

Table 12-3. Percentage of items with over 5 percent missing data: Teacher Questionnaires

Variable label		Percent of items	
		Grade 4	Grade 9
(7:7)	Number of teacher education courses	23	15
(8:)	Percent of time learn teaching of reading	6	
(32:34)	Different texts per student	13	14
(34:30)	Rate aims		
	lasting interest		9
	reading comprehension	6	7
	extend vocabulary	6	10
	critical thinking	6	8
	expand read choice	6	8
	deepen emot devel	6	
	word attack skill	6	
	make reading enjoyable	6	
	research/study skills		10
	expand world view		8
	increase speed of reading		12
	apply strategy to other subjects		10
	appreciation of literature		8
	interpret diagrams		6
(42:26)	Percent		
	teach narration	6	10
	teach exposition	6	10
	teach documents	6	10
(29:)	Time teaching ESOL	14	
(35:)	Strategies/graded text difficulty	7	
(39:)	Freq use		
	age groups	14	
	interest groups	12	
	other groups	41	
(44:)	Encourage students/other	32	
(:24)	Assess methods		
	teacher quizzes		9
	multiple choice		8
	student interests		7
	oral discussion		10
	discuss material read		10
	open-ended question		10
	essays about literature		9
(:27)	Encourage stud/read assign		6
(:44)	Percent/other topics		6

NOTE: The first number in the parentheses refers to the item number on the fourth grade Teacher Questionnaire; the second number refers to the item number on the ninth grade questionnaire.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Rank Order Items. In the case of question 34 from the fourth grade Teacher Questionnaire and question 30 of the ninth grade Teacher Questionnaire (the questions are identical), the nonresponse is due in part to the nature of the question itself, which asks teachers to rank 5 out of 12 items. If a teacher gave any response but failed to indicate exactly five aims, appropriately ranked, this was then regarded as nonresponse to the remaining items. As such, it was easy to generate the observed levels of nonresponse shown in Table 12-3.

34. Please rank five of the following aims of reading instruction in order of the importance you attach to each of them. (Place "1" next to the most important and so on to "5" for the least important. Choose only five aims, and use each rank only once.)

Importance

a. developing skill in reading aloud __________
b. developing a lasting interest in reading __________
. ...
. ...
. ...
l. making reading enjoyable. __________

Question 24 of the ninth grade Teacher Questionnaire also asks teachers to rank order. Their responses, in this case on a question about assessment methods, are to be a ranking of all items. In addition to the difficulty with ranking questions, the wording of this question implies that all teachers use all of these methods. However, this is not necessarily the case; therefore, all methods would simply not be ranked.

24. What assessment methods do you use most often in this English/Language Arts/Reading class? (Rank order by assigning a "1" to the most frequent, "2" to the next, and "7" to the least frequent.

Frequency

a. teacher quizzes __________
b. multiple-choice questions __________
c. records of student interests __________
d. oral discussions __________
e. oral discussions on material read __________
f. written open-ended questions on material read __________
g. essays in response to literature. __________

The difficulty with question 29 from the fourth grade Teacher Questionnaire rests in the question itself. Teachers were asked about the time devoted to the teaching and learning of ESOL. However, no definition of ESOL was provided. Inspection of the responses in the actual questionnaires themselves demonstrated that it was incorrect to assume that teachers would know what ESOL was. In a few cases, teachers who only taught native English speakers filled in the blanks with time allocations, probably assuming that we meant the more general language arts curriculum described in the parentheses. Additionally, a number of teachers took what appears to be the "when in doubt, leave it out" approach.

29. How much time per school week is typically devoted to the teaching and learning of ESOL (including, reading, writing, speaking, literature, listening, and other language skills) for the class?

____________hours and __________minutes per week

Information Retrieval Items. Question 7, which was the same on both Teacher Questionnaires, clearly asks for information that few would remember accurately, if they knew it in the first place. Considering that teachers would most likely need a transcript to figure out the correct response and that they were asked to respond during the testing period, it is not surprising that we have a nonresponse rate of 23 percent and 15 percent, respectively, for fourth and ninth grade teachers.

7. How many teacher education courses did you complete?

________________courses

In looking at question 26 from the ninth grade Teacher Questionnaire and question 42 from the fourth grade Teacher Questionnaire, which are identical, we note that there is a difference in non-response rates across the grade levels. For ninth grade teachers, this question may be inappropriate, because this group, who would generally be classified as English teachers, tend to emphasize almost exclusively narrative texts and poetry (which is not a possible response), and would have little reason to teach students about either exposition or documents. As with all the explanations generated in this section, this is speculation. These items were imputed using hot-deck imputations (see Appendix 1 to Chapter 12).

26. What percentage of classroom time is devoted to teaching each of the following kinds of text?

	Percent
a. narration	________
b. exposition	________
c. documents	________
	100%

12.2.3. Student Questionnaires

Although they have much in common, there are differences between the fourth and ninth grade Student Questionnaires and the nonresponse rates for each. To accommodate those differences, we look at the fourth and ninth grade Student Questionnaires separately (Tables 12-4 and 12-5, respectively).

The highest levels of nonresponse in the tables are where one would predict them to be--in the reporting of parental educational attainment, in the list pertaining to persons living in the household, and for students of an age where homework is not regularly assigned by schools, who helps with homework. The questions directed at non-English-speaking students also show response rates somewhat higher than the modal values, but still less than 20 percent.

Table 12-4. Percentage of items with over 5 percent missing data: Grade 4 Student Questionnaire

Item	Percent of items
5 Number years father's education	12
6 Number years mother's education	10
11i Home possession/recreational vehicle	10
14 In household, male guardian	13
Female guardian	13
Brother(s)	8
Sister(s)	8
Grandparent(s)	11
Other relative(s)	12
Non-relative(s)	12
17 Other language, understand	10
18 Other language, speak	11
19 Other language, read	11
20 Other language, write	11
21 Other language, classes	12
22 Use other language-parents	16
Siblings	18
Friends	18
Relatives	17
23 Read to in other language at home	14
24 Read to in other language elsewhere	14
25 English, used in school work	15
26 Speak English at home	15
27 Understand English	16
28 Speak English	15
29 Read English	14
30 Write English	14
43 Before reading, remember	6
Guess	7
Somebody else	8
44 While reading, make notes	7
Think	7
Read over	8
Guess	7
45 After reading, write notes	6
Related selection	6
New ideas	6
Somebody else	6
Write	6
53 Homework, reading and writing	7
Reading only	9
Choose reading & report	9
Reading choice	10
Other	30
54 Homework help, mother	8
Father	13
Sibling	15
Tutor	18
Other	17
56 Frequency reading books for fun	6
58 Frequency reading comics	6
59 Reading a magazine last week	6
60 Frequency reading a magazine	7
61 Reading a newspaper last week	6
62 Frequency reading newspapers	7
63 Frequency reading directions	7
64 Frequency read textbooks in school	7
65 Frequency reading story books in school	7
66 Frequency reading workbooks in school	7
67 Frequency practice exercises	7
68 Frequency looking up information	7

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 12-5. Percentage of items with over 5 percent missing data: Grade 9 Student Questionnaire

Item	Percent of items
20 In Household	
b male guardian	15
d female guardian	16
e brother(s)	10
f sister(s)	11
g grandparent(s)	18
h other relatives	17
i non-relatives	18
46 Homework	
a reading and writing	6
b reading only	11
c choose reading & report	12
d reading choice	12
e other	35
47 Homework help	
a mother	6
b faiher	9
c sibling	12
d tutor	16
e other	13
61 Own writing	
a poetry	11
b diary	12
c letters	7
d messages	11
e stories	12
f computer programs	12
g other	30

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Other Category Items. Question 53 is analogous to this type of item on both the Principal and Teacher Questionnaires. As expected, it is particularly problematic for fourth grade students. If we set aside the 30 percent nonresponse to the residual category of question 53, nonresponse does not exceed 20 percent on any question.

53. Which kinds of reading work do you normally do for homework?

	Yes	No
a. we read and write answers to the teacher's questions	1	2
b. we read but do not have questions to answer	1	2
c. we choose what to read and report back to the teacher or class	1	2
d. we choose what to read but do not report back to the class	1	2
e. other (specify)	1	2

List Checking Items. The remaining questions tend to be, as before, elements of multipart questions. Questions 11, 14, 53, and 54 follow the same format as question 28 on the Principal Questionnaire. With fourth graders, it is even more likely that students would ignore the direction to circle one response on each line and would mark answers only on those items that were in the affirmative.

43. How often do you do each of the following before you begin to read? (Circle one number on each line.)

	Never	Once in a while	Quite often	Most of the time
a. look at the title, illustrations and...	1	2	3	4
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
f. pinpoint issues that you are interested in exploring	1	2	3	4

Questions 43, 44, and 45 follow the same format, a multipart question with an estimate of frequency. This combination may be beyond the capabilities of a number of fourth grade students, who then revert to circling only those things that they do on a regular basis.

Note that with regard to questions 18 through 30, the majority of students bypass these questions through a skip pattern. The questions themselves are designed for students from non-English-speaking backgrounds.

Other Category Items. As before, the highest levels of nonresponse are recorded for items that are residual "other" categories in multipart questions (questions 20h, 46e, 47e, and 61g), with the remainder of the items showing less than 20 percent nonresponse.

List Checking Items. As we have seen before, most of the nonresponse is located in multipart questions (20, 46, 47, and 61) requiring either a "yes" or "no" answer. Respondents seem to consistently answer only those parts of the question that they can answer affirmatively, leaving the rest blank.

12.2.4. The Source of Items Having Greater Than Five Percent Nonresponse Rates

In the preceding discussion, we looked in some depth at the relatively small number of survey questions where item nonresponse rates exceeded 5 percent. The history regarding how those items came to be included in the main study instruments is summarized in Table 12-6.

A reader of this table should note that there are two kinds of items -- U.S. national and international. For the purposes of this discussion, the essential difference between the two item types is the degree of control that the U.S. team had over the item construction. In the case of U.S. national items, the U.S. Steering Committee decided to include those items, and the committee approved the wording. In the case of international items, whose inclusion in the questionnaire is obligatory, the International Steering Committee had control over the content and wording of the items.

Table 12-6. Item history: Grade 4 and Grade 9 questionnaires[1]

Test history	U. S. National		International	
Grade 4 School Questionnaire				
pilot tested			*9*	*number of books in school library*
			28	*teacher evaluation procedures*
not pilot tested	3	race/ethnicity[2]		
	20	satisfaction with assessment information		
	37	number of courses in language arts		
Grade 9 School Questionnaire				
pilot tested			35	years as a principal
			9	*number of books in school library*
			10	*number of books added in the last year*
			14	*total instructional time*
			28	*teacher evaluation procedures*
not pilot tested	19	tests for evaluation		
	20	satisfaction with assessment		
	37	courses in language arts		
Grade 4 Teacher Questionnaire				
pilot tested	8	time devoted to learning about reading	35	instructional strategies
			7	*teacher education courses*
			32	*texts per student*
			34	*reading aims*
			39	*grouping*
			44	*encourage students*
not pilot tested	29	time teaching ESOL		
	42	teaching text types		
Grade 9 Teacher Questionnaire				
pilot tested			*7*	*teacher education courses*
			27	*encourage students*
			30	*reading aims*
			34	*different texts per student*
			44	*percent/other topics*
not pilot tested	26	teaching text types	24	assessment method
Grade 4 Student Questionnaire				
pilot tested	*43*	*reading before*	11	home possession
	44	*reading while*	23	read in other language
	45	*reading after*	24	read in English
	53	*homework assigned*	56-68	reading for fun
	54	*homework who helps*	*5*	*parental education*
			6	*parental education*
not pilot tested	14	persons in the household[3]		
	17-22	facility in other language[3]		
	25-30	English facility[3]		
Grade 9 Student Questionnaire				
pilot tested	*46*	*homework*		
	47	*homework who helps*		
	61	*own writing*		
not pilot tested	20	persons in the household[3]		

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

---

[1] Items in *italics* were piloted tested and revised based on information gathered during pilot testing.

[2] This item was taken from the NCES School and Staffing Survey.

[3] This item was taken from the NELS for 8th graders.

Also note that not all items included in the main study were pilot tested in their final form. Items that appear in *italics* were revised after the pilot testing and were not further tested prior to the main study. Some additional items were never pilot tested as part of this study, although they had been included in other surveys conducted by NCES.

Of the 68 items having nonresponse rates greater than 5 percent, 36 were designed and worded by the International Steering Committee. Half of these items were revised by the International Steering Committee based on information gathered during the pilot test. However, they were not field tested again before inclusion in the main study. In effect, these items function as new items where our "best guess" was the basis for inclusion.

In contrast, the international items that were pilot tested (question 35, ninth grade School Questionnaire; question 35, fourth grade Teacher Questionnaire; questions 11, 23, 24, 56-68, fourth grade Student Questionnaire) and used as worded did not previously appear to be problematic because in the pilot test, the nonresponse rates varied from 0 percent to 7 percent in the United States. Why then did the nonresponse rate increase in the main study? The best explanation available is the difference in the samples used for the pilot and main studies. For the pilot study, we used a much smaller judgment sample in which nonresponse across all parts of the instruments was low. In contrast, for the main study, the sample size was four to six times larger, and the time schedule for data collection was tighter, which may have led to the greater variability in nonresponse across items.

The remaining 32 items with nonresponse rates greater than 5 percent were controlled by the U.S. Steering Committee. Of these, only one item that had been pilot tested and had a high nonresponse rate (33 percent) during the pilot was included in the same format. This was question 8 on the fourth grade Teacher Questionnaire, asking teachers to provide an estimate of the amount of time during their preservice training devoted to learning about reading instruction, which was included because the U.S. Steering Committee believed that it was extremely important to differentiate between general teacher training and training specifically targeted on reading. Similarly, question 37 on the Principal Questionnaire was added so that we could look at issues of instructional supervision.

The U.S. Steering Committee was particularly interested in the impact of second language learning on reading achievement. However, this interest was not identified until our review of the pilot test data. To gather data on students for whom English was a second language, we looked for items that had previously been used in other NCES surveys. The NELS instruments designed for eighth graders provided the best available model. Therefore, we chose to take a chance with these items. What is apparent from the nonresponse rates is that these items are fine for the intended eighth graders, and worked well with the ninth graders included in our study. However, the questions had nonresponse rates that exceeded 5 percent for fourth graders. This leads us to believe that these items should be reworded, reformatted, or both if they are to be used with students below eighth grade in the future.

The U.S. Steering Committee was also interested in preserving a number of items that had been part of the international instruments during pilot testing. While a number of these items did have high nonresponse rates internationally as well as in the U.S. during pilot testing, the Steering Committee believed that with revision we might be able to capture good data on issues of importance. For example, questions 43, 44, and 45 on the fourth grade Student Questionnaire represent the only available measure of a schema-based approach to reading instruction. This theoretic stance represents a major thrust in U.S. instruction during the last 10 to 15 years and should be included in our analyses.

In conclusion, the overall item nonresponse rates were low. Where they were high, it was in places that were predictable. This was best exemplified by the fourth grade Student Questionnaire, in which students had difficulty with items of parental education, those that were long multiple-part questions, and those with skip patterns. We have chosen to impute for all items with missing data to improve estimates that we will include in our study.

12.3. Choice of Imputation Methods

This section is a review of available methods of imputation and our selection of methods. We briefly discuss the concept that most imputation techniques can fit under the framework of a regression model. The main difference between methods is the way they treat the residuals in the model. Models that include a residual term are called stochastic; those without are called deterministic. This study used a stochastic procedure, hot-deck imputation, for most of the items, but in some cases deterministic procedures were chosen as more appropriate. We begin with a brief overview as to why it is desirable to impute missing data.

12.3.1. Problems That Result From Missing Item Data

As research indicates, item nonresponse may affect the precision of estimates as well as yielding biased estimates. Estimates from items with missing data may be biased because the answers from respondents may not be representative of the answers from nonrespondents. Therefore, the observed responses may not truly represent the target population. There is also some loss of precision because the effective sample size is reduced when cases with missing data are ignored.

Because of the problems of bias and reduced precision, analysts of a dataset with item nonresponse are confronted with a number of pertinent questions. The most prominent is, what is the effect of the nonresponse bias on various statistics? Kalton (1983) has shown that the effect of nonreponse on different statistics, such as population means, totals, variances and covariances, can be fairly diverse. Further, for univariate estimates of a single item (such as mean, total, and proportions), the loss of precision depends on the amount of missing data in that item. For bivariate statistics, such as the correlation between two variables, the loss of precision depends on the joint amount of missing data in both items. For multivariate statistics, like regression coefficients, factor loading, and canonical correlations, the extent of the loss is the cumulative amount of missing data from the set of variables being analyzed.

In typical situations, an item nonresponse rate of 10 percent or lower may not result in substantial loss of precision or bias. These typical situations are predominant in conducting univariate analyses. However, in more complex analyses, which generally will require joint distributions of variables, these otherwise low levels of item nonresponse may become problematic. In the case of the IEA Reading Literacy Study, we are primarily interested in more complex statistical analyses, such as factor analysis, multivariate linear regression, and hierarchical linear modeling. Therefore, our aim is to prevent loss of precision and introduction of bias. To do so, we have considered various ways of compensating for missing item data.

Traditionally, analysts consider whether to apply pair-wise or list-wise deletions to handle missing item data. Pair-wise deletion (which uses all the nonmissing pairs of values for each pair of variables) is generally preferable, since it results in less data loss for a given analysis. In certain situations (e.g., estimating factor scores) list-wise deletion, which uses an observation only if none of the variables is missing, would be more appropriate since factor scores will not be estimated for a case with any missing item data. However, list-wise deletion can lead to a substantial loss of information. An analyst using list-wise deletion may be throwing away observations that are 90 percent complete.

For the IEA Reading Literacy Study, the amount of missing data was judged to be small (see Section 12.2). Thus, the loss of precision of estimates and bias associated with estimates in general, and univariate and bivariate statistics in particular, should be quite small. For multivariate statistics and estimates of constructs, however, the loss in precision and bias may not be negligible. Because multivariate analyses and estimation of constructs typically involve large numbers of variables (e.g., 10 or more variables), the effect of item nonresponse in this case will be compounded.

The derivation of the construct family wealth for fourth grade students illustrates this point. This construct was derived from 11 variables (Q11A-A11E, Q11G, Q11H, Q11J-Q11M), and the item nonresponse for this set of items ranges from 2 to 5 percent. Since missing data for any of these 11 variables will result in a missing value for family wealth, the percentage of cases with missing data on family wealth combined is 18 percent. Thus, there is a significant increase in the amount of missing data for the construct family wealth as compared to the missing values for the items that constitute family wealth.

A more appropriate alternative to pair-wise or list-wise deletions in this instance is to impute values for items with missing data. In the next section, we elaborate on approaches for imputation.

12.3.2. Unifying Framework for Various Imputation Methods

A wide variety of imputation methods can be used to assign values for missing item responses in surveys. Kalton and Kasprzyk (1982), for instance, list nine different forms of imputation that include deductive imputation, overall or class mean imputation, random imputation, various hot-deck procedures, regression imputation, and distance function matching. While these procedures appear fairly diverse, the authors show that they can nearly all fit within the general structure of a regression model.

Let us suppose that the item to be imputed is the number of books in the school library (from the Principal Questionnaire). A simple regression model can be written to predict number of books in the school library from student enrollment as follows:

$$y_i = \beta z_i + \varepsilon_i \qquad (1)$$

where y_i is the number of books for the ith sampled school, z_i is the number of students enrolled in the school, β is the regression coefficient of y on z, and ε_i is the error term (or residual from the predicted values) for the ith sampled school.

When the value of y is missing from a sampled unit, it can be replaced by its imputed value, $\hat{y}_i$. We can estimate the regression coefficients, b, from the respondents and determine what, if any, residual to add to the predicted value. The imputed value is thus

$$\hat{y}_i = bz_i + e_i \qquad (2)$$

where the residual, e_i, may be set to zero or may be determined in some other way.

12.3.3. Stochastic Versus Deterministic Imputations

Under the framework of a regression model, imputation methods differ in the ways they treat the error term. If the residual, e_i in equation (2), is set to be zero, the method is a deterministic procedure. If e_i is chosen randomly according to some specific procedure, then the method is stochastic. Sometimes, the auxiliary variables used in the regression model are dummy variables indexing subgroups of the sample (e.g., public or private schools, regions of the country). In this case, the predicted values for the regression are the subgroup means. Thus, the regression model covers the mean within class imputation as a special case. If a residual is chosen randomly from the set of all observed residuals from a particular subgroup, the regression model produces an imputation procedure that is like hot-deck imputation.

For general purpose imputations, a stochastic procedure is preferred over the deterministic approach, because it preserves the distribution of the data as observed from respondents. A deterministic imputation scheme assigns values on the regression line, and hence does not reflect the residual variation around the line. In consequence, deterministic imputations attenuate the variance of the distribution subject to imputation and distort the shape of the distribution. However, in some circumstances a deterministic approach is preferable. For example, if the researcher has a model for the nonresponse mechanism that indicates that a certain type of nonrespondent is very likely to have a particular true value, then it may be preferable to impute that value directly for such respondents.

12.3.4. Imputation Methods Used in the IEA Reading Literacy Study

In this study, we employed a combination of a hot-deck imputation procedure and deterministic imputations to assign values for missing responses for the data items. The general structure of hot-deck procedures is to define imputation classes according to the cross classification of the auxiliary variables chosen for use in imputing for missing responses to a particular item, and then to assign for a missing response the value from a respondent in the same imputation class. Hot-deck imputation procedures correspond roughly to a regression model in which the z's are dummy variables that specify the classes, and in which the stochastic term is a residual chosen from one of the respondents in the same imputation class. Expressed in terms of the auxiliary variables, the regression model for hot-deck imputation includes not only the main effects but also all the interaction terms between these variables. The hot-deck procedure implemented in this study used WESDECK, a SAS macro developed by Westat.

Hot-deck imputation procedures were used to handle missing responses for most items. For some of the remaining items, the missing responses were completed from information available in other data sources; for some items, it was possible to deduce the missing response from the responses to other items on the questionnaire; and for other items, the overall modal response for respondents was assigned for all missing responses. The latter technique, which was employed for operational expediency, was used only when the item nonresponse rate was very small.

The methods used to impute each data item on the questionnaires are shown in Appendix 1 to Chapter 12, tables 1 through 6. In this section, we describe the process of how these imputations were conducted and our monitoring of the quality of the outcome, and give examples drawn from the appendices, which are available from Westat.

12.3.4.1. Hot-Deck Imputations Using WESDECK

WESDECK begins by sorting the data files sequentially, according to a list of variables stated in the SORTVAR option. These variables define the imputation classes, some of which are defined as critical while others are not. With a critical class, a respondent donor has to be found within that class. With a noncritical class a search is made for a donor within that class, and if one or more donors are found, one of them is used; if none is found, then a search is made for donors in the next class. WESDECK processes the sorted data file in order. Values from records with valid data are stored in a donor pool, which can hold up to three donors at a time. The donor pool is continuously updated; when an imputation class had three records and a further record from that class is encountered, the new record replaces the oldest record in the donor pool. When a record with a missing response to the item (denoted by the variable indicated by the IMPVAR keyword) is encountered, one of the values stored in the donor pool for the appropriate class is assigned.

The advantages of WESDECK are that it is tested, relatively fast, easy to use, and can impute more than one variable at a time. If an imputee is missing a value to any of the variables in the list indicated by the keyword ADDVAR, then the donor value for each such variable is imputed also. This capability is especially useful given the many variables to be imputed in this study. When several variables are to be imputed in one run, we specified that WESDECK would use as donors only records with complete information on all variables subject to imputation. When a record had more than one missing value in the set of items for which joint imputations were being made, WESDECK replaced all the missing values from the same donor.

The Imputation Classes

The imputation classes defined for each hot-deck imputation were formed to meet two goals. First, we formed classes such that we felt that it was reasonable to treat the missing responses as missing at random (MAR) within classes. Although required for unbiased estimation, the MAR assumption is seldom tested in practice (see Little and Rubin 1987 for a discussion of this assumption).

Second, imputation classes were constructed in a way that produced fairly homogeneous values for the item within each class. Known relationships reported in the literature, expert opinions, and a small amount of exploratory analyses of the dataset were used to help form homogeneous imputation classes. The imputation classes were also chosen to take account of the edit constraints. For instance, in the imputation for missing responses on the gender of a student, individual schools were used to form imputation classes so that a response of "boy" would not be imputed for a student in a girls' school.

Monitoring the Hot-Deck Imputations

For each hot-deck imputation, the following attributes were carefully monitored:

- How many times each record was used as a donor;
- Which donor was involved in a particular imputation;
- Were all records with missing data imputed in one run; and
- How many runs were needed to complete the imputation.

We specified in WESDECK that a donor's value should be used no more than three times. Because of this constraint, some missing responses might not have been imputed because of a lack of donors. When this happened, we adjusted the sort sequence of the file to locate new donors. This adjustment was required only in a couple of imputations.

An Example of Hot-Deck Imputation with Multiple Variables in the IEA Reading Literacy Study

Consider the data from questions 8 and 10 of the fourth grade Student Questionnaire. The items are shown below, along with their response frequencies.

8. Does your family regularly get or see a newspaper at home? (Circle one only.)

		FREQUENCY	
No	1	1,305	(19.6%)
Yes	2	5,292	(79.6%)

For 49 cases (0.7%), no response was given.

10. About how many books are there in your home? (Do not count newspapers, comic books, or magazines; circle one only.)

		FREQUENCY	
None	1	121	(1.8%)
1-10	2	495	(7.4%)
11-50	3	1,128	(17.0%)
51-100	4	1,358	(20.4%)
101-200	5	1,172	(17.6%)
More than 200	6	2,304	(34.7%)

For 68 cases (1.0%), no response was given.

There were 26 cases missing responses to both questions 8 and 10, 23 missing responses to questions 8 only, and 42 missing responses to question 10 only.

Because it was expected that the answers to these two questions would be highly correlated, and it was considered important to preserve that correlation in the imputed data for analyses, responses were imputed simultaneously for the missing data from these questions. The sort variables used were parents' education, school type, race, the response to question 13 (How many people live in your home?), community size, and school. The imputation was constrained so that a "donor" of an imputed value always had the same value for parents' education and school type as the case that required imputation.

By sorting on the other variables listed as well, we attempted to get a donor that was as similar as feasible to each imputee on these characteristics, but there was no constraint that the values of these other variables had to be identical.

The imputation proceeded by imputing a response as needed for questions 8 and 10 for the 91 cases that were missing at least one of these responses. All donors had complete data for both questions. For the 26 cases missing both responses, the pair of responses was imputed from a single donor. This helped to preserve the correlation between the two sets of responses. For the 23 cases missing a question 8 response, the donor provided a response only for question 8, with the imputee retaining its original response for question 10. The 42 cases that were missing a question 10 response only were handled in a similar manner. The variable for the question 8 response is termed WASNEWS in the Chapter 12 appendices, while the variable for the question 10 response in WASBOOKS.

The table below shows the summary of the results of imputation for these two questions. The frequency distributions for the two questions before and after imputation are shown.

Q8 Daily Newspaper at Home

	Before Imputation		After Imputation	
No	1,305	(19.6%)	1,318	(19.8%)
Yes	5,292	(79.6%)	5,328	(80.2%)
Missing	49	(0.7%)	----	

Q10 Number of Books at Home

	Before Imputation		After Imputation	
None	121	(1.8%)	122	(1.8%)
1-10	495	(7.4%)	503	(7.6%)
11-50	1,128	(17.0%)	1,140	(17.2%)
51-100	1,358	(20.4%)	1,377	(20.7%)
101-200	1,172	(17.6%)	1,181	(17.8%)
201+	2,304	(34.7%)	2,323	(35.0%)
Missing	68	(1.0%)	----	

In each case, the reader can see that the imputed responses have been spread across the range of possible responses via the hot-deck procedure. No given donor was used more than three times to impute, and just one case was used as a donor three times. Thus, the imputed responses were derived from a broad range of donors, which helps to ensure that the imputation scheme does not add substantial response variance to analysis.

12.3.4.2. Other Imputations

In addition to hot-deck imputations, three other forms of imputation were used with the U.S. samples of the IEA Reading Literacy Study: use of external sources of data, deductive imputations, and modal imputations.

Use of External Data

For some items we located data from other data sources from which the missing data could be completed. For example, some principals did not report enrollment data for their schools (Questions 1-3 of the Principal Questionnaire). We used the Quality Education Data, Inc., data files to provide the missing information.

For some data items, we checked hard copies of the questionnaires for additional information. For example, some schools did not report metropolitan status of the community in which the school was located. For these schools, the hard copies of the questionnaires were retrieved and the address of the school was used, in combination with Census Bureau tables, to derive the missing information.

Deductive Imputation

In several cases, the values for the missing responses to an item were deduced from responses to other items on the questionnaire. This method was often used with skip patterns and incomplete responses to items with multiple parts. For example, if a ninth grade student reported having no regular jobs (question 8) but failed to answer the question about the time spent on jobs (question 9), then the nonresponse to the latter was imputed as "not applicable."

Combination of Deductive Imputation and Hot-Deck Imputation for a Single Item

For some questions with multiple parts, there were responses with missing data for some parts only. In these cases of partial nonresponse to the question, responses for the missing parts were imputed as "no," "not applicable," or "never," as appropriate. Hot-deck imputation was used for cases of full nonresponse to the question. This type of question was associated with a substantial proportion of the total amount of missing data and is discussed extensively in Section 12.2.

An example was the imputation for question 43 of the fourth grade Student Questionnaire. This question asked about the frequency with which a student did a list of activities before reading (i.e., read the title, think about the topic). For students who responded to some of the activities on the list but failed to respond to others, the nonresponses were imputed to indicate "never" (i.e., they never did the activity before reading). The rationale for this decision was that such students interpreted the question to mean that they should indicate a response only for those activities that they sometimes undertook. For students who left the entire question unanswered, the hot-deck procedure was used to impute a response for each of the six activities listed from among the pool of respondents. All six parts of the question were imputed using a single donor for each nonrespondent in this case.

The table below shows the format of question 43 from the fourth grade Student Questionnaire and the frequency distribution prior to any imputation.

Question 43. How often do you do each of the following before you begin to read? (Circle one number on each line.)

		Never	Once in a while	Quite often	Most of the time	Missing
a.	Look at the title, illustrations and heading to find out what it is likely to be about.	932 (14.0%)	1,925 (29.0%)	1,068 (16.1%)	2513 (37.8%)	208 (3.1%)
b.	Think about what you already know about the topic.	1,756 (26.4%)	2,228 (33.5%)	1,318 (19.8%)	1,000 (15.0%)	344 (5.2%)
c.	Remember other selections about the same topic.	1,791 (26.9%)	2,088 (31.4%)	1,376 (20.7%)	984 (14.8%)	407 (6.1%)
d.	Try to guess what will happen or what information you might learn.	1,179 (17.7%)	1,657 (24.9%)	1,349 (20.3%)	2,008 (30.2%)	453 (6.8%)
e.	Talk to somebody else about it.	2,276 (34.2%)	1,832 (27.6%)	983 (14.8%)	1,004 (15.1%)	551 (8.3%)
f.	Pinpoint issues that you are interested in exploring.	1,787 (26.9%)	1,831 (27.6%)	1,177 (17.7%)	1,489 (22.4%)	362 (5.4%)

There were 136 cases in which all six responses were missing. The missing values associated with other cases (ranging in number from 72 for part a to 415 for part e) were set to a value of 1 (a response of "never"). The 136 cases were imputed using the hot-deck procedure. The use of a single donor (having complete response to question 43) for all six missing responses in a given case of full nonresponse for the question is indicated in Table 5 of Chapter 12 Appendix 1 by the notation reference to multiple variables in one hot-deck, Q43A,...., Q43F. The cases (excluding the 720 cases with partial missing data that were imputed deterministically) were sorted prior to imputation by the variables school type, parents' education, number of books in the home (collapsed into a reduced set of classes), the response to question 37 (Are you in a special class to help you read at your grade level?), and the response to question 35 (How good are you at reading?). The imputation was constrained so that the donor always belonged to the same school type and parents' education class as the imputee. Table 12-7 shows the percentage distribution of responses to each part of the question before and after the combined steps of imputation. For comparisons, the distribution of responses before imputation was computed based on observed cases only.

Table 12-7. Percentage distribution of responses to question 43 of the grade 4 Student Questionnaire before and after imputation

Part of question	Before imputation				After imputation			
	Never	Once in a while	Quite often	Most of the time	Never	Once in a while	Quite often	Most of the time
a	14.5	29.9	16.6	39.0	15.5	29.7	16.4	38.5
b	27.9	35.4	20.9	15.9	30.2	34.2	20.1	15.5
c	28.7	33.5	22.1	15.8	31.7	32.1	21.1	15.1
d	19.0	26.8	21.8	32.4	22.8	25.7	20.6	30.8
e	37.3	30.1	16.1	16.5	41.2	28.1	15.1	15.6
f	28.4	29.1	18.7	23.7	30.9	28.3	18.1	22.7

NOTE: Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Clearly in this case a large proportion of the imputed response is in the "never" category, as a result of the deterministic rule that applied in the cases of partial missing data.

Modal Imputation

The modal imputation was used with some items with a small amount of missing data. With this procedure, all records with missing values for the item were assigned the modal value of the respondents. The advantage of this deterministic procedure is that it is easy to apply. However, the disadvantage is that it distorts the shape of the distribution of the items. It was used for convenience in situations where it would not cause any appreciable distortions. Its use was therefore reserved for items for which the item nonresponse rate was low and for which the modal category included a high proportion of the respondents. It was mainly used in the Teacher Questionnaires, in which the item nonresponse rates were about 1-2 percent. This method was not used with student data.

An example of the application of modal imputation is given by question 44 of the fourth grade Teacher Questionnaire (see Table 12-8). This item asked whether teachers regularly do a list of activities to encourage their students to read outside school. The response categories were either a "yes" or "no." Only one or two teachers failed to respond to the parts of the item, and teachers who responded had a strong tendency of selecting the same option.

Table 12-8. Examples of modal imputation: Percentage distribution of responses to question 44 of the grade 4 Teacher Questionnaire before and after imputation

44. Do you regularly (i.e., at least once a week) do the following activities to encourage your students to read outside school?					
Activity	Percent before imputation			Percent after imputation	
	No	Yes	Missing	No	Yes
a) Suggest books (to student) to read	18.3	81.4	0.3 (1 case)	18.3	81.7
b) Suggest newspaper articles to students to read	54.2	45.4	0.3 (1 case)	54.7	45.4
c) Read stories to students	8.2	91.5	0.3 (1 case)	8.2	91.8
d) Hold discussions about books	32.7	66.7	0.7 (2 cases)	32.7	67.4
e) Encourage students to borrow library books	1.6	98.0	0.3 (1 case)	1.6	98.3

NOTE: Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

12.3.4.3. Usage of Different Imputation Methods

Table 12-9 shows the relative extent to which the different imputation procedures were used with the IEA Reading Literacy Study data. It can be seen that for the student data, which reflected the greatest extent of item nonresponse, hot-deck imputation was used extensively, often in conjunction with a deductive approach for partial nonresponse to a multipart question, as discussed above. For the Principal and Teacher Questionnaires, other approaches were used to a greater extent. Again, it can be seen that the hot-deck procedure was the primary method of imputation for these types of questions. Tables 12-10 through 12-13 show the methods used for each item having more than 5 percent missing data. Chapter 12 Appendix 1 tables 1-6 show the imputation methods for all items below and above 5 percent.

Table 12-9. Number and percentage of items imputed by various imputation methods: Grades 4 and 9

Imputation method	Percent imputed			
	Grade 4		Grade 9	
	Number	Percent	Number	Percent
Student Questionnaires				
Hot-deck imputation	84	63	143	59
Hot-deck and deduction	45	34	77	32
Deductive imputation	5	4	15	6
Use external data	0	0	0	0
Modal imputation	0	0	0	0
No imputation required	0	0	6	3
Teacher Questionnaires				
Hot-deck imputation	30	12	23	15
Deductive imputation	45	18	24	16
Use external data	0	0	0	0
Modal imputation	159	64	88	58
No imputation required	16	6	16	10
Principal Questionnaires				
Hot-deck imputation	29	26	39	33
Deductive imputation	41	36	41	35
Use external data	10	9	11	9
Modal imputation	12	11	15	13
No imputation required	21	18	11	9

NOTE: Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Consider as examples two questions discussed in Section 12.2. Question 3 from the School Questionnaire for grade 4 asks about distribution by race/ethnicity of the fourth grade. Each of the five categories had 11 percent nonresponse. Information was available from the survey frame about the race/ethnicity distribution for the whole school, as of about 1989. These data were used to impute the missing cases. This is indicated in Table 12-10 by the notation of "external data" as the method of imputation.

Table 12-10. Items with over 5 percent missing data in the grade 4 and grade 9 Principal Questionnaires

Item		Grade 4		Grade 9	
		Percent missing	Method of imputation	Percent missing	Method of imputation
Q3A	RACE/ETHN:ASIAN, PAC ISLAND	11	External data		
B	RACE/ETHN:AM IND, ALASKAN	11	External data		
C	RACE/ETHN:HISPANIC	11	External data		
D	RACE/ETHN:WHITE (NON-HISP)	11	External data		
E	RACE/ETHN:BLACK (NON-HISP)	11	External data		
Q9	SCHL LIB/ N BOOKS CONTAINED	7	Hot-deck	6	Hot-deck
Q10	SCHL LIB/ N BOOKS ADDED LAST			7	Hot-deck
Q14H	INSTRUCT TIME/WEEK, HOURS			6	Hot-deck
Q19B	TEST EVAL/CURRICULUM			7	Hot-deck
C	TEST EVAL/TEACHERS	8	Hot-deck	7	Hot-deck
D	TEST EVAL/TEXTBOOKS			7	Hot-deck
Q20A	SATISF,NORM-REF TEST SCORES	7	Hot-deck	9	Hot-deck
B	SATISF,CRITERION-REF SCORES	17	Hot-deck	24	Hot-deck
C	SATISF,STUD WORK SAMPLE			10	Hot-deck
D	SATISF,TEACHER JUDGMENT			8	Hot-deck
E	SATISF,GRADE REPORT			10	Hot-deck
Q28A	PROCEDURES/INTERVIEWS	7	Hot-deck		
B	PROC/SELF REPORTS BY TCHRS	10	Hot-deck		
D	PROC/STUDENT RATINGS	12	Hot-deck	8	Hot-deck
E	PROCEDURES/OTHER	13	Hot-deck	8	Hot-deck
Q35T	NO. YEARS PRINCIPAL, CAREER			6	Hot-deck
Q37	COURS IN ENG/LANG ARTS/READ	7	Hot-deck	7	Hot-deck

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 12-11. Items with over 5 percent missing data in the grade 4 and grade 9 Teacher Questionnaires

Item		Percent missing	Method of imputation
Grade 4			
Q7	N OF TEACHER ED COURSES	23	Hot-deck
Q8	PCT OF TIME LEARN TCHG OF READ	6	Hot-deck
Q29H	TIME TEACH ESOL/HOURS	14	Hot-deck
M	TIME TEACH ESOL/MINUTES	14	Hot-deck
Q32	DIFF TEXTS PER STUDENT	13	Mean within class
Q34C	RATE AIMS/READING COMPREH	6	Deductive imputation
E	RATE AIMS/EXTEND VOCAB	6	Deductive imputation
F	RATE AIMS/CRITICAL THINKING	6	Deductive imputation
H	RATE AIMS/DEEPEN EMOT DEVEL	6	Deductive imputation
I	RATE AIMS/WORD ATTACK SKILLS	6	Deductive imputation
K	RATE AIMS/EXPAND READ CHOICE	6	Deductive imputation
L	RATE AIMS/MAKE READING ENJOY	6	Deductive imputation
Q35D	STRATEGIES/GRADED TEXT DIFF	7	Modal imputation
Q39A	FREQ USE AGE GROUPS	14	Deductive imputation
C	FREQ USE INTEREST GROUPS	12	Deductive imputation
D	FREQ USE OTHER GROUPS	41	Deductive imputation
Q42A	PERCENT TEACH NARRATION	6	Hot-deck
B	PERCENT TEACH EXPOSITION	6	Hot-deck
C	PERCENT TEACH DOCUMENTS	6	Hot-deck
Q44F	ENCOURAGE STUD/OTHER	32	Deductive imputation
Grade 9			
Q7	N OF TEACHER ED COURSES	15	Hot-deck
Q24A	ASSESS METH/TEACHER QUIZZES	9	Deductive imputation
B	ASSESS METH/MULTIPLE-CHOICE	8	Deductive imputation
C	ASSESS METH/STUD INTERESTS	7	Deductive imputation
D	ASSESS METH/ORAL DISCUSS	10	Deductive imputation
E	ASSESS METH/DISCUSS MAT READ	10	Deductive imputation
F	ASSESS METH/OPEN-ENDED QUES	10	Deductive imputation
G	ASSESS METH/ESSAYS ABOUT LIT	9	Deductive imputation
Q26A	PERCENT TEACH NARRATION	10	Hot-deck
B	PERCENT TEACH EXPOSITION	10	Hot-deck
C	PERCENT TEACH DOCUMENTS	10	Hot-deck
Q27D	ENCOURAGE STUD/READ ASSIGN	6	Hot-deck
Q30A	RATE AIMS/LASTING INTEREST	9	Hot-deck/deductive
B	RATE AIMS/READING COMPREHEN	7	Hot-deck/deductive
C	RATE AIMS/RESEARCH/STUDY SKI	10	Hot-deck/deductive
D	RATE AIMS/EXTENDING VOCAB	10	Hot-deck/deductive
E	RATE AIMS/CRITICAL THINKING	8	Hot-deck/deductive
F	RATE AIMS/EXPAND WORLD VIEW	8	Hot-deck/deductive
G	RATE AIMS/INCR SPEED OF READ	12	Hot-deck/deductive
H	RATE AIMS/EXPAND CHOICE	8	Hot-deck/deductive
I	RATE AIMS/STRAT TO OTH SUBJ	10	Hot-deck/deductive
J	RATE AIMS/APPREC OF LIT	8	Hot-deck/deductive
K	RATE AIMS/INTERPRET DIAGRAMS	6	Hot-deck/deductive
Q34	DIFF TEXTS/STUDENT	14	Median within class
Q44FP	PERCENT/OTHER TOPICS	6	Deductive imputation

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 12-12. Items with over 5 percent missing data in the grade 4 Student Questionnaires

Item	Percent missing	Method of imputation
Q5 N YEARS FATHERS EDUCATION	12	Hot-deck/deduction
Q6 N YEARS MOTHERS EDUCATION	10	Hot-deck/deduction
Q11I HOME POSSESS/REC VEHICLE	10	Hot-deck
Q14B IN HOUSEHOLD, MALE GUARDIAN	13	Hot-deck/deduction
D IN HOUSE, FEMALE GUARDIAN	13	Hot-deck/deduction
E IN HOUSEHOLD, BROTHER(S)	8	Hot-deck/deduction
F IN HOUSEHOLD, SISTER(S)	8	Hot-deck/deduction
G IN HOUSEHOLD, GRANDPARENT(S)	11	Hot-deck/deduction
H IN HOUSE, OTHER RELATIVE(S)	12	Hot-deck/deduction
I IN HOUSE, NON-RELATIVE(S)	12	Hot-deck/deduction
Q17 OTHER LANGUAGE, UNDERSTAND	10	Hot-deck (n=628)
Q18 OTHER LANGUAGE, SPEAK	11	Hot-deck (n=628)
Q19 OTHER LANGUAGE, READ	11	Hot-deck (n=628)
Q20 OTHER LANGUAGE, WRITE	11	Hot-deck (n=628)
Q21 OTHER LANGUAGE, CLASSES	12	Hot-deck (n=628)
Q22A USE OTHER LANGUAGE-PARENTS	16	Hot-deck (n=628)
B USE OTHER LANGUAGE-SIBLINGS	18	Hot-deck (n=628)
C USE OTHER LANGUAGE-FRIENDS	18	Hot-deck (n=628)
D USE OTHER LANGUAGE-RELATIVES	17	Hot-deck (n=628)
Q23 READ TO IN OTHER LANG AT HOM	14	Hot-deck (n=628)
Q24 READ TO OTH LANG ELSEWHERE	14	Hot-deck (n=628)
Q25 ENGLISH, USED IN SCHOOL WORK	15	Hot-deck (n=628)
Q26 SPEAK ENGLISH AT HOME	15	Hot-deck (n=628)
Q27 UNDERSTAND ENGLISH	16	Hot-deck (n=628)
Q28 SPEAK ENGLISH	15	Hot-deck (n=628)
Q29 READ ENGLISH	14	Hot-deck (n=628)
Q30 WRITE ENGLISH	14	Hot-deck (n=628)
Q43C BEFORE READING, REMEMBER	6	Hot-deck/deduction
D BEFORE READING, GUESS	7	Hot-deck/deduction
E BEFORE READING, SOMEBODY ELSE	8	Hot-deck/deduction
Q44B WHILE READING, MAKE NOTES	7	Hot-deck/deduction
C WHILE READING, THINK	7	Hot-deck/deduction
D WHILE READING, READ OVER	8	Hot-deck/deduction
E WHILE READING, GUESS	7	Hot-deck/deduction
Q45B AFTER READING, WRITE NOTES	6	Hot-deck/deduction
D AFTER READ, RELATED SELECTION	6	Hot-deck/deduction
E AFTER READING, NEW IDEAS	6	Hot-deck/deduction
F AFTER READING, SOMEBODY ELSE	6	Hot-deck/deduction
G AFTER READING, WRITE	6	Hot-deck/deduction
Q53A HOMEWORK, READING AND WRITING	7	Hot-deck/deduction
B HOMEWORK, READING ONLY	9	Hot-deck/deduction
C HMWK-CHOOSE READING & REPORT	9	Hot-deck/deduction
D HOMEWORK, READING CHOICE	10	Hot-deck/deduction
E HOMEWORK, OTHER	30	Deductive imputation
Q54A HOMEWORK HELP, MOTHER	8	Hot-deck
B HOMEWORK HELP, FATHER	13	Hot-deck
C HOMEWORK HELP, SIBLING	15	Hot-deck
D HOMEWORK HELP, TUTOR	18	Hot-deck
E HOMEWORK HELP, OTHER	17	Hot-deck
Q56 FREQ READING BOOKS FOR FUN	6	Hot-deck
Q58 FREQUENCY READING COMICS	6	Hot-deck
Q59 READING A MAGAZINE LAST WEEK	6	Hot-deck
Q60 FREQUENCY READING A MAGAZINE	7	Hot-deck
Q61 READING A NEWSPAPER LAST WEEK	6	Hot-deck
Q62 FREQUENCY READING NEWSPAPERS	7	Hot-deck
Q63 FREQUENCY READING DIRECTIONS	7	Hot-deck
Q64 FREQ READ TEXTBOOKS IN SCHOOL	7	Hot-deck
Q65 FREQ READING STORY BKS IN SCH	7	Hot-deck
Q66 FREQ READING WKBS IN SCHOOL	7	Hot-deck
Q67 FREQUENCY PRACTICE EXERCISES	7	Hot-deck
Q68 FREQ LOOKING UP INFORMATION	7	Hot-deck

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 12-13. Items with over 5 percent missing data in the grade 9 Student Questionnaires

Item		Percent missing	Method of imputation
Q20B	IN HOUSEHOLD, MALE GUARDIAN	15	Hot-deck/deduction
D	IN HOUSE, FEMALE GUARDIAN	16	Hot-deck/deduction
E	IN HOUSEHOLD, BROTHER(S)	10	Hot-deck/deduction
F	IN HOUSEHOLD, SISTER(S)	11	Hot-deck/deduction
G	IN HOUSEHOLD, GRANDPARENT(S)	18	Hot-deck/deduction
H	IN HOUSE, OTHER RELATIVES	17	Hot-deck/deduction
I	IN HOUSEHOLD, NON-RELATIVES	18	Hot-deck/deduction
Q46A	HOMEWORK, READING AND WRITING	6	Hot-deck
B	HOMEWORK, READING ONLY	11	Hot-deck
C	HMWK-CHOOSE READING & REPORT	12	Hot-deck
D	HOMEWORK, READING CHOICE	12	Hot-deck
E	HOMEWORK, OTHER	35	Deductive imputation
Q47A	HOMEWORK HELP, MOTHER	6	Hot-deck
B	HOMEWORK HELP, FATHER	9	Hot-deck
C	HOMEWORK HELP, SIBLING	12	Hot-deck
D	HOMEWORK HELP, TUTOR	16	Hot-deck
E	HOMEWORK HELP, OTHER	13	Hot-deck
Q61A	OWN WRITING, POETRY	11	Deductive imputation
B	OWN WRITING, DIARY	12	Deductive imputation
C	OWN WRITING, LETTERS	7	Deductive imputation
D	OWN WRITING, MESSAGES	11	Deductive imputation
E	OWN WRITING, STORIES	12	Deductive imputation
F	OWN WRITING, COMPUTER PROG	12	Deductive imputation
G	OWN WRITING, OTHER	30	Deductive imputation
Q63C	SCH OR HWK/FREQ READ FOR LAN	6	Hot-deck/deduction
E	SCH OR HWK/FREQ READ VO-TECH	7	Hot-deck/deduction
Q68AA	SCHOOL PLACE/GET UPSET	6	Hot-deck
BB	SCHOOL PLACE/FEEL GREAT	6	Hot-deck
C	SCHOOL PLACE/LIKE TO GO	6	Hot-deck
E	SCHOOL PLACE/FEEL IMPORTANT	6	Hot-deck
O	SCHOOL PLACE/LOOK UP TO ME	6	Hot-deck
X	SCH PLACE/GET SATISFACTION	6	Hot-deck

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

For question 53 of the fourth grade Student Questionnaire, part e ("other") elicited a 30 percent nonresponse rate, while parts a through d had from 7 to 10 percent missing. As discussed in Section 12.2, we concluded that the missing responses to part e constituted a "no" response--the students had no other kind of reading work for homework, other than that listed in parts a through d, and so did not respond. This is indicated in Table 12-12 as deductive imputation. For the remaining four parts, with much lower response rates, hot-deck imputation was used to impute all four parts simultaneously. Thus, for these parts we attempted to preserve in the data set the distributional characteristics of the respondents.

12.3.5. Identifying the Imputed Values

The imputed values were identified by imputation flags that have the following values: 0 = not imputed, 1 = imputed by hot-deck, and 2 = imputed by other methods. There is a flag variable that corresponds to each imputed item so the users can decide whether the imputed responses should be included in their analysis.

12.4. Effects of Imputation

A simple indication of the effects of imputation for an item can be obtained by comparing the distribution of the original responses to the distribution of the responses after imputations have been made for missing values. These distributions have been compared, and the results for each data item are included in a separate volume of technical appendices. As an illustration, we show the percentage distributions of several items from the fourth grade Student, Teacher, and School Questionnaires (Table 12-14). These items were chosen because they were often used in analyses and they had a relatively high nonresponse rate. In general, we found little difference in the distributions before and after imputation. Where there are differences, in most cases these are the realization of differences between respondents and nonrespondents to the item, as reflected in their values for the sort variables used to control the hot-deck imputation procedure. In the following discussion, we briefly examine the effects of imputation on the bias of univariate statistics, on the relationships between pairs of variables, and on sampling errors.

Table 12-14. Percentage distribution of selected items before and after imputation

Variable	Label	Categories	Percentage before imputation	Percentage after imputation
WASFEDUC	Father's education (Percent missing = 12%)	Less than high school	10	10
		High school	24	24
		Some college	18	18
		College or university	48	48
WASTV	Hours watching TV (Percent missing = 4%)	Low (0-1 hours)	17	16
		Moderate (2-4 hours)	50	50
		High (5 or more hours)	33	34
WATTZACO	Number of teacher education courses (Percent missing = 23%)	Low (10 or less)	30	32
		Moderate (11-20)	42	41
		High (20 or more)	28	27
WATPCTNA	Percentage classroom time teaching narrative text (Percent missing = 6%)	Low (less than 40%)	19	19
		Moderate (40-60%)	40	40
		High (60% or more)	41	41
WASLIBC	Number of library books at school (Percent missing = 7%)	Low (less than 5,000)	28	27
		Moderate (5,000-10,000)	51	52
		High (10,00 or more)	21	21

NOTE: Percentages were computed excluding missing records.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1994.

12.4.1. Effects of Imputation on Bias of Univariate Statistics

In general, we expect the imputation procedures applied in this study to improve the estimates of univariate statistics such as means, totals, and ratios. The imputations should reduce the nonresponse bias in the background variables used to define subclasses for which such univariate statistics are presented, and thus reduce the bias in these subclass estimates themselves. Note that no imputation has been used for the cognitive items used to derive the proficiency scale scores, nor for the scores themselves. Although the nonresponse bias in this study should be relatively small because of the small amount of missing data, we recommend that users employ the imputed data to produce population estimates. An acceptable alternative for simple summary statistics is to use the raw responses and include an explicit category for missing cases in the resulting tabulations.

To examine the effect of imputation, we used the imputed and unimputed data to compute the means and the standard errors of the means for the narrative scaled scores for fourth grade students. These statistics were estimated for subgroups classified by variables from the Student, Teacher, and School Questionnaires. The results showed that the estimates were very similar (Table 12-15).

12.4.2. Effects of Imputation on Relationships

The hot-deck imputation used in this study preserves the relationship between the classifying variables used in the imputation model and the item to be imputed. Discussions on the effects of imputation on relationships are provided by Santos (1981), Kalton and Kaspryzk (1982), and Little (1986). It is possible that the imputations may have attenuated some multivariate relationships. However, the effect is likely to be negligible in view of the small amount of missing data. The analyst conducting multivariate analyses who is concerned about the effects of imputation may want to compare results using the imputed and the original respondent datasets, respectively. This can be accomplished by using the imputation flag for each variable of interest.

As an example, we estimated a regression equation predicting the narrative scaled scores for fourth grade students using the data with and without imputation (Table 12-16). Using a list-wise deletion of students with missing information, we found that the sample size without imputation was effectively cut in half (there were 3,184 students in the unimputed data set and 6,248 in the imputed set). Results from the regression equations are fairly comparable. Most of the parameters found to be significant in the imputed data set are also significant with the unimputed data set with the exceptions of the parameters "mother and step-father families" and "language at home-other first and other." These parameters are significant with the imputed data set, but not significant with the unimputed data set. A possible explanation is that students who speak a language other than English at home are likely to have more missing responses because of a language barrier. Therefore, they were probably excluded in the analysis with the unimputed data set, and their difference from the other students were not identified. This example illustrated that the imputations conducted in this study is effective in maintaining the sample size for multivariate analyses.

Table 12-15. The mean and standard error of the narrative scaled score for grade 4 students by various characteristics: A comparison of the results using data with and without imputation

Characteristics	Estimates using data	
	With imputation	Without imputation
Father's education		
Less than high school	233 (±3.1)	233 (±3.4)
High school	247 (±1.8)	248 (±1.8)
Some college	252 (±2.1)	254 (±2.3)
College or university	258 (±1.8)	259 (±1.8)
Time watching TV		
Low (0-1 hour)	256 (±2.3)	257 (±2.2)
Moderate (2-4 hours)	256 (±1.5)	257 (±1.5)
High (5 or more hours)	242 (±2.2)	243 (±2.2)
Number of teacher education courses taken by teacher		
Low (10 or less courses)	251 (±2.0)	251 (±2.3)
Moderate (11-20 courses)	254 (±2.0)	253 (±2.8)
High (more than 20 courses)	251 (±2.7)	252 (±2.8)
Percent of classroom time teachers spent teaching narrative text		
Low (less than 40 percent)	247 (±2.8)	247 (±3.0)
Moderate (40-60 percent)	254 (±2.2)	255 (±2.1)
High (60 percent or more)	251 (±2.2)	250 (±2.3)
Number of library books school contained		
Low (less than 5,000)	248 (±3.5)	247 (±3.7)
Moderate (5,000 - 10,000)	253 (±2.5)	252 (±2.4)
High (10,000 or more)	255 (±4.2)	254 (±3.9)

NOTE: 1. Standard error of the mean is shown in parenthesis. Standard errors were computed using a jackknife variance estimation procedure. The program WESVAR was used for these calculations.

2. The narrative scaled score used in this analysis was scaled for the U.S. population. For international comparisons, this score has been rescaled with a mean of 500 and standard deviation of 100. The international scale was used in the rest of this report.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 12-16. Estimated regression coefficients predicting grade 4 student's narrative scale score, using data with and without imputation

Predictor variable	Data with imputation		Data without imputation	
	Regression coefficient	Standard error	Regression coefficient	Standard error
Intercept	379.2*	2.10	373.2*	2.82
Student's age	-0.9*	0.01	-0.8*	0.02
Student's sex (1 = female, 0 = male)	9.0*	0.27	9.8*	0.41
Asian	4.9*	0.47	4.6*	0.79
American Indian	-5.4*	0.87	-4.7*	1.12
Hispanic	-10.5*	0.43	-18.7*	0.60
Black	-25.9*	0.33	-28.2*	0.64
Father's education less than high school	-10.7*	0.43	-15.6*	0.66
Father's education high school only	-8.0*	0.24	-9.7*	0.39
Father's education some college	-4.8*	0.32	-8.6*	0.37
Mother's education less than high school	-8.1*	0.52	-4.8*	0.73
Mother's education high school only	-0.6	0.37	0.9*	0.48
Mother's education some college	-0.4	0.30	-0.3	0.42
Family wealth index	102.5*	3.22	97.5*	3.92
No parents in household	-6.5*	0.70	-11.3*	1.05
Stepparent(s)	-14.3*	0.57	-21.2*	1.13
Mother only	1.2*	0.40	-1.0	0.60
Mother and stepfather	-5.7*	0.45	-9.2*	0.67
Father only	-14.5*	0.63	-18.4*	0.99
Father and stepmother	-13.3*	0.56	-22.0*	0.73
Unknown parent arrangement	-14.9*	0.42	-13.8*	0.62
Extended family	-11.3*	0.25	-14.8*	0.37
Language at home English first and other	-14.3*	0.61	-12.2*	0.68
Language at home other first and English	-4.0*	0.23	-2.8*	0.40
Language at home other first and other	-6.8*	0.52	-0.3	0.79
Model R^2	0.1		0.1	.
Sample size	6,248		3,184	

NOTE: The "*" next to the regression coefficient indicates significance at a 0.05 level using a 2-tailed t-test. The standard error of the regression coefficients were computed using the program WESREG with 33 sets of replicate weights developed for jackknife variance estimation.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

12.4.3. Effects of Imputation on Sampling Error

By assigning values for missing responses, imputation acts to inflate artificially the apparent reliability of the data, to some extent. As a result, the routine application of standard error calculations will tend to overstate the precision of survey estimates (i.e., understate the level of sampling error). However, with the small extent of imputation used in this study, the degree of overstatement will not be sizable here.

There is no quick and easy way to measure the effect of imputation on sampling error. Rubin (1987) advocates the use of multiple imputations in which the dataset is completed not once, but several times, using the same imputation model. The major drawback to this routine is the additional amount of computing it entails. Users concerned about sampling error should also see recent work by Rao and Shao (1991) and Sarndal (1991), who have made progress toward dealing with the problem of standard error estimation with imputed data sets without resorting to multiple imputations. Users can gain an indication of the upper bound of the inflation in the reliability of estimates resulting from the use of imputed data by comparing the standard error obtained from using the unimputed data (by dropping cases with missing data from the analysis) with the standard error estimate using the imputed data, for a series of key estimates.

References

Kalton, G. (1983). *Compensating for missing survey data.* Research Report Series. Ann Arbor: Institution for Social Research. University of Michigan.

Kalton, G., and Kasprzyk, D. (1982). Imputing for missing survey responses. *Proceedings of the section on survey research methods*, 22-31. American Statistical Association.

Little, R.J.A. (1986) Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 139-157.

Little, R.J.A., and Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Rao, J.N.K., and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika*, 79 (4), 811-22.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley.

Santos, R.L. (1981). Effects of imputation on regression coefficients. *Proceedings of the section on survey research methods*, 140-14. American Statistical Association.

Sarndal, C.E. (1992). Methods for estimating the precisions of survey estimates when imputation has been used. *Survey Methodology*, 18, 241-152.

13. CONSTRUCTS AND DATA

13.1. Introduction

This chapter focuses on the reading research literature. While it is concerned with the data developed in the U.S. portion of the International Reading Literacy Study and how we organized that data, its more primary emphasis is on the research about reading. In principle, this chapter should provide the context for understanding what the Reading Literacy Study data can tell us about learning to read. It describes how we organized the data -- both logically and empirically -- our attempts to relate the data to what the literature says about learning to read, and finally how we selected certain sets of data, organized in certain ways, for inclusion in a model of reading that we will develop in Chapter 14. The variables included in the study have, in fact, limited and structured the review of the literature.

We began with the extensive data set developed by the International Steering Committee, the 32 National Research Coordinators, and the U.S. National Steering Committee. Together these groups of people, with differing perspectives on what influences how successful children are at learning to read, developed questionnaires that included over 500 discrete items that they believed were related to learning to read. From this large base, we organized the variables into categories that would systematically relate to the literature (a further description is provided in Section 13.2.1) and would serve as the basis for our model of reading developed in Chapter 14.

Given the extensive nature of the ground covered in this or any other omnibus survey, we have selectively gone into the literature with an eye on explaining or elaborating on just those variables we used to model reading proficiency. We explain how we inferred the latent conceptual structure of the data and how we developed and estimated measurement models consistent with this structure and the data. But our overall intention here is to embed the constructs identified in the theoretical and substantive literature on reading. The purpose of this exercise is to clarify the range of possible interpretations of the data and findings. It is also to provide the reader with sufficient general information to place this data in a larger context, recognizing when the variables and data appropriately address important issues.

The organizing structure for the chapter is reflected in these activities. The first sections describe the methods used to infer the latent structure of the data from Student, Teacher, and School Questionnaires. In contrast, the remaining sections of the chapter describe the creation and estimation of latent variables, and the assignment of substantive meaning to them, describing their relationship to reading proficiency, the relationship of the findings to those from other surveys, and the linkages of these variables and relationships to the research and policy literature concerning children's reading comprehension.

13.2. Organizing the Data

13.2.1. The First Framework

The instrumentation was extensive, with over 500 separate responses from students, teachers, and schools in three questionnaires administered at each grade level. In many ways this data set as a whole has many of the characteristics of an omnibus survey such as High School and Beyond (HSB), National Educational Longitudinal Study of 1988 (NELS:88), or National Assessment of Educational Progress (NAEP). To develop a conceptual understanding of the data, and to begin the more detailed work on the definition of variables and constructs, we developed a simple classification system as the beginning frame.

It seemed possible to think of the data in terms of two dimensions: to whom and to what they referred. In the case of the "who" dimension, the data seemed to describe **students**, their **families**, their **teachers**, the **classes** they were in, their **principals or instructional leaders**, their **schools**, and their **communities**. On the "what" dimension, each of these can be considered in terms of their **attributes** and the kinds of **environments** they provided. Cast in the form of a matrix, the next task was to determine how to distribute the 500 or so discrete items from the questionnaires within each of the 10 cells. These two dimensions produce 14 cells, although only 10 cells were needed to capture the measures used (Figure 13-1).

13.2.2. Constructs and Rules of Thumb

Within each of the categories defined above, conceptual and statistical explorations of the data aimed at producing meaningful constructs were organized around four rules of thumb that we used to guide our judgments.

The first of these concerned blocks of items having the same response scales, grouped together as a single "question" in the questionnaire, and with an apparently common theme tied to the literature on reading. These items were assumed to be tapping an identifiable substantive domain. The matter of the latent structure of the item group was resolved through exploratory factor analyses. However, the rule did not hold in all cases. In some instances the items grouped together were only loosely related. For example, in question T4Q53[1] the 30 subsumed items, which are all generally tied to teaching practices, tap more than one aspect of instructional practice. A number of items are statements solely related to student- or teacher-directedness (e.g., "Students have a choice in what they will do"), while other items relate more closely to the content of reading instruction (e.g., "Specific skills are taught at certain times").

The level of intuition required to assign meaning to items varied across the items. In some instances the linkages of item groups to the literature could be readily recognized. The three questions to do with reading strategies (S4Q43, S4Q44, and S4Q45) in the Student Questionnaire provide an illustration. Each has less than 10 items and each item has a common stem, a common response scale, and corresponds closely to steps in a directed reading lesson -- a prevailing instructional strategy consistently included in teacher editions of basal readers and numerous methods books.

At the other extreme, question T4Q43 in the Teacher Questionnaire is less clearly linked to a single substantive body of literature. This question, consisting of 26 separate items, has its origins in various positions on reading instruction. Represented in this array are notions of hierarchical skills approaches (e.g., reading learning materials should be carefully sequenced in terms of language structures and vocabulary), basal approaches (e.g., class sets of graded reading material should be used as the basis for the reading program), whole language approaches (e.g., students should always choose their own books to read), and the relationship between reading and writing (e.g., students should be encouraged to read texts they have written), to name just a few. The intended latent structure of these items is not obvious, nor are all positions on reading instruction equally represented or measured.

---

[1] Note that the items in question are identified by questionnaire, sample, and item number as follows:

S4 = fourth grade Student Questionnaire
S9 = ninth grade Student Questionnaire
T4 = fourth grade Teacher Questionnaire
T9 = ninth grade Teacher Questionnaire
P4 = fourth grade School (principal) Questionnaire
P9 = ninth grade School (principal) Questionnaire

Q1...Qn = questionnaire item numbers

Figure 13-1. Conceptual structure of Reading Literacy Study data

	ATTRIBUTE	ENVIRONMENT	ACHIEVEMENT
STUDENT	age gender race/ethnicity language		narrative comprehension expository comprehension document comprehension
FAMILY	family composition augmented family father's education mother's education family wealth	regularity of meals study facilities literacy resources TV watching homework help parental interest	
TEACHER	gender race/ethnicity teacher training formal education teaching experience	type emphasis reading homework grouping practices assessment emphasis teacher's reading preference transmission/translation emphasis	
CLASS	size reading resources instructional time (all) instructional time (reading)	% remedial students % remedial assisted	
PRINCIPAL	gender race/ethnicity training (administration) training (reading) experience (total) experience (present school) instructional leadership staff development		
SCHOOL	sector instructional time library resources % specialist teachers % minority teachers size		
COMMUNITY	urbanicity resources region cooperation		

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

A second and similar rule of thumb was used in dealing with items that seemed to have a common focus, used a common response scale, but were treated as separate questions in the questionnaire; the items concerning reading activities in the fourth grade Student Questionnaire are one example (S4Q55-S4Q68). We tended to treat item configurations of this type as though they were the same as those noted immediately above--multiple indicators of one or more constructs to be determined through exploratory analyses.

A third configuration of items treated as a group for the purposes of analysis was drawn from the residue of single items, some in sequence within the questionnaire, and some not, thus incorporating those that were seen as having a common theme. For example, in the Student Questionnaire we combined a series of contiguous questions to do with reading aloud at home (S4Q38, S4Q39, S4Q40, S4Q41), an or item on whether parents asked the student about his or her reading at school (S4Q42), and an item on whether parents helped with homework (S4Q54a, S4Q54b). The rationale in this instance is that these items are likely to tap parental interest in and encouragement for the student's reading. In this case, the approach is informed to a degree by notions of what one would have liked or expected to see measured in a model that sought to explain variation in children's reading comprehension.

A fourth rule of thumb was developed in connection with item groups in which we inferred that a measure capturing a pattern of responses across several items was the intent, rather than the measurement, of a latent variable through multiple indicators. Examples include items requesting the respondent to rank order a series of alternatives--or a subset of the items (S4Q36, for example). Similarly, questions about the school's use of standardized test results in evaluation (P4Q19) or about the teacher's involvement in subject-specific training (T4Q10-T4Q12) were treated as indicators of a pattern rather than a single construct.

13.2.3. Constructs and Statistics

The statistical treatment of the data for constructs defined in these ways took three forms:

1. In the case of single item indicators we simply looked at the distribution of responses to see if it looked reasonable and, if the data seemed to demand it, collapsed the tails of some of the distributions to combine contiguous categories containing relatively small numbers of respondents and so reduce the possibility that extreme outliers would distort the results. The parent education questions (S4Q5, S4Q6), the question about the number of persons in the household (S4Q13), and the item on self-rating of reading ability (S4Q35) are examples.

2. In the case of groups of items thought to be tapping one or more constructs from a single domain, we engaged in exploratory factor analyses to get at the latent structure of these items. As a general strategy, a principal factor solution was obtained and, in the first instance, factors with eigenvalues greater than one rotated to an oblique solution. In subsequent analyses more or fewer factors were rotated until a solution was obtained that exhibited good simple structure and whose factors could be assigned meaning consistent with the theory and substance of reading and reading instruction. Factor scores were estimated to provide measures of the latent variables identified.

3. In some instances we created single categorical variables out of several items by inspecting the patterns of response across the several items (this amounts to displaying the cells of a multiway tabulation). For example, in the case of the three-item question

about meals eaten during the day (S4Q7), the distribution of the responses suggested that we could combine these items into a single variable capturing whether the respondent dined once, twice, or three times each day on a regular basis.

The specifics of these statistical treatments of the questionnaire items are dealt with below in some detail on a variable-by-variable basis. Each variable is associated with the appropriate literature.

13.3. The Constructs

In the following sections we describe each of the categories of variables listed in Figure 13-1. Each section covers one of the categories, including a listing of the variables related to that category and a discussion of each resulting construct. This is done by providing a description of the variables in terms of what they measure, how they relate to reading proficiency, how that relationship compares to findings from other surveys, and how the data relate to the larger research and policy literature. The data are, therefore, placed in the context of issues confronting policymakers.

13.3.1. Student Attributes

Items categorized as student attributes focused on preexisting traits of students. Four attributes were identified:

- Age (S4Q1, S4Q2; S9Q1, S9Q2),
- Gender (S4Q3; S9Q3),
- Race/ethnicity (S4Q4; S9Q4), and
- Language (S4Q15, S4Q16; S9Q21, S9Q22).

Age (S4Q1, S4Q2; S9Q1, S9Q2). Students were asked to report their age in years on their last birthday along with the month, day, and year of their birth. Based on experience with NAEP, the two questions would provide the most accurate measure available. Age in months at the time of the survey (March 1991) was calculated from these data.

In the grade 4 sample, students in the U.S. ranged in age from 87 months (7 years, 3 months) to 153 months (12 years, 9 months) with a mean of 120 months (10 years). In the grade 9 sample, students in the U.S. ranged in age from 157 months (13 years, 1 month) to 250 months (20 years, 10 months) with a mean of 180 months (15 years; Table 13-1).

The Reading Literacy Study data related to age reflect the application of more than just an age criteria to grade placement. In both grades, the age range represented is wide, spanning more than 4 years. In looking at the relationship between reading proficiency and age in grade 4, we note that as the age of the student increases toward 10 years there is an improvement in the mean reading proficiency scores (Figure 13-2). After age 10 1/2 there is a sharp decline in reading proficiency. This is most likely to be related to retention policies, where students who have not met certain standards repeat the grade. In grade 9 (Figure 13-2), the decline in scores after age 15.5 is also apparent. Here it is even more likely to be associated with retention and class placement. In contrast to grade 4, where there was an increase in proficiency associated with the children in the low end of the age range, this trend was not the case in grade 9. This might be related to the difference in emphasis in instruction related to reading across the grades.[2]

The most common method for grouping students into classes across the nation is by age. Although there is some variation in compulsory school starting age across the 50 states and the District of Columbia, the overwhelming majority of states have compulsory school attendance beginning at age 6 (17 states) and 7 (27 states).[3] However, placement, promotion, acceleration, and retention policies, most of which are formulated at the district level, typically reflect developmental differences in the early grades and achievement standards in the later grades as well.[4] In fact, single grade enrollment figures reflect the intermingling of these three decision criteria. While there is clearly a modal age for a given grade, large numbers of students who are both younger and older are also enrolled in any given grade. For example, according to the 1990 October school enrollment figures there were 18,000 7-year-olds, 191,000 8-year-olds, 2,462,000 9-year-olds, 843,000 10-year-olds, 114,000 11-year-olds, 16,000 12-year-olds, 6,000 13-year-olds, and 1,000 14-year-olds in grade 4 (Kominski and Adams 1992).

Table 13-1. Mean reading proficiency scores, by age of student: Grades 4 and 9

Grade 4					Grade 9				
Age	Percent	Narrative	Expository	Document	Age	Percent	Narrative	Expository	Document
< = 9	0.9	515 (9.0)	502 (12.4)	498 (9.6)	< = 14.0 ...	.2	563 (2.1)	572 (26.2)	548 (26.2)
< = 9.5	9.7	552 (4.6)	535 (4.7)	546 (4.9)	< = 14.5 ...	12.3	556 (10.3)	562 (13.1)	540 (8.2)
< = 10.0 ...	36.3	571 (3.5)	553 (3.0)	560 (2.6)	< = 15.0 ...	37.9	558 (4.5)	558 (5.2)	544 (3.7).
< = 10.5 ...	32.7	567 (3.2)	549 (3.2)	562 (2.8)	< = 15.5 ...	31.5	550 (5.7)	554 (6.2)	536 (4.2)
< = 11.0 ...	13.7	523 (4.8)	512 (4.3)	530 (4.8)	< = 16.0 ...	9.9	500 (7.1)	498 (7.1)	497 (7.0)
< = 11.5 ...	5.3	491 (7.7)	489 (5.1)	505 (5.3)	< = 16.5 ...	5.0	476 (8.1)	470 (9.3)	472 (7.4)
< = 12.0 ...	1.0	481 (9.6)	483 (9.5)	489 (9.4)	< = 17.0 ...	1.6	456 (8.0)	476 (10.0)	484 (10.0)
>12.0	0.4	458 (11.1)	464 (10.3)	499 (9.2)	>17.0	1.6	435 (9.1)	431 (14.0)	439 (10.6)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

---

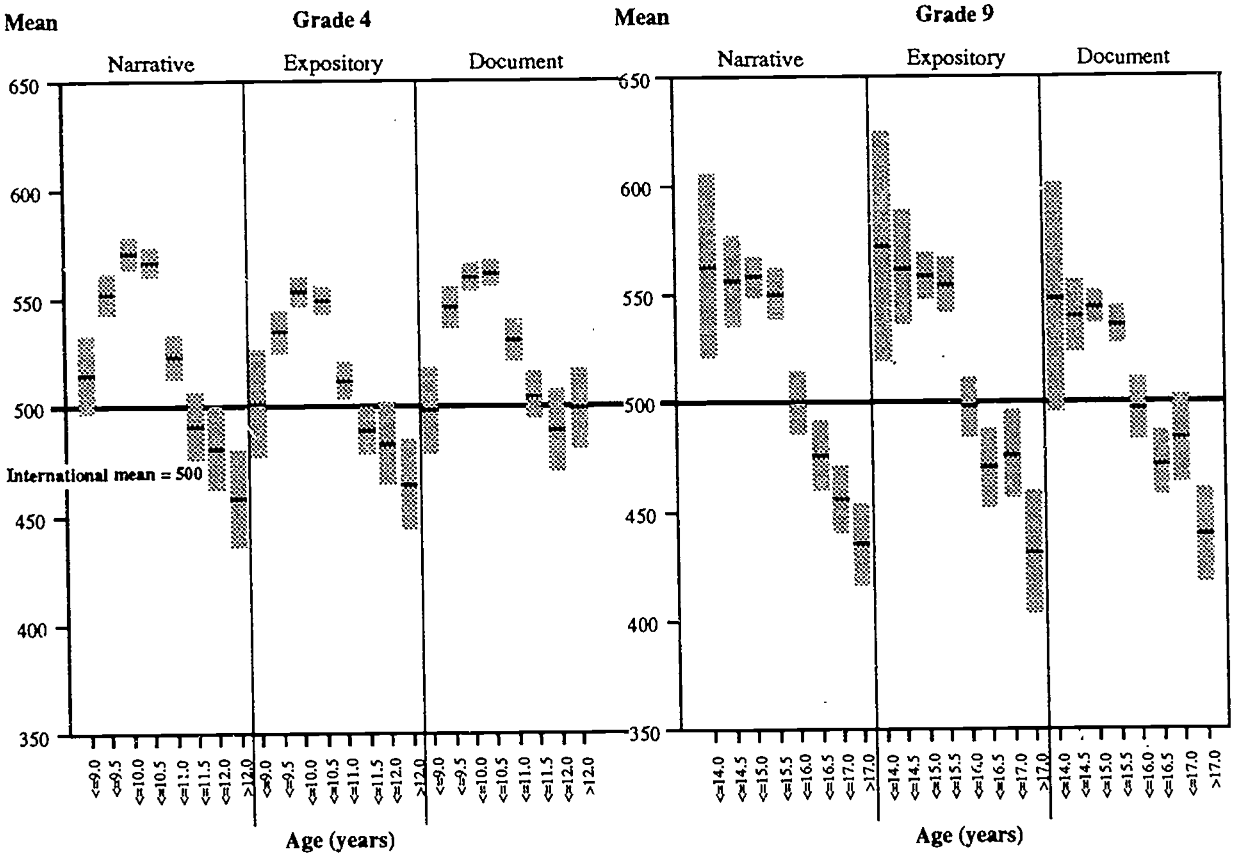
[2] In the earlier grades students are learning to read, while in the later grades they are expected to read to learn. Although the transition between the two emphases is between grades 3 and 4, one might still see the effects of increased instruction resulting in a growth pattern related to development and instruction.

[3] Four states (Arkansas, Delaware, South Carolina, and Virginia) mandate age 5 (although South Carolina and Virginia permit parental waiver of kindergarten at age 5) and the remaining 3 states (Arizona, Pennsylvania, and Washington) do not begin compulsory schooling until age 8 (*Digest of Education Statistics* 1992, p.

[4] Although the policies may vary. the following example from the Board of Education of Montgomery County, Maryland, is typical. Their regulations state that "Although students of the same age share similarities as a group, they differ remarkably from one another... Most children pass through the same sequence of developmental stages, but the pace of development varies from child to child... Acknowledgement of developmental differences should be the basis for placement, promotion, acceleration and retention decisions." Despite the implied flexibility, the regulations stipulate that:

> "a. In prekindergarten through grade two, placement and promotion should be based on age...
> b. In grades three through eight, placement and promotion should be based on academic progress and attainment of objectives assigned to the student...
> c. In grades nine through twelve, placement and promotion of students should be based on the number of credits earned..."

Figure 13-2. Mean reading proficiency scores, with 95 percent confidence intervals, by age of student: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In the literature, age is most often associated with developmental stages. As many psychologists (for example, Piaget 1959) have argued, children progress through certain cognitive stages that influence the way and what they learn. While there are likely to be individual differences that affect exactly when particular children pass through specific developmental stages, these tend to occur at roughly the same age. Similarly, we might expect certain types of reading behaviors to be apparent at certain ages; for example, the notion of readiness to learn phonics being associated with children between the ages of 6 1/2 and 7 represents this line of reasoning (Morphett and Washburne 1931; Dolch and Bloomster 1937).

While the early research appeared to tie readiness for reading instruction (and particularly phonics instruction) to the age of 6 or 7, more recent research has challenged this association between age or even more specifically mental age and beginning instruction (Adams 1990). For example, as early as 1967 researchers had established that factors such as prereaders' letter knowledge and ability to discriminate phonemes auditorially were much better predictors of success in early reading achievement than age or mental age (Chall 1967; Bond and Dykstra 1967). Further, a number of studies demonstrated that phonics (which would include these indicators) could successfully be taught to young or slow learners (Bateman 1979; Wallach and Wallach 1979; Williams 1979).

Yopp and Singer (1985) went even further in challenging this relationship. As a result of their review of the research, they concluded that what actually made the difference in reading achievement was not age or mental age, but rather the ways in which children of any particular age had been taught. The more recent research on early readers would support this conclusion (Durkin 1966; Clark 1976; Tobin and Pikulski 1988; Heath 1982, 1983; Heath and Branscombe 1985; Ninio 1980; Teale 1984; Sulzby and Teale 1987).

Considered in this light, age would not necessarily be associated with a developmental stage, but rather with the increased probability of exposure to instruction that may have taken a variety of forms -- i.e., informal instruction at home or systematic instruction in a preschool setting. This raises the question of what effect earlier and more instruction might have on reading achievement.

Although the state policies stipulate an age for compulsory school attendance, they do not necessarily limit or restrict the possibility that younger children will participate in either formal schooling or in informal instructional activities. A large and growing number of children between the ages of 3 and 5 are being enrolled in preprimary programs. As early as 1965, 3.4 million children (27 percent of the age cohort) were enrolled. By 1979, over 50 percent of the age cohort were enrolled in preprimary programs. In 1991 this had increased to 6.3 million (55.7 percent of the age cohort; U.S. Department of Education 1992, Table 47, 61).

As part of the 1988 NAEP reading assessment, fourth graders were asked whether they had attended preschool, nursery school, or day care, and whether they had attended kindergarten. Comparisons between students who had an earlier start in school and those who had not indicate that those who did were more likely to have higher reading proficiencies (Langer et al. 1990, 19).

While attendance in preprimary programs might be seen as a way to ameliorate differences in home factors affecting reading achievement, the data suggest that the opposite may be occurring because of actual attendance or participation patterns. For example, the more education parents have, the more likely it is that their child will participate in some kind of preprimary experience. This is also reflected in the choice of program -- the more educated parents tend more often to choose nursery school over day care (Snyder and Hoffman 1992, 62).

Unlike other surveys, where the relationship between age, placement policies, and achievement were specifically to be studied, we have no measures of preschool experience or grade retention. Consequently, we cannot further enlighten the debate on instruction versus developmental stage with these data.

Gender (S4Q3; S9Q3). The measure of gender comes from answers to the question "Are you a boy or a girl?" At both age levels, the distribution of boys and girls within the U.S. was approximately equal, as was to be expected (Table 13-2 and Figure 13-3).

Table 13-2. Mean reading proficiency scores, by gender: Grades 4 and 9

Gender	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Male . . .	50.4	546 (3.6)	535 (3.2)	552 (3.0)	49.3	530 (6.2)	541 (7.5)	530 (4.9)
Female . .	49.6	564 (3.1)	544 (3.0)	550 (2.8)	50.7	554 (5.0)	546 (5.7)	531 (4.0)

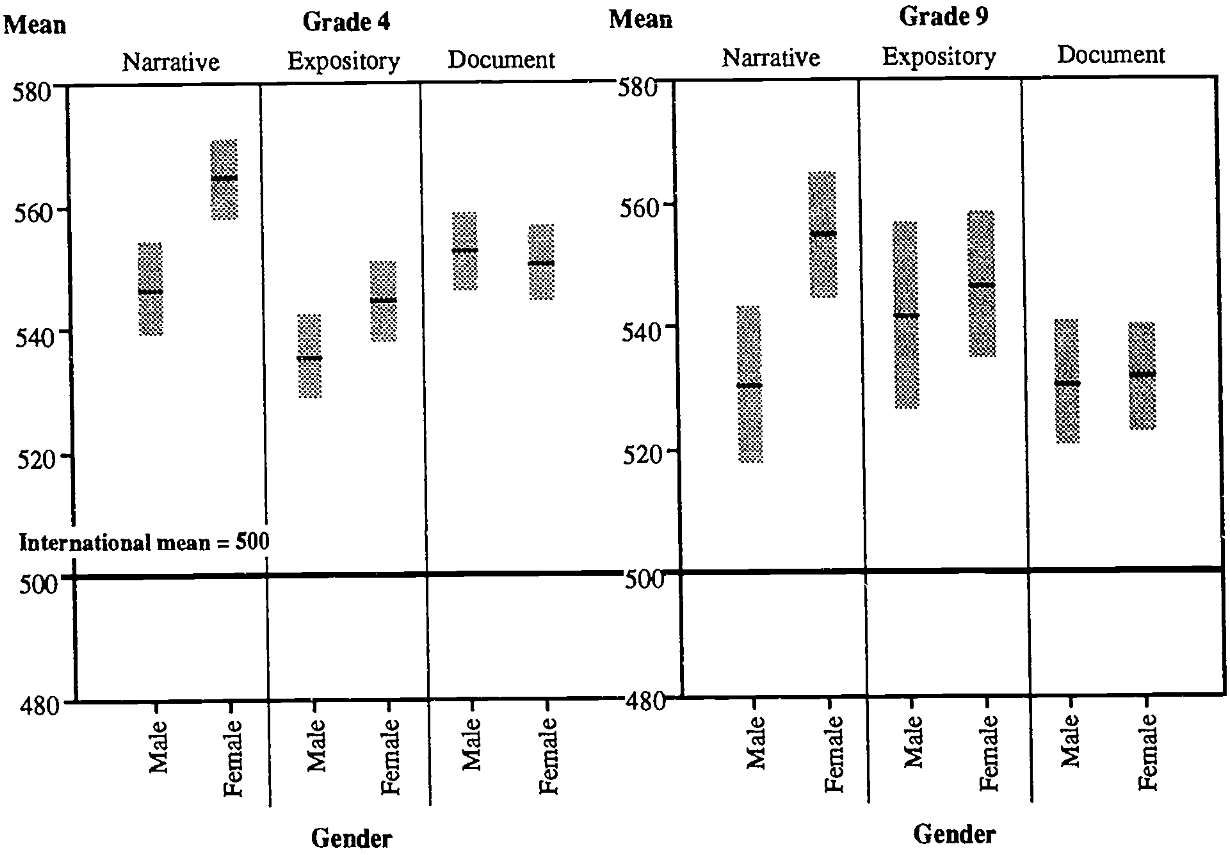
NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As discussed in Section 9.3.1, a comparison of the means of males and females for each domain and grade reveals that females outperform males on the narrative scale at each grade level. In the expository domain, while females appear to have a somewhat higher mean score than males at both grade levels, there is no statistical significance to these differences. With regard to the document scale, the means for each gender at both grades are essentially the same.

There appears to be no fundamental reason why one or the other gender should necessarily possess greater reading proficiency. In contrast to findings in math and science, where males are generally more proficient than females, the gender differences in literacy are more minimal and seem to favor females. The reading literacy data reported above are consistent with findings of gender differences on other surveys. For example, according to the data reported in the NAEP *Reading Report Card, 1971-88*, females at all three ages tested outperformed their male counterparts. However, while female performance across the assessments has remained fairly constant, males have shown a significant net gain over time. Still, in the 1992 NAEP assessment, the pattern of females outperforming males at all three grade levels was maintained (Mullis et al. 1993, Table 3.4, 105).

Figure 13-3. Mean reading proficiency scores, with 95 percent confidence intervals, by gender: Grades 4 and 9



NOTE: Shaded bands indicate the 95 percent confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Most research in gender differences with regard to reading proficiency indicates that they might be related to differences in interest (Guthrie and Greaney 1991). Girls tend to favor literature pertaining to people, their relationships, home, and fashion, while boys tend to be more interested in nonfiction and topics related to science, sports, and transportation (Southgate, Arnold, and Johnson 1981; Huus 1979; Meisel and Glass 1970; Ashley 1970; Heather 1981). Limited evidence suggests that boys are more interested in reading newspapers than girls (Flodin, Hedinsson, and Roe 1982; Norvell 1966), while girls appear to favor magazines (Gorman et al. 1983). Girls devote more time to reading (Robinson and Weintraub 1973), and tend to read more than boys (Jenkinson 1940; James 1987; Ashov and Fishbach 1973; Chiu 1984; Greaney and Hegarty 1987).

These differences in patterns of interests and attitudes might explain the difference in performance on NAEP reading assessments. When the 1992 NAEP assessment results are reported by type of reading, the figures vary slightly across the scales in ways that may be considered consistent with interest -- the gap between males and females is smaller in relation to texts that are read to gain information than in texts read for a literary experience.[5]

A related pattern is apparent in the National Adult Literacy Survey (NALS). The authors of *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey* (Kirsch et al. 1993) report that "on the prose scale, the average proficiencies of women and men are about the same. ...In contrast, men's average document and quantitative proficiencies are significantly higher than those of women (p. 46).[6]

Across time there has been a shift in how literacy tests have been constructed. While these assessments have previously used narrative texts, or texts specifically constructed for testing purposes, there has been a move to include a more balanced selection of various genre. When the assessment results are reported in relation to the genre, the differences favoring females begin to diminish. While females tend to do better with narrative materials, as is consistent with their interests, males do as well on expository text, as is consistent with their interests, and about the same or better on documents.

Race/Ethnicity (S4Q4; S9Q4). Students in both grades were asked to indicate their race and ethnicity according to the standard designation included on NAEP, NELS, and other NCES surveys. The categories were Asian or Pacific Islander, American Indian or Alaskan Native, Hispanic, white, and black. The distribution of students across the ethnic/racial groups parallels that reported in NAEP and NELS.[7]

Within the Reading Literacy Study, the mean scores for white students significantly exceeded those for black and Hispanic students for each domain at each grade level. The mean scores for Asian/Pacific Islander students significantly exceeded those of black and Hispanic students for all three domains at grade 4[8] (Table 13-3 and Figure 13-4).

---

[5] According to NAEP, the fourth grade public school students' average reading proficiency for reading for literary experience for males was 214 (1.6) and for females, 223 (1.1). In contrast, the spread was less on the reading to gain information scale, where the males had an average proficiency of 210 (1.5) and females, 216 (1.4).

[6] NALS does not have a separate narrative or literary experience scale as this genre does not figure prominently in adult literacy requirements. Instead of discriminating between narrative and expository text types, both are considered to be part of a prose scale. Examination of the test instruments and the definition of the domain indicates that there is very little if any narrative included in that measure.

[7] This is based on a comparison with the NAEP 1992 reading assessment data (p. 101) and the NELS:88 data (*Profile of the American Eighth Grader*, p. 1).

[8] For a complete discussion of these comparisons, see Section 9.3.2.

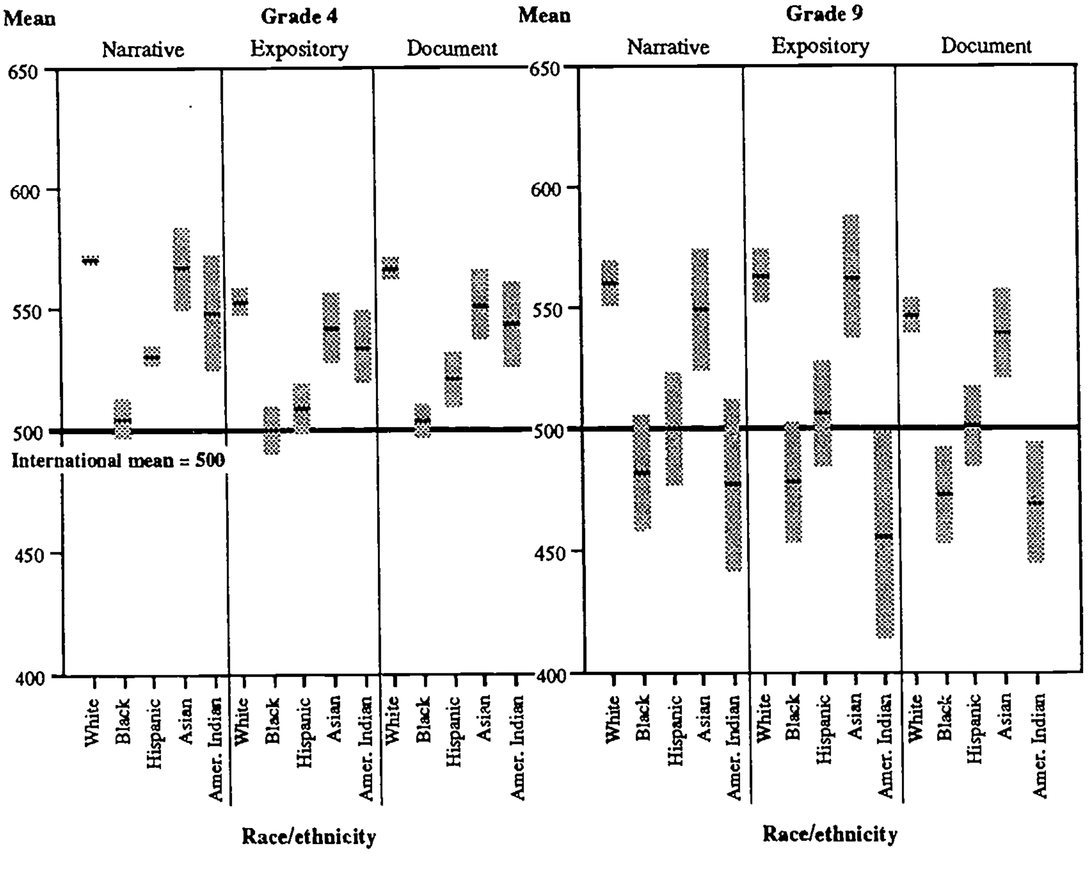
Table 13-3. Mean reading proficiency scores, by race/ethnicity: Grades 4 and 9

Race/ethnicity	Grade 4							Grade 9						
	Percent	Narrative		Expository		Document		Percent	Narrative		Expository		Document	
White	67.5	570	(2.3)	553	(2.6)	566	(2.1)	72.5	560	(4.5)	563	(5.4)	546	(3.7)
Black	16.8	505	(4.3)	500	(4.8)	504	(3.1)	12.7	482	(11.6)	478	(13.1)	473	(9.5)
Hispanic	8.7	528	(3.9)	509	(5.1)	521	(5.3)	8.0	500	(11.3)	506	(10.6)	501	(8.0)
Asian	3.9	567	(8.4)	542	(7.2)	551	(7.1)	3.1	549	(12.1)	562	(12.5)	539	(9.2)
American Indian	3.1	548	(11.7)	534	(7.4)	544	(8.6)	3.7	477	(17.6)	456	(21.2)	469	(12.3)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-4. Mean reading proficiency scores, with 95 percent confidence intervals, by race/ethnicity: Grades 4 and 9



NOTE: Shaded bands indicate the 95 percent confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

It is not surprising to find this pattern of performance across the racial/ethnic groups. Since its inception, NAEP has tracked the differences in reading proficiency among racial/ethnic groups and has found that the results have consistently tended to favor white students (Langer et al. 1990, 13).

Comparisons of NAEP results between 1971 and 1990 show that white students have maintained a constant level of performance. However, there have been significant increases in the average proficiency for both blacks and Hispanics. This has lessened the gap between whites and the minority populations (Mullis et al. 1991).

However, not all minority racial or ethnic groups do less well than whites. In the Reading Literacy Study data there is no significant difference in performance between whites and Asian/Pacific Islanders at either grade level on any of the scales. Similarly, in the 1992 NAEP reading assessment, white and Asian/Pacific Islander students had essentially the same average reading proficiency at every grade level tested. The NELS data present the same picture as well. The percentage of white and Asian students who demonstrate advanced levels of proficiency is double that of the Hispanic and black students (Hafner et al. 1990, 27).

Researchers have tried to determine why there are differences in performance patterns across the racial/ethnic groups. Research during the 1960s and 1970s documented the patterns of difference. For example, a number of researchers have argued that black children experience more difficulty with reading than white children, and the discrepancy increases across the school years (Wigfield and Asher 1984; Coleman et al. 1966; Singer, Gerard, and Redfearn 1975).

In explaining these findings, researchers have noted that these differences are often associated with other factors. For example, the authors of *Learning to Read in Our Nation's Schools* (Langer et al. 1990, 14) suggest that "in large part, these differences in performance (across ethnic groups) reflect differences in socioeconomic status." Others have suggested that achievement motivation may differ across ethnic and racial groups because of differences in child rearing practices (McClelland 1961). Parental aspirations and expectations may vary, thus negatively affecting the performance of nonmainstream students (Brook et al. 1974; Dreger and Miller 1968; Rosen 1959; Resnick and Robinson 1975; and Wolff 1966).

Another perspective that has been considered is the issue of relative standing within the society. Friere (1973), Cummins (1986), and Pearl (1991) have argued for "a pedagogy of empowerment" that will help to overcome the oppressive social conditions to which children from linguistically and culturally diverse groups are subjected. Bernstein (1971), Laosa (1983), and Wilson (1987) identified the organizational structure of schools and instruction as a factor in perpetuating the failure of these students. Ogbu (1987), Ogbu and Matute-Bianchi (1986), Levin (1988), and Snow (1987) point to the presence of an "under class" whom teachers and school professionals in general do not believe or expect will or can excel academically or economically. Further, it is believed that these students ultimately come to internalize these limited expectations.

A similar line of argument that may have more direct bearing on reading achievement has focused on the disparity between home and school language and culture (Tharp 1989; Heath 1983; Boykin 1986; Boggs 1972; Vogt, Jordan, and Tharp 1987; Wiesner, Gallimore, and Jordan 1988; Philips 1972; Sindell 1988; Delgado-Gaitan 1987; Garcia 1989; Rivera-Medina 1984; Nine Curt 1984; Nguyen 1984; Wong-Fillmore 1983). These researchers have pointed to differences in discourse patterns, nonverbal communication, socialization, and learning styles.

Heath (1991, 13-14) clearly illustrates this position when she describes the difference in stance between the way mainstream groups who tend to "enculturate their young to fundamental beliefs and customs that undergird the school's criteria for successful displays of reading, writing, and critical thinking" by reinforcing their children as "individuals who have the right to voice their own judgements, ...[who are encouraged] to compare, complement, and supplement the information in the books they read,

...[who are expected] to display knowledge, consider its relevance for action, and challenge the ideas of others in gradually molded acceptable verbal forms," and that of many nonmainstream sociocultural groups who "traditionally orient their young to group membership and adherence to age and gender roles rather than to individual status."

Within schools, this difference in language socialization among ethnic groups is believed to play a major role in influencing academic achievement, particularly with regard to reading, writing, and forms of argument (Applebee 1981; Cazden 1988; Goodlad 1984; Heath 1983, 1985, 1986; Slaughter 1989). At least two opposing views have emerged with regard to how these differences might be accommodated. On the one hand, schooling activities might be brought into alignment with home values and practices. This position may be best typified by the Kamehameha Elementary Education Program (KEEP), which has led to academic achievement gains (Gallimore and Au 1979; Vogt, Jordan, and Tharp 1987). Alternatively, some argue that instructional programs must "ensure the appropriate application of general effective principles of instructional conditions that have academically significant effects across various conditions and groups" (Walberg 1986; Baden and Maehr 1986; Bloóm 1984; Slavin, Karweit, and Madden 1989; Rosenshine 1986; Bloom 1984). It is their contention that if there were systemic and effective implementation of these principles of instruction, minorities would no longer fail.

Although this survey does contain measures of instruction, at best, these measures offer only a limited look into instruction and do not address the interactions within classrooms. As we will argue, these measures are insufficient for the purposes of shedding much light on appropriate instruction for culturally and linguistically diverse students.

Language (S4Q15, S4Q16; S9Q21, S9Q22). Non-English speaking and limited-English-proficiency (LEP) students present a special, and perhaps extreme, category of the racially or ethnically diverse student. Because of the obvious connection between English language proficiency and English reading proficiency, this issue has been given somewhat greater emphasis in the survey instruments.

In an attempt to capture the effect that the use of a language other than English has on reading achievement within the U.S., two questions about language use were asked of the total sample. In response to the first question, which asked if the student spoke any language other than English at home, 76 percent of students in grade 4 and 79 percent of students in grade 9 reported always speaking English at home. On each of the scales, at both grade levels, the mean proficiency of these students was higher than that of students who spoke another language at home.

Responses to the second question, which asked students to identify the first language spoken as a child, showed a wide range of language usage within the United States. Among those students who reported that their first language was other than English (9 percent of all students in grade 4 and 10 percent of all students in grade 9), more than half identified Spanish as this language. These children represented between 5 to 7 percent of the total number of students included in the study. Of the remaining 4 percent of the students in grade 4, about half were distributed across the 10 languages included in most U.S. census categories. The other 2 percent reported speaking a collection of other languages. In grade 9, the remaining 3 percent of the population was divided such that 1 percent reported speaking the 10 identified languages and 2 percent reported using other languages.[9]

---

[9] Although students who reported using a language other than English were asked to respond to additional questions regarding their level of fluency, given their small number and the very wide distribution across language groups, we have chosen not to pursue any analysis regarding differences in fluency or differences among language groups.

These questions were then cross-classified so that we might have a measure of the extent to which a language other than English, either as the language spoken at home and/or as the first language learned, affected the performance on an English-language reading test. These two measures gave rise to four categories of response, which were then collapsed to three meaningful groups. The first group, those who always speak English, are not problematic since they have had no exposure to any language other than English. The second group consists of children who speak a language other than English at home or at school and are likely to have some level of English proficiency. In contrast, the third group is made up of the children of fairly recent immigrants and/or those who make a special effort to resist the acculturation pressures of the melting pot, and these children are less likely to be proficient in English (Table 13-4 and Figure 13-5).

As seen in the table, the data seem to indicate that students whose only language was English consistently outperformed students who had some exposure to or used languages other than English at home. Research has indicated that the reasons for this difference in performance may have many complex and interrelated causes. How to deal with this seeming disadvantage among students who are in the linguistic minority also poses a complex issue.

Although the Reading Literacy Survey identified almost 25 percent of the population who were likely to have had exposure to languages other than English at home (either as their first language or because it was spoken at home), only a few of these students could be characterized as either truly bilingual, of limited English proficiency, or as having no English proficiency. Research supporting this conclusion by Skrabanek (1970), Waggoner (1984), Hakuta (1986), Veltman (1988), and Hakuta and D'Andrea (1992) suggests that while school-aged children in the U.S. may continue to be bilingual, over time and across generations there is a shift toward the majority language. As reported in the *Condition of Education: 1992* (Alsalam et al., 22), as of 1989 only about 9 percent of the student population spoke a language other than English at home. Of that 9 percent, only about 28 percent (about 3 percent of the total population) would have been considered to be of limited English proficiency. According to another U.S. Department of Education report, in 1990-1991 there were approximately 2.3 million elementary and secondary students who could be classified as LEP (1992, 29-30). Yet another report cites estimates that range from 3.5 million to 5.3 million LEP students nationwide (CCSSO 1990). The actual number varies due to the lack of a uniform definition of limited English proficiency, compounded by the fact that states use different assessment measures to obtain these estimates of the target population.

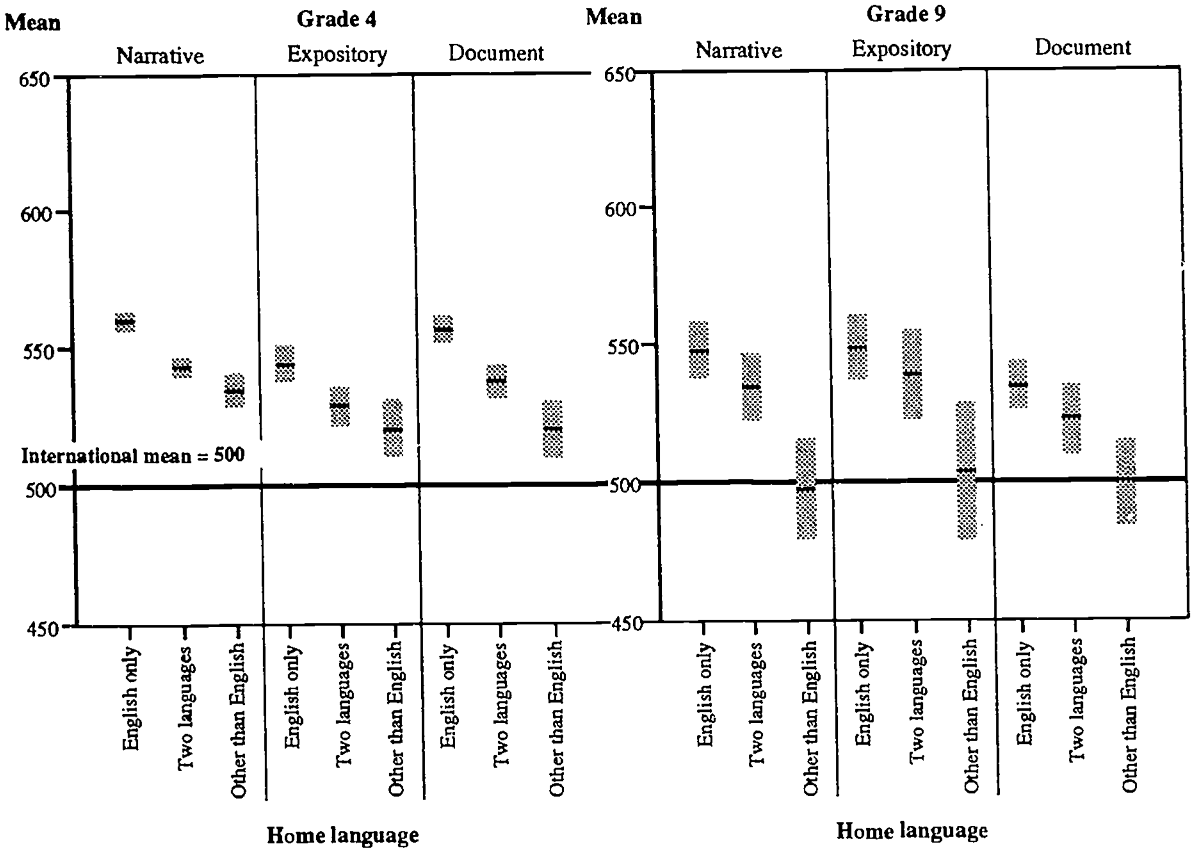
Table 13-4. Mean reading proficiency scores, by language usage: Grades 4 and 9

Language	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
English only	74.9	560 (3.1)	544 (3.1)	557 (2.6)	76.8	548 (5.0)	549 (5.9)	535 (4.0)
Two languages	17.1	543 (3.2)	529 (3.3)	538 (2.8)	14.2	535 (6.0)	539 (8.0)	523 (6.4)
Language other than English	8.0	535 (5.7)	521 (5.1)	521 (4.8)	9.0	498 (9.0)	504 (12.3)	500 (7.8)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-5. **Mean reading proficiency scores, with 95 percent confidence intervals, by language: Grades 4 and 9**



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Despite the proportionally small size of the LEP population, no matter how it is defined and counted, there has been a sharp increase in the number of these students within the last decade. According to the report *Language Characteristics and Schooling in the United States, A Changing Picture: 1979 and 1989* (McArthur 1993, p. 12), the National Center for Education Statistics reports that, "among school-age children (5 to 17 years old) there had been great changes in the numbers of speakers of languages other than English. Although the total number of children in this age group dropped by about 3 million over the decade to about 42.2 million in 1989, the number of speakers of languages other than English grew from 3.8 million to 5.2 million (from 8 percent to 12 percent of all school-age children." The Hispanic enrollment increased from 6.4 percent to 10 percent as a proportion of the total enrollment. During this same period, the number of Asian and Pacific Islander students increased by more than 116 percent, from 535,000 to 1,158,000 (Ogle et al. 1991). But perhaps more importantly, this growth is having a disproportionate impact on concentrated geographic regions. Approximately 75 percent of these students reside in just eight states.[10] Their presence in concentrated groups is transforming schools

---

[10] These states are Arizona, California, Colorado, Florida, New Jersey, New Mexico, New York, and Texas (Dorothy Waggoner, "Language Minority Census Newsletter." Available from Numbers and Needs, Box G1 H/B, 3900 Watson Place NW, Washington, DC 20016, 1991).

dramatically, and this is particularly true in the nation's 10 largest school districts, which happen to be situated in states that are fast becoming minority/majority states.

The overall increase of LEP students across the nation, coupled with their concentration in many school districts such that they constitute a majority of the student body, underscores the need to examine the policies directed at the growing number of LEP students in our schools.

The public policies mandated for limited-English-proficient students have changed dramatically during the past 25 years. Today, there is little discussion about the eradication of a student's linguistic and cultural heritage into the American melting pot. Neither is there discussion among psychologists that "the child reared in a bilingual environment is handicapped in his language growth" (Thompson 1952).

This change in attitude can be attributed, in large part, to the two primary Federal interventions targeted at improving the educational opportunity of LEP students. The first was Title VII of the Elementary and Secondary Education Act (1968), which provided funds for innovative bilingual education programs, the training of teachers, and the development of instructional materials. It was limited to students from poor backgrounds and did not prescribe use of the native language or culture in instruction. By 1974, Title VII was expanded to include students regardless of their economic background, and it required the use of the child's native language and culture "to the extent necessary to allow a child to progress effectively through the educational system."

This legislation was followed by the U.S. Supreme Court decision known as *Lau v. Nichols* (414 U.S. 563,1974), which held that school districts had to "take affirmative steps to rectify the language deficiency (of LEP students) in order to open its instructional program to these students." The court decision was based on Title VI of the Civil Rights Act and ruled that students with limited English proficiency, in the absence of treatment, were "effectively foreclosed from any meaningful education."

These policies may have found their basis, in part, from research in language acquisition and the cognitive effects of bilingualism. The early research by Peal and Lambert (1962), followed by Bain and Yu (1980) and Hakuta and Diaz (1984), showed that bilingual children who continue their cognitive development in both languages throughout the elementary school years frequently outperform monolinguals on measures of cognitive flexibility, linguistic and metalinguistic abilities, concept formation, divergent thinking skills, and creativity. More recent work both internationally (Swain and Lapkin 1991) as well as in the U.S. (Kessler and Quinn 1985, 1987; Galambos and Hakuta 1988) continues to support this position. Bilingualism in these studies is viewed as a cognitive asset, not an intellectual handicap (Hakuta 1986; McLaughlin 1984, 1985). However, Garcia (1991) cautions that "any detailed conclusions concerning the relationship between the bilingual character of children and cognitive functioning must remain tentative."

In addition to cognitive factors, it is important to consider the social/communicative aspects of language learning. The learner's attitude toward the second language and members of that cultural group can affect the language acquisition process. For example, in looking at the impact of attitudes on language learning, Gardner and Lambert (1972) found that the positive attitude of English-speaking Canadians toward their French peers led to high integrative motivation to learn French. Canale (1983), Cazden (1988), Halliday (1975), Heath (1983), Hymes (1974), and Ramirez (1985) demonstrated that there is a strong correlation between positive attitudes toward the target language and successful language learning. In addition to attitude, the relationship between the two cultures is also important. Studies conducted by Schumann (1976) revealed that the greater the social distance between the two cultures in question, the greater the difficulty the student will have in learning the target language.[11]

According to Ramirez and Merino (1990), within the U.S., four alternative instructional programs typify the services available to LEP students:[12]

- **Submersion,** where students are in ordinary mainstream classrooms with all instruction in English.
- **English as a second language (ESL),** where language minority students receive most of their instruction in English and additional instruction in English as a language.
- **Transitional bilingual education (TBE),** where students study subject matter in their primary language until they develop sufficient proficiency in English so that they can survive in English-only classrooms. Reading is taught in the primary tongue first and English is taught as a separate subject. These programs take two forms: early exit, in which development and/or maintenance of the native language is not a goal; and late exit, in which the native language continues to be developed in addition to English.
- **Structured English immersion strategy (SEIS),** where students receive all instruction in English, but teachers tailor their English to what LEP students can understand. SEIS teachers are bilingual and may speak to students in their native tongue; students may respond to the teacher and to other students in their native tongue.

The debate hinges on the use of the native language and English as the vehicle for instruction (for a thorough discussion of this issue, see August and Garcia 1988; Baker 1990; Hakuta 1986; Hakuta and Gould 1987; Rossell and Ross 1986; Secada 1990; Troike 1981; and Willig 1986). When specifically considering learning to read, questions regarding bilingualism seem to take on even greater importance. Researchers generally agree that learning to read is a complicated process, but is it more complicated when moving across languages? Should children be taught to read in their native tongue or in the language they are learning, and are reading skills transferable from language to language?

At one end of the continuum, the student's native language is used for instructional purposes, and mastery of that language is viewed as critical before English is introduced into the curriculum. This method is based on the notion that competencies in the native language, particularly with regard to academic learning, provide the necessary cognitive and social foundation for second language learning. According to Hudelson (1987), "You really only learn to read once." Following along this line of reasoning, some would argue that bilingual programs should concentrate on providing literacy skills in the home language, particularly for those children whose parents have little education and poor literacy skills, and once the basic skills of reading have been acquired in the home language, reading skills can be transferred to a second language (Goldenberg and Gallimore 1991; Goldman and Trueba 1987; Snow 1983). But this assumption has been challenged by other researchers (Goodman, Goodman, and Flores 1989).

---

[11] Schumann defines social distance as the relative status of two cultures (i.e., two cultures that are equal in political status have less social distance than those where one is dominant and the other subordinate).

[12] In addition, there has been some recent growth in "two-way bilingual" and "developmental bilingual" programs whose goals are to produce bilingual/biliterate students (Garcia 1992).

At the other end of the continuum, English is introduced at the beginning of the student's educational experience, with little or no use of the native language. The theory behind this approach is based on the concept of time on task; that is, the more the student is exposed to English, the greater the likelihood of learning the language more rapidly (Baker and deKanter 1983; Rossell and Ross 1986). In certain instances, where the diversity of languages represented in the student population of the district or even the school building itself is great (i.e., New York City where students from over 180 different languages are currently enrolled in the public schools), this may be the only feasible solution.

There have been a number of evaluation studies designed to measure the comparative advantage of one approach versus the other (e.g., Ramirez et al. 1991). In general the findings have not been conclusive. However, Garcia (1993) argues that these studies, while focusing primarily on language acquisition, have overlooked the larger social issues associated with being a cultural minority (see discussion on race/ethnicity). He suggests that a rethinking of how instruction is situated in the larger social context may be in order.

In line with Garcia's argument for considering issues related to language acquisition within a broader social context, in the succeeding chapters we address his concern by examining the effect of language after having controlled for all student and family variables. This essentially does what he requests by examining only the residual effect, as if all other things were equal.

13.3.2. Family Attributes

The role of the family in a child's education is too often underestimated. Many would argue that "Families play the central role in nurturing very young children, in shaping the character of older children, in inculcating the habits that make learning possible, and in choosing the specific knowledge valued by that specific family. A case can be made that, when 'education' is broadly conceived, families play the most important role" (Weston 1989, 2). In the Reading Literacy Study, a number of variables related to families, their attributes, and the environment created by them were included. These variables have their basis in the research literature and are commonly organized in two ways: family structure and status, and family process. For the purposes of this report, the two sets of variables will be considered as family attributes and family environment. This first section addresses issues of family attributes and includes family composition, size, and socioeconomic status (SES) as measured by family wealth and parental education.

Items categorized as family attributes focused on those preexisting attributes of families that are not amenable to change by schools or public policy. These attributes colored students' experiences outside of school. Four constructs were derived from among the variables:

- Family composition (S4Q13, S4Q14; S9Q19, S9Q20);
- Augmented family (S4Q13, S4Q14; S9Q19, S9Q20);
- Parental education
    - Father's education (S4Q5, S9Q5)
    - Mother's education (S4Q6, S9Q6); and
- Family wealth (S4Q11, S9Q17).

Family Composition (S4Q13, S4Q14; S9Q19, S9Q20). In the Reading Literacy Study, students were asked two questions with regard to family composition. The first was an indication of how many people lived in the household, and the second indicated the relationship of those people to the respondent.

With data of this kind a variety of distinctions between families become possible. We chose to make the following distinctions: first, based on the configuration of parents/guardians present, to capture two- and one-parent families, along with biological and guardian parents in their various combinations; and second, to consider the distinction between "nuclear" families involving only parents/guardians and children, and "augmented" families that include one or more of grandparents, other relatives, and/or other nonrelatives.

In the interests of simplification we adopted the categories of parental status shown in Table 13-5 and Figure 13-6. Students indicating that their father and/or mother was present are referring to what we may assume are biological parents in the eyes of the child. Where a student indicates a male and/or female guardian present, we assume that these are stepparents or de facto stepparents. What is at issue here is whether the family might be considered a "regular" family or one that has been disrupted in some way.

In Table 13-5 we capitalize the first letter in biological parents and show them as Mother/Father. In contrast, guardians are shown without capitalization as mother/father. The various combinations of biological parents and guardians are identified in this way. For example, Mother+father indicates a biological mother and a stepfather present; Mother indicates a single-parent family headed by a biological mother; and so on.

In the interests of further simplification, the three types of families involving two parents where one or both were guardians are treated as a single grouping, as are the parental combinations reported as nonconventional. The results show clearly that students in nonconventional families have substantially lower scores than those with two biological parents present, across grades and scales. The mean scores estimated for the other family composition types consistently fall between these two extremes.

The available survey data for comparison purposes come from NELS:88. The only thing to be discerned from the comparison is that the distributions across the parental configurations in both data sets are basically analogous. About 64 percent of students live with their own biological mother and father, and another 14 percent live with a mother and male guardian or father and female guardian, for a total of 78 percent living with two parents.

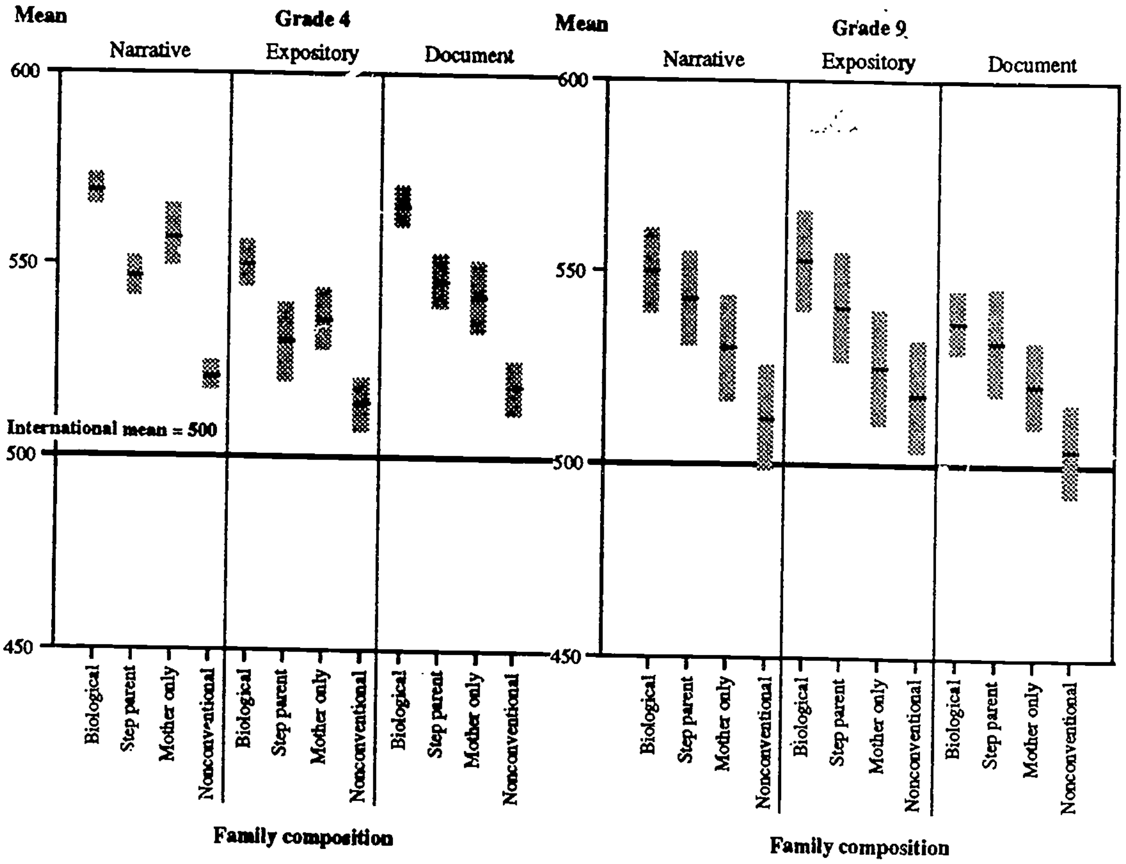
Table 13-5. Mean reading proficiency scores, by family composition: Grades 4 and 9

Family composition	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Two Parent (Mother + Father)	57.4	569 (3.1)	550 (2.8)	565 (2.6)	60.6	550 (5.5)	553 (6.3)	537 (4.0)
Two parent (M+f, m+F, m+f)	9.5	547 (4.8)	530 (5.2)	546 (3.5)	13.5	543 (6.1)	541 (7.0)	532 (6.8)
Mother only	10.7	556 (6.3)	536 (4.0)	542 (4.8)	13.2	530 (6.7)	525 (7.5)	521 (5.5)
Nonconventional (no parents, m-only, f-only, F-only, other)	22.3	521 (3.6)	514 (3.5)	518 (3.3)	12.8	512 (6.8)	518 (7.3)	504 (5.0)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-6. **Mean reading proficiency scores, with 95 percent confidence intervals, by family composition: Grades 4 and 9**



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As is the case in most of the research literature, children from two-parent families tend to do better on the reading literacy tests than those from families that have been disrupted in some way (Milne 1989). Even so, the differences between two parent (mother and father) and mother only for grade 4 narrative, and between two parent (mother and father) and two parent (other) for all three scales for grade 9, are not statistically significant. Interestingly, at grade 4, children in mother-only families do no worse on the three scales than those in families where parents have remarried.

Augmented Family (S4Q13, S4Q14; S9Q19, S9Q20). A further classification can also be made based on the data available, differentiating between augmented and nuclear families. This variable focused on the question of whether people outside the nuclear family were present in the household. Two categories were developed: parent(s)/siblings only, and parents/siblings/grandparents/relatives/others.

In making a distinction between nuclear and non-nuclear families we resisted the use of "extended" to describe the non-nuclear families. "Extended family" has a special meaning not captured completely in the measure used here since the term tends to be used to refer to an extended network of mutually supportive social relations not necessarily restricted to persons within the same household. Since all we know is whether or not there are non-nuclear family members of the household, and whether these people are relatives or not, we adopted the term "augmented families" to describe such situations. The distribution of the two types of families is displayed in Table 13-6 and Figure 13-7.

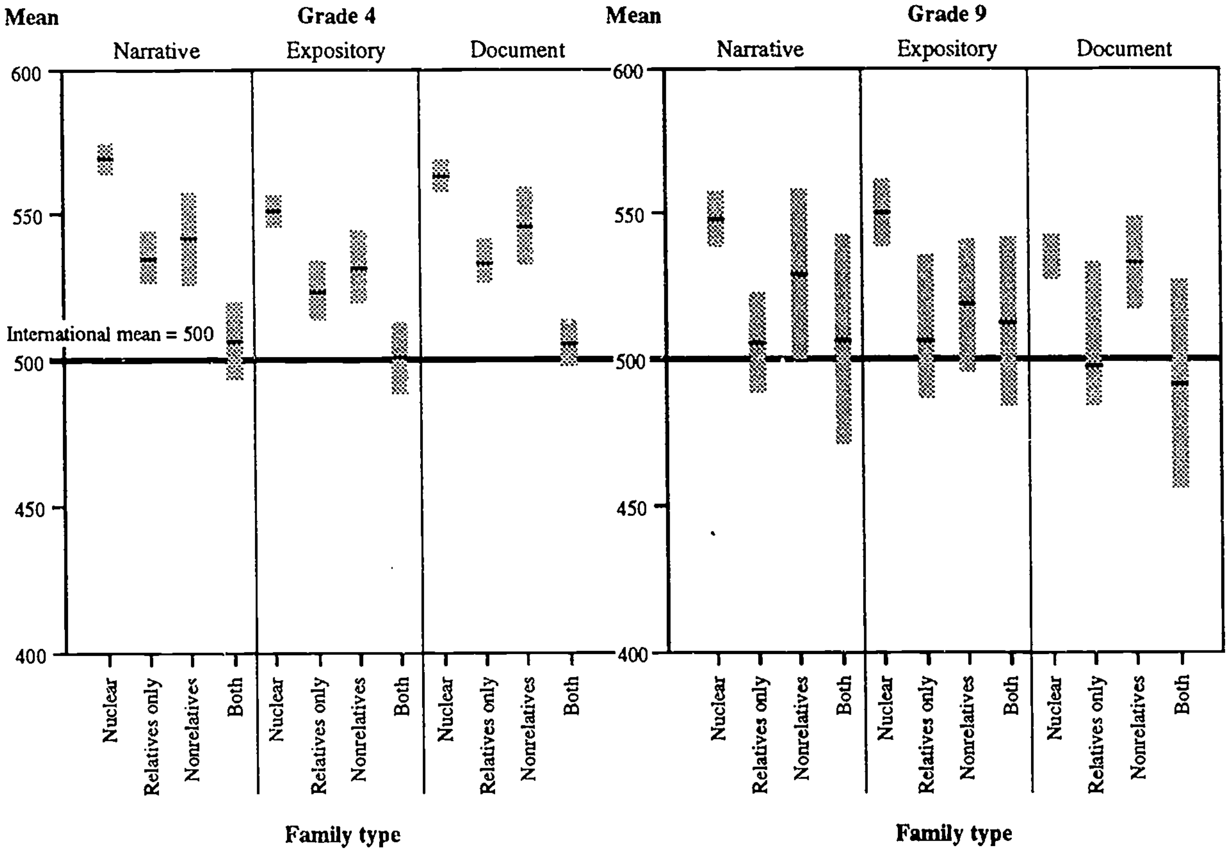
Table 13-6. Mean reading proficiency scores, by nuclear and augmented families: Grades 4 and 9

Type of family	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Nuclear	64.3	569 (2.6)	551 (2.7)	563 (2.5)	83.9	548 (4.8)	550 (5.7)	535 (3.8)
Relatives only	22.3	535 (3.5)	522 (3.6)	533 (2.9)	11.3	506 (8.3)	507 (10.1)	498 (6.7)
Nonrelatives only .	13.2	542 (7.9)	532 (6.0)	546 (6.6)	4.8	529 (14.5)	519 (11.2)	533 (7.8)
Both rel. and nonrel.	7.6	507 (6.3)	501 (6.0)	506 (4.0)	1.4	507 (17.7)	513 (14.3)	492 (17.5)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-7. Mean reading proficiency scores, with 95 percent confidence intervals, by nuclear and augmented families: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Data regarding the presence of others in the household are not available from either NAEP or NELS. Looking at the Reading Literacy Study data, we note that children living with just their immediate family appear to do better than those who live in families where more adults are present. This may appear to be somewhat counterintuitive because with more adults present, the child might be more likely to receive more attention. However, one might hypothesize that the presence of more adults could be associated with lower economic status, and that might be the more important factor.

Given the amount of time children spend at home, both before they begin school and then during the time that surrounds school attendance, it is not surprising that the Commission on Reading reported that "parents play roles of inestimable importance in laying the foundation for learning to read" (Anderson et al. 1985, 27). But even beyond the parent, the family environment more generally is known to have an impact on student performance.

There have been numerous studies related to family structure and school-related performance. Family structure has been looked at from many perspectives. These have included the number of parents in the family, the family size, birth order, and gender distribution. These structural variables have also been considered in relation to changes in economic status, parental time availability, and parental role models. In addition, the reasons for a particular configuration, perhaps disruption due to divorce or death, are considered to have an important impact on school-related performance.

Currently we are witnessing a dramatic change in the configuration of the American family, and these changes are thought to be having an impact on how proficient children are as readers when they enter school. The picture of the "ideal" family as a father, mother at home, and two children is eroding as divorce rates, the number of children born to unwed mothers, and the number of mothers entering the labor force climb. The annual number of divorces has increased approximately 120 percent between 1965 and 1989, and about 1 million children are involved in divorces each year (Snyder 1991, 17). While overall birth rates have been declining since 1950 (106.2 per 1,000 in 1950 to 63.0 per 1,000 in 1988), the number of births to unmarried women has been increasing (14.1 per 1,000 in 1950 to 38.6 per 1,000 in 1988; Snyder 1991, 20, 22). And the labor force participation rate of women with children under 18 years of age has markedly increased from 11.8 million in 1970 to 22.3 million in 1991. In total, 57 percent of women aged 16 and over were part of the labor force in 1991. Fifty-two percent of the women in this age cohort who had no children were in the labor force, as compared to 66.6 percent of those with children under the age of 18 (U.S. Department of Labor 1991).

But perhaps as importantly, the change in family structure from the traditional is not consistent across various racial/ethnic groups and socioeconomic groups. For example, while only 16.8 percent of white families were single-parent families in 1989, the corresponding figure for Hispanics was 28.4 percent and for blacks, 54.2 percent (Snyder 1991, 29). Consequently, the number of children who might be considered at risk of educational failure is growing as more children are confronted by combinations of risk factors.[13]

The most common perspective in the research literature has been to look at whether the father is present or absent, with only a few studies looking at different parental configurations. Although the bulk of the evidence favors the children of two-parent families (Milne 1989), even in early research syntheses, reviewers cautioned that the effects of father absence were generally less significant when SES and race were taken into account (Herzog and Saudia 1973; Shinn 1978). Despite the more profound impact of economic or racial factors, it was still believed that the absence of a father had a detrimental

---

[13] Pallas, Natriello, and McDill (1989) define an at-risk child as one who has been exposed to certain background factors or experiences in formal schooling, family, or community. Examples of these are single-parent homes with low incomes and parents with limited English proficiency who have no high school diploma.

impact, particularly on boys who would lack a role model, and would more negatively affect quantitative skills (Shinn 1978). Two more recent reviews of the literature (Hetherington, Featherman, and Camara 1982; and Salzman 1987) continue to support a position favoring two-parent families--children in one-parent families received lower achievement scores than did children in two-parent families, although the overall differences were quite small.

There have been only a few studies of different family configurations. Shilling and Lynch (1985) compared single-parent families headed by men with those headed by women and concluded that children living with fathers achieved less well than those living with mothers. Again, the difference was slight. Ganong and Coleman (1984) compared children in intact and/or single-parent families with those in families where parents have remarried across 38 studies and concluded that there were no differences in cognitive outcomes. Studies by Hett (1983), Boyd (1984), Collins (1981), and Gray (1980) seem to indicate a trend favoring children of intact families, followed by children whose parents have remarried, then by children in single-parent families.

What is important about family configuration may not be who is present, but rather how parent and child interact. With regard to learning to read, more often than not it is the parent who provides the first major access to literacy events through storybook reading. It is this significant adult who mediates the reading in response to the child's reactions. It is the social link between the literacy act and the child that brings about learning (Pellegrini, Brody, and Sigel 1985; Heath 1982; Ninio 1980; Sulzby and Teale 1987).

Heath's (1982) work comparing middle-class and working-class parents suggests that it is more than just presence and the interaction, but also the quality of interaction that affect reading achievement. She notes that although parents from both groups read to their children, and the children from both communities did well in elementary school, where there was a focus on recitation and low-level skills, the children of middle-class parents outperformed those of working-class parents when the children reached upper elementary school and the focus shifted to higher level comprehension. The social interaction between parent and child depended on the adult's ability to mediate between the child and what was to be learned. The more educated the parents, the more likely they were to be able to assist the child beyond the information given. So perhaps more important than the absence or presence of a particular parent, it is the environment provided and the time for and qualities of interaction that may more greatly influence reading achievement.

Parental Education: Father's and Mother's Education (S4Q5,S4Q6; S9Q5,S9Q6). Measures of father's and mother's education were obtained through the respondent's report of the highest level of schooling completed by each parent or guardian. The measure itself allowed for responses on a six-point scale ranging from elementary school to college or university. To simplify, we have reduced the categories so that elementary school, junior high school, and some high school form one category -- less than high school. In both grades, this combined category represents less than 10 percent of all students sampled (Tables 13-7 and 13-8; Figures 13-8 and 13-9).

Similar data on parental education is available from two sources, NAEP and NELS:88. While the response categories differ between the Reading Literacy Study and NAEP, the NAEP data provide the only comparison available for fourth graders. For eighth graders the comparison to NELS is more appropriate as the exact item was used. The data in NELS were compiled from responses by parents about their own educational attainment and, as such, provide a benchmark for other surveys. Even a cursory comparison makes it obvious that, except for children of high school dropouts, students in the Reading Literacy Study and NAEP tended to inflate the educational attainment of their parents. However, despite the inaccuracies in their reporting, the same basic pattern between parents' education and students' reading proficiency is attained. The higher the perceived level of parents' education, the higher the reading proficiency of the child.

Indicators included in *The Condition of Education 1993* (Alsalam et al. 1993) provide a good picture of the relationship between parental educational attainment and overall student achievement. These indicators show fairly consistent differences in the educational achievement and attainments of students from families whose educational backgrounds differ. For example, the following comparisons originated in that report:

- "Students whose parents had not completed high school were much less likely then those whose parents had completed college to describe their high school program as academic/college prep (28 percent compared to 68 percent) and much more likely to describe their program as general (52 percent compared to 28 percent) or vocational/technical (20 percent compared to 4 percent).
- "High school students whose parents have not completed high school have lower average academic achievement than students whose parents have completed some college.
- "Among students who take the SAT, both verbal(V) and math(M) scores were higher among those whose parents had more education."

Table 13-7. Mean reading proficiency scores, by father's education: Grades 4 and 9

Educational attainment	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
No father	3.0	543 (10.0)	522 (8.3)	535 (7.3)	1.7	507 (14.0)	498 (17.4)	493 (12.9)
Less than high school	9.5	520 (6.1)	508 (4.2)	518 (4.8)	11.9	497 (7.8)	496 (7.8)	502 (8.4)
Completed high school	23.5	546 (3.5)	531 (3.4)	545 (2.7)	33.3	530 (5.5)	529 (5.5)	519 (4.4)
More than high school	16.9	557 (4.1)	542 (3.0)	552 (3.3)	19.0	548 (7.1)	546 (7.1)	531 (5.9)
Completed college or university	47.1	567 (3.4)	550 (3.4)	561 (3.1)	34.1	567 (5.6)	575 (5.6)	553 (4.2)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

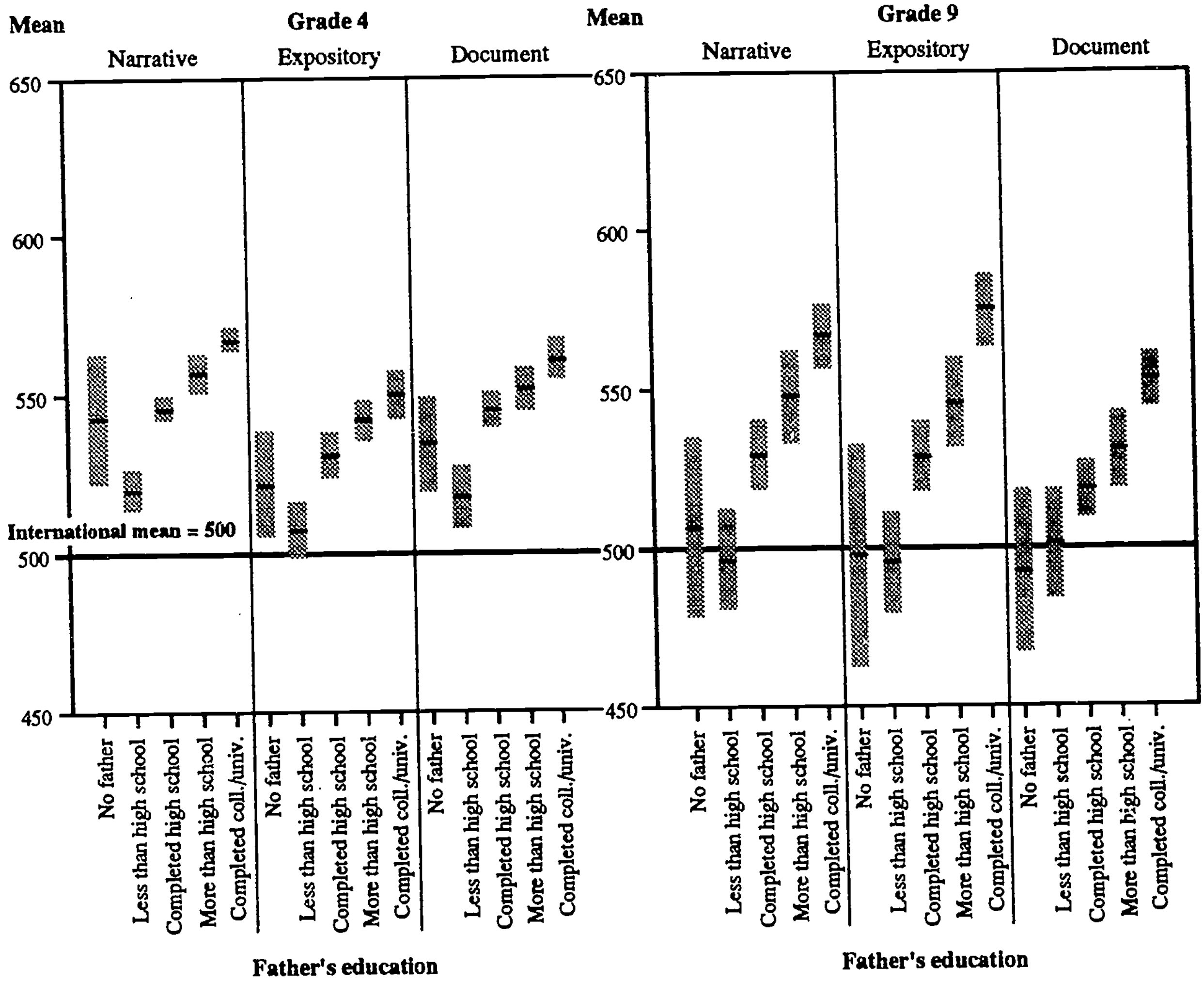
Table 13-8. Mean reading proficiency scores, by mother's education: Grades 4 and 9

Educational attainment	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
No mother	0.8	502 (10.3)	516 (10.1)	501 (8.1)	0.3	438 (30.0)	408 (20.0)	444 (44.6)
Less than high school	8.6	516 (4.7)	509 (4.5)	516 (4.1)	10.8	503 (8.2)	498 (7.1)	501 (6.8)
Completed high school	26.3	552 (4.1)	534 (3.2)	546 (3.1)	34.4	531 (6.1)	530 (7.5)	523 (4.6)
More than high school	20.3	557 (4.4)	539 (3.7)	551 (3.2)	24.3	552 (5.5)	550 (6.8)	534 (4.9)
Completed college or university	44.0	565 (3.1)	549 (3.1)	562 (3.0)	30.2	562 (4.8)	574 (5.3)	549 (3.8)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

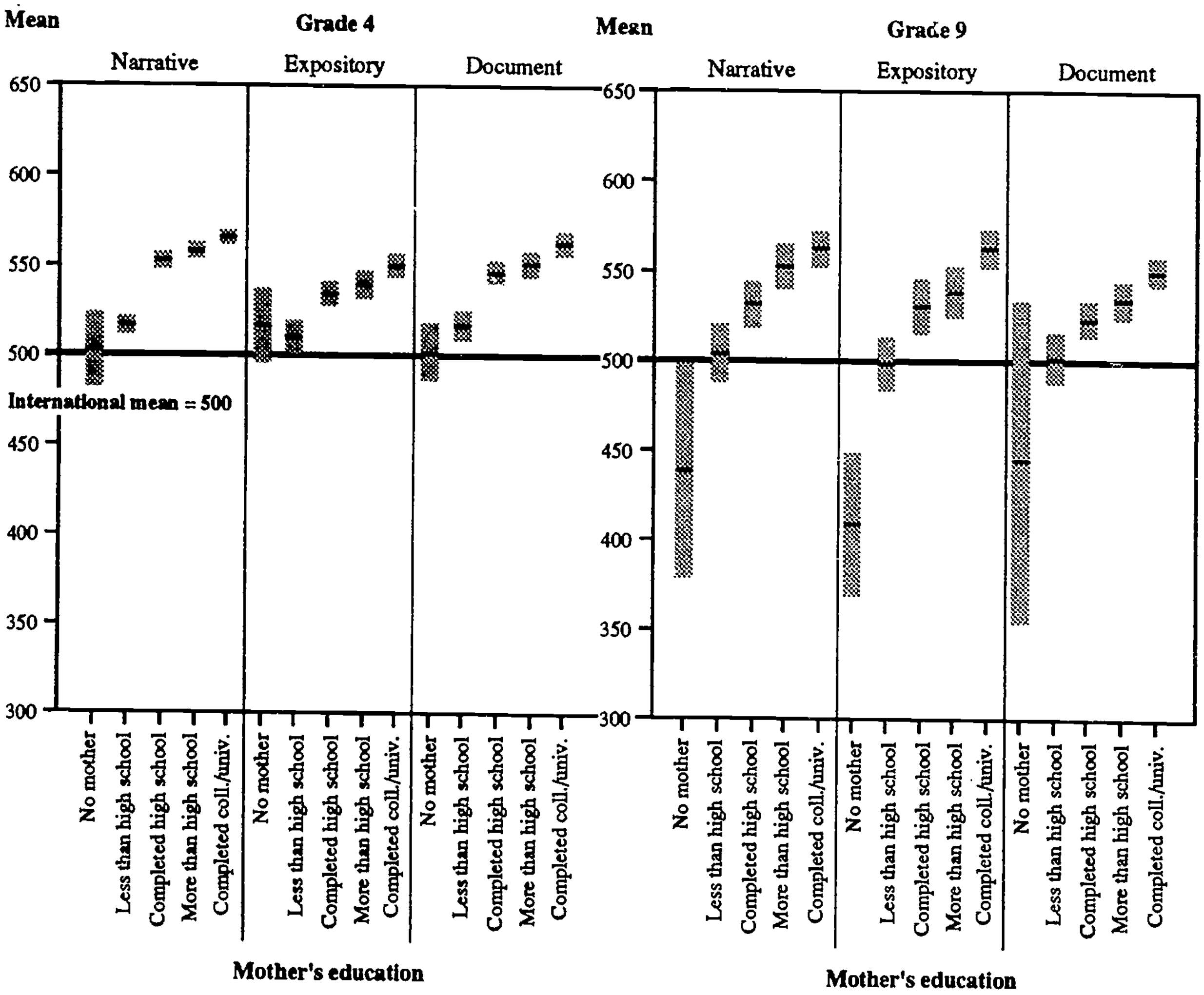
Figure 13-8. Mean reading proficiency scores, with 95 percent confidence intervals, by father's education: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-9. Mean reading proficiency scores, with 95 percent confidence intervals, by mother's education: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

NELS:88 parent survey data provide a good picture of how parental educational attainment relates to their expectations for their child's future education. The findings indicate that as parents' levels of education rose, so did their expectations for their child's educational attainment; for example, 45 percent of single mothers with college degrees expected their eighth grader to graduate from college, compared with 34 percent of single mothers with a high school diploma and 21 percent of those who did not graduate from high school (Horn and West 1992, 40).

The 1979 review of the literature by Hess and Holloway is more pertinent when considering reading proficiency specifically. Their conclusions suggest that more highly educated mothers produce children with higher preschool oral language skills, and that this is accomplished by the mother speaking more with the child, in a more intelligible manner, and about a wider range of topics known to the child. In addition, the more highly educated parents are more likely to spend more time reading books to their children, thereby exposing them to a literate manner of speech. They also are likely to have more reading materials around the house and to spend more time reading, thus, modeling the value of reading and how reading materials are used.

In looking at ways to ameliorate or change the patterns of intergenerational illiteracy or less than full literacy, a number of researchers have studied whether it was possible to improve the interactive style between parents and children. Edwards (1992) and Heath and Thomas (1984) demonstrated that parents can be taught to read with their children in the interactive fashion that would facilitate greater learning. Whitehurst et al. (1985) trained parents to increase their rates of asking open-ended questions, commenting on functions and attributes, and expanding children's contributions and to decrease asking questions that could be answered simply.

This literature points to the importance of the underlying relationship, the interaction at home, rather than the fact of a particular level of education attainment by the parents. These kinds of interactions seem to be highly associated with level of education, but are not necessarily caused by that level of education (Sulzby and Teale 1991, 743).

Family Wealth (S4Q11; S9Q17). Since children's reports of family income and/or assets are likely to be unreliable, if obtainable, an indirect measure of family wealth was devised based on household possessions of the respondent's family. Students were required to respond to each of 18 items describing family possessions by indicating whether none, one, or more than one was present in their home. The items were consumer durables in the main and included automobiles, TVs, stereo receivers, VCRs, telephones, calculators, encyclopedias, and the like.

The 18 items broke logically into two distinct subgroups: a set of 12 to do with household items of the consumer durable kind (automobiles, TVs, clothes dryers, dishwashers, refrigerators, and the like); and a set of 5 that appeared to tap the kinds of school-related possessions that families make available for the children (computers, dictionaries, calculators, typewriters, and encyclopedias). (The 18th item, recreational vehicle, did not load on any factor.) Since the items separated in this way in factor analyses as well, and because previous work with family wealth indicators of this kind has concentrated on the consumer durables, we considered only the 12 consumer durable items for the purpose of developing a wealth indicator.

A further consideration of these items suggests that, as indicators of family wealth, they are confounded by family composition, family size, and the number of persons in the household. For example, poor multifamily households and/or large families might well be indistinguishable from wealthy small nuclear families in terms of the number of possessions they have. In order to address this matter to the extent possible using these data, each possession item score was divided by the number of persons in the household to create a possessions per person measure as a better indicator of family wealth.

A factor analysis of the relations among the items indicated only one factor with an eigenvalue greater than one, this factor explaining some 45 percent of the variance. Table 13-9 summarizes the results of these analyses.

Table 13-9. Family wealth: Factor pattern

Factor loading	Item
0.79	phone
0.72	color TV
0.69	bicycles
0.67	recorder
0.62	automobile
0.60	refrigerator
0.59	microwave oven
0.58	VCR
0.57	stereo
0.56	clothes dryer
0.42	dishwasher

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

On this basis, factor scores were estimated to create a measure of the family wealth construct for each student. Where this continuous measure of family wealth was categorized for the purpose of presenting observed and adjusted means, the distribution was grouped into four categories, given by the population quartiles. This maintained sizable samples in each group, while giving four groups of students who differed substantially with regard to this factor (Table 13-10 and Figure 13-10).

Table 13-10. Mean reading proficiency scores, by family wealth: Grades 4 and 9

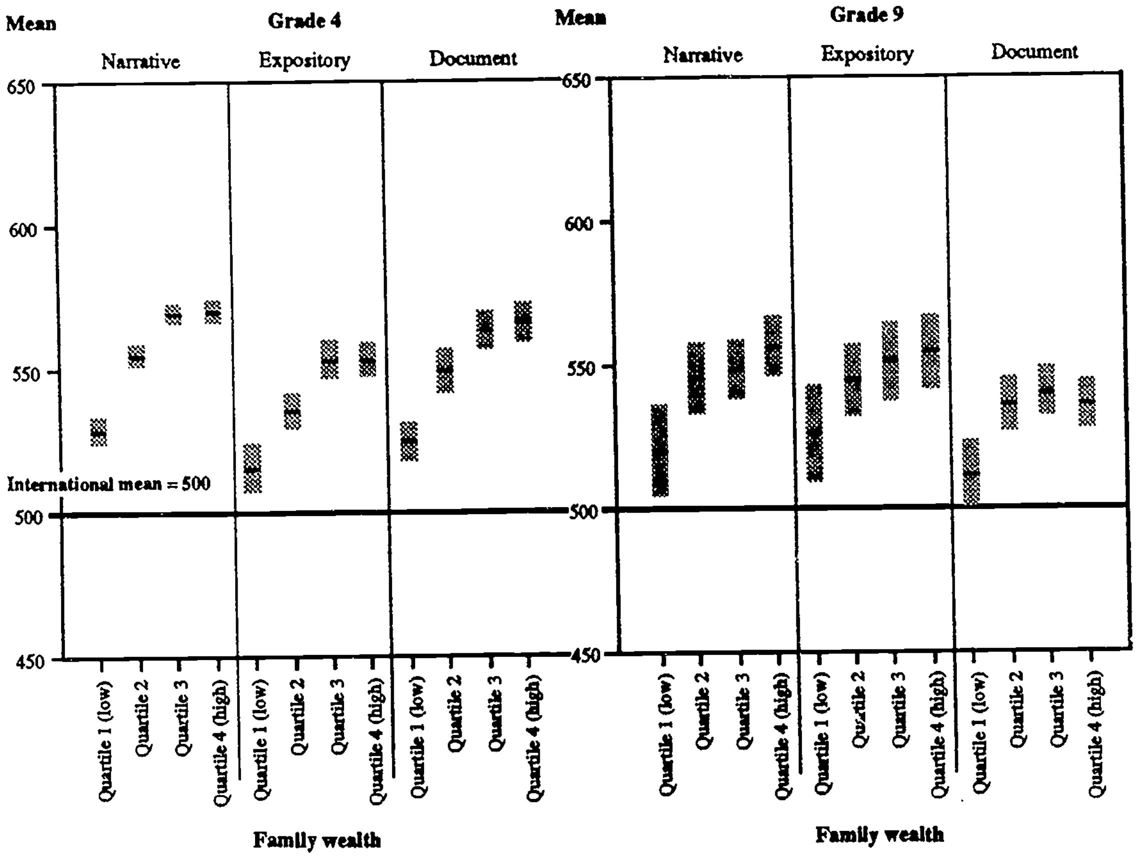
Family wealth	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Quartile 1 (low) . .	24.1	528 (4.8)	515 (4.1)	524 (3.2)	25.9	520 (7.9)	526 (8.5)	511 (5.7)
Quartile 2	25.4	554 (3.7)	535 (3.2)	549 (3.8)	25.0	545 (6.3)	544 (6.2)	536 (4.6)
Quartile 3	25.6	569 (3.1)	553 (3.3)	563 (3.2)	24.3	548 (5.0)	551 (6.7)	540 (4.5)
Quartile 4 (high) .	24.9	570 (3.3)	553 (2.9)	566 (3.5)	24.8	556 (5.3)	554 (6.3)	536 (4.3)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The truth of the commonly held expectation that the wealthier one's family is, the higher the probability that one will achieve at higher levels is apparent in the Reading Literacy Study data. Clearly, children from families in the highest quartile of family wealth have higher mean scores on all the reading scales at both grade levels than do children in the lowest quartile. The students in the highest (fourth) quartile of family wealth did not have significantly higher means than those in the third quartile, for either grade, indicating either that the benefits of family wealth diminish rapidly above the median, or that the method of reporting family wealth is not reliable for distinguishing wealth differences at higher levels. For grade 4, the third quartile notably had higher means than the second quartile, but this finding was not evident for grade 9. This suggests that apart from the effects of outright poverty, family wealth, as measured in this study, has little impact at older ages. The differences between the second and lowest quartiles were the most marked among successive quartiles for both grades and were statistically significant in all cases, with the exception of the expository scale at grade 9, where the estimated difference falls just short of being statistically significant.

Figure 13-10. **Mean reading proficiency scores, with 95 percent confidence intervals, by family wealth: Grades 4 and 9**



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As documented in *The Condition of Education 1993*, "a student's family has a strong influence on his or her educational achievement" (Alsalam et al., 3). This seems to be particularly true when we examine differences between children in low-, middle-, and high-income families. Consistently, children from low-income families were seen to be at a disadvantage. For example,

- "Children from low income families were less likely than children from high income families to get a head start. In 1991, 22 percent of low income children were enrolled in pre-K compared to 53 percent of high income children (Table 2-3). This may be due to the expense of nursery school -- 63 percent of pre-K enrollment was in private schools.

- "First grade children from low income families were more likely than children from high income families to be 7 or older in October -- 27 percent compared to 18 percent. This may have been due to a higher percentage of children from low income families repeating kindergarten or first grade.
- "Fourth grade children from low income families were more likely to be 10 or older -- 39 percent compared to 20 percent; and seventh grade children from low income families were more likely to be 13 or older -- 48 percent compared to 18 percent.
- "The difference between the percentage of children from low and high income families who are above the typical age for their grade increases from 9 percent in first grade, to 19 percent in fourth grade, to 30 percent in seventh grade."

This pattern of findings is repeated in the NELS:88 reports. For example, among eighth graders in 1988, 3.2 percent of students from the lowest SES quartile had repeated kindergarten compared to 1.8 percent from the highest SES quartile, and 8.5 percent of students from the lowest quartile had repeated first grade compared to 1 percent of students from the highest quartile (Hafner et al. 1990, Table 1.3). Similarly, 31 percent of students from the lowest SES quartile had repeated at least one grade compared to only 8 percent of those from the highest SES quartile.

This trend of differing school outcomes for children from low and high income families continues throughout schooling. As reported in *The Condition of Education 1993* (Alsalam et al.), "a higher percentage of high school students from low income families drops out of school each year than students from high income families. Between October 1990 and October 1991, 11 percent of students from low income families dropped out compared to 1 percent from high income families. Consequently a larger percentage of 19- to 20-year-olds from low income families was out of school and had not finished high school -- 30 compared to 3 percent." As reported in *Dropout Rates in the United States: 1991* (U.S. Department of Education 1992, 17, Table 10), differences in family income may account for most of the differences in dropout rates between racial/ethnic groups. When comparisons are drawn across racial/ethnic groups within an income level, there was no difference in status dropout rates of white and black 16- to 24-year-olds in 1991. The rate for Hispanic 16- to 24-year-olds were, however, higher than for whites within each income level.

As Dornbusch and Wood (1989) point out, "the huge literature in sociology that relates student performance to various statuses of the parents--parental education, family income, race and ethnic background, and family structure--produces a careful statistical portrait of education in our society, showing how educational performance is correlated with background variables such as ethnicity, family income, parental education, or family structure." The data in this study replicate this statistical portrait. However, as Williams et al. (1987, 1993a, 1993b) argue, while each of these variables contributes to the statistical picture, they are likely to have differing effects on producing that picture. For example, family wealth is only part of the group of interrelated family characteristics -- the socioeconomic status of the family. Generally, wealthier families include parents with higher levels of education, who also work in higher status occupations. Given the constellation of attributes, one must consider which is the true contributing factor in bestowing greater benefits to the child. If the overall family wealth were the basis for differences in the educational attainment of children, then a simple redistribution of wealth within society would ameliorate the differences between groups. However, it is more likely that it is the position of the family within the social structure that is making the difference. The effect of the difference in social status is carried through differences in the aspirations of parents, the types of interactions that occur in the family, and the interaction between the family and the school. We explore these relationships further in two ways. First, in the next section, we look at how some common family processes relate to reading performance. Second, in the next chapter, we look at the relative impact of each of these variables in explaining differences in reading proficiency in a way that takes account of this confounding.

13.3.3. Family Environment

While the literature abounds with studies that relate family structure or status to student achievement, more recent emphases have pointed toward family processes as perhaps being more important in explaining differences in students' achievement in that these processes are likely to have a more direct impact, to be more malleable, and possibly to be changed through the intervention of social agencies or the school. Dornbusch and Wood (1989, 67) argue that what is needed is descriptions of those family behaviors associated with better school performance. It is their contention that "less-educated parents, single parents, stepparents, minority parents, and poorer parents could at least be aware of behaviors that more 'advantaged' families, as well as the more successful among the less-advantaged families, use more often with greater success" (also see Clark 1983).

However, studies of family processes have suffered from three main methodological limitations. First, researchers tend to focus on a particular context and may not study the same behavior in other contexts (Steinberg, Elmen, and Mounts 1989). Therefore, they may overlook that which is common across settings (Leichter 1974). Second, these studies have almost exclusively focused on the impact of family processes on young children. Some would argue that parental influence wanes as children age. However, Boger, Richter, and Paolucci (1986), Walberg and Marjoribanks (1976), and Epstein (1985) have shown that the impact of parental encouragement and stimulation may be as great for adolescents as it is for young children. Third, according to Bronfenbrenner (1979), most studies of familial influence focus on the microsystem (i.e., influence on child's personality), or the exosystem (i.e., influence on the social structure within which the child lives), rather than on the mesosystem where the linkage between experiences in the home and school environments occurs.

Like most research in this area, the Reading Literacy Study has some of these limitations. First, it is limited to the questions asked, and due to limitations of time and space, these do not cover the full range of familial processes. Second, a number of the questions have been drawn from the literature associated with what we know makes a difference with young children -- specifically preschoolers. However, the populations studied are no longer that young and, consequently, we may have missed capturing more appropriate types of interactions.

Items placed in the category of family environments were of two kinds -- those things families provide for their children, and the kinds of interactions or activities that occur in the family home:

- Regularity of meals (S4Q7, S9Q7),
- Study facilities (S4Q12, S9Q18),
- Literacy resources (S4Q8, S4Q10, S4Q11; S9Q17, S9Q13, S9Q15),
- TV watching (S4Q9, S9Q14), and

- Homework help and parental interest (S4Q31, S4Q32, S4Q38, S4Q42 S4Q54; S9Q47, S9Q37, S9Q38, S9Q16, S9Q48).

Regularity of Meals (S4Q7, S9Q7). Students were asked to note the regularity with which they ate breakfast, lunch, and dinner. Each of the three questions regarding meals was dichotomized into regular daily meal/less than regular meal. Cross-tabulation of the three variables yielded a single variable capturing whether the student ate zero, one, two, or three regular meals per day. This variable was then simplified again, capturing whether or not the student had three meals a day or less (Table 13-11 and Figure 13-11).

It is clear that for grade 4 students, those reporting that they regularly receive three meals a day have noticeably higher mean achievement across the three scales than other students. For grade 9 this phenomenon persists to a similar order of magnitude, although the finding for this population for the narrative scale shows no significant difference between the two groups.

Within the context of the international component of the IEA Reading Literacy Study, a measure of adequate levels of nutrition as it related to achievement was perceived as important, particularly within developing countries. Within developed countries, such as the United States, it was believed that this variable would not serve as a measure of adequate nutrition, but rather as an indicator of the organization and perhaps stability of family life, the argument being that the more students' meals were provided, the more likely the family was well integrated.

For the most part, measures of this kind are not seen in the U.S. survey literature. There are no corresponding data in NAEP or NELS, and no mention is made in either the *Condition of Education* or the *Digest of Education Statistics*. However, the research literature makes tangential reference to meals at home. Hess and Holloway (1984), for example, report that better school performance was found among students who participate in mealtime conversations and among those who were asked for information by their parents. The issue at the core of their finding is not whether the student had a meal, but rather the qualities of conversation the student has with his/her parents.

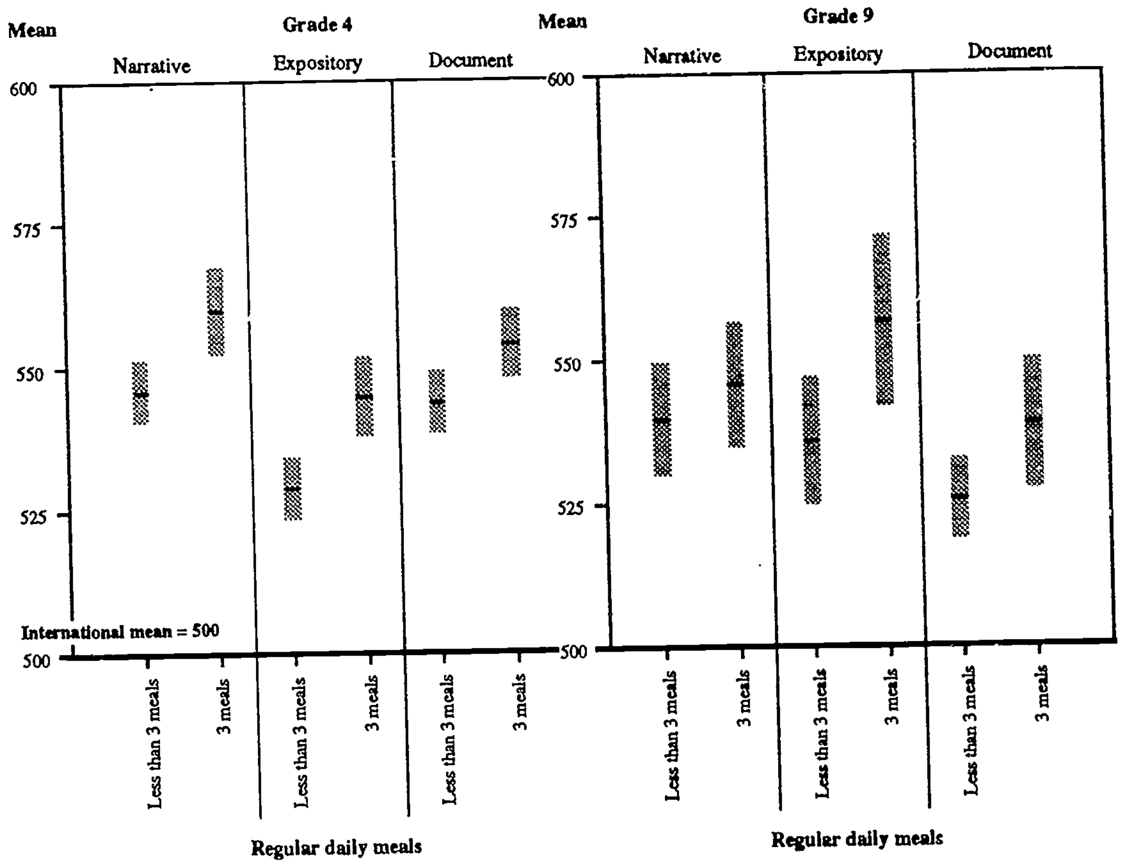
Table 13-11. Mean reading proficiency scores, by regularity of daily meals: Grades 4 and 9

Meals per day	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Less than 3 meals	35	546 (2.6)	529 (2.7)	544 (2.7)	66	540 (4.7)	536 (5.6)	526 (3.5)
3 meals	65	560 (3.7)	545 (3.4)	554 (3.0)	34	546 (5.4)	557 (7.4)	539 (5.6)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-11. Mean reading proficiency scores, with 95 percent confidence intervals, by regularity of daily meals: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Study Facilities (S4Q12, S9Q18). Students at both grade levels were asked whether they had a specific place to study at home (Table 13-12 and Figure 13-12).

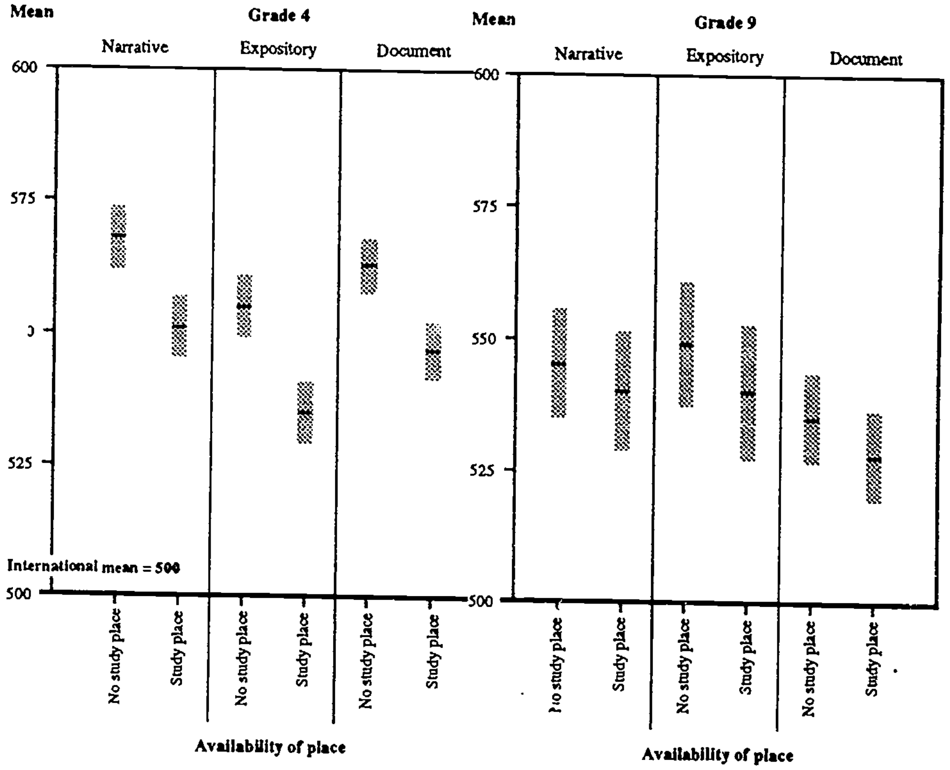
Table 13-12. Mean reading proficiency scores, by availability of study place: Grades 4 and 9

Availability	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
No study place . . .	24	568 (2.9)	555 (2.8)	563 (2.6)	36	545 (5.1)	549 (5.8)	535 (4.1)
Study place	76	551 (2.9)	535 (2.8)	547 (2.6)	64	540 (5.6)	540 (6.4)	528 (4.2)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-12. Mean reading proficiency scores, with 95 percent confidence intervals, by availability of a study place: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

For grade 4, clearly students who report that they have no regular place to study (24 percent of students) have substantially higher mean achievement across the scales than those who do. At grade 9, there is no significant difference between the 36 percent of students with no study place and those with a regular study place for any of the three scales.

The fact that more grade 4 students (76 percent) than grade 9 students (64 percent) indicated that they had a regular place to study suggests that perhaps grade 4 students did not understand the question, or at least did not interpret it in the same way as the older students. This must be considered when interpreting the differences in achievement between those reporting having a special study place and those not having one.

Literacy Resources (S4Q11, S4Q8, S4Q10; S9Q13, S9Q15, S9Q17) The provision of print resources within families as a reading resource, and as a stimulus and encouragement to read, figures prominently in explanations of between-family differences in children's reading capabilities. Books in the home as a measure of this resource has become a standard measure in IEA surveys. In this instance, we have developed a measure of literacy resources within families by considering the nonconsumer-durable items tapped in the family possessions question--possession of computers, dictionaries and the like--but omitted from the family wealth analyses. To these, we have added the books in the home item and the measure of whether the family sees a newspaper regularly.

Preliminary analyses suggested that the newspaper item was not really part of this set and it was eliminated from subsequent analyses. A factor analysis of the relations among the remaining seven items suggested a single construct as responsible for their covariation. We have called this construct "home literacy resources" and consider it to be a measure of the support families provide for reading through the provision of literacy-related materials in the home. Table 13-13 summarizes the results of these analyses. Only one factor was found to have an eigenvalue greater than one, and that factor explains close to 40 percent of the variance.

Table 13-13. Literacy possessions: Factor pattern

Factor 1	Item
0.71	Dictionary
0.61	Calculator
0.60	Atlas
0.57	Encyclopedia
0.46	Typewriter
0.37	Computer
0.36	Number of books at home

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

For both grades the table and figure give a strong indication of an increasing mean score on each scale, with increasing levels of literacy possessions, as characterized by the different population quartiles (Table 13-14 and Figure 13-13). Most striking is that the mean for students in the lowest quartile for literacy possessions is substantially below the means for the other groups, across scales and populations, with these differences being highly statistically significant.

As part of the 1988 NAEP reading assessment, students were asked about the reading materials in their home. At that time, about half of the 4th graders and two-thirds of the 8th and 12th graders reported having their own magazine subscriptions, and these students tended to have higher reading proficiency than their classmates. In addition, the students who reported having more of their own books at home also tended to have higher reading proficiency. At each of the grade levels, the more books students reported having, the higher their average reading proficiency was (Langer et al. 1990).

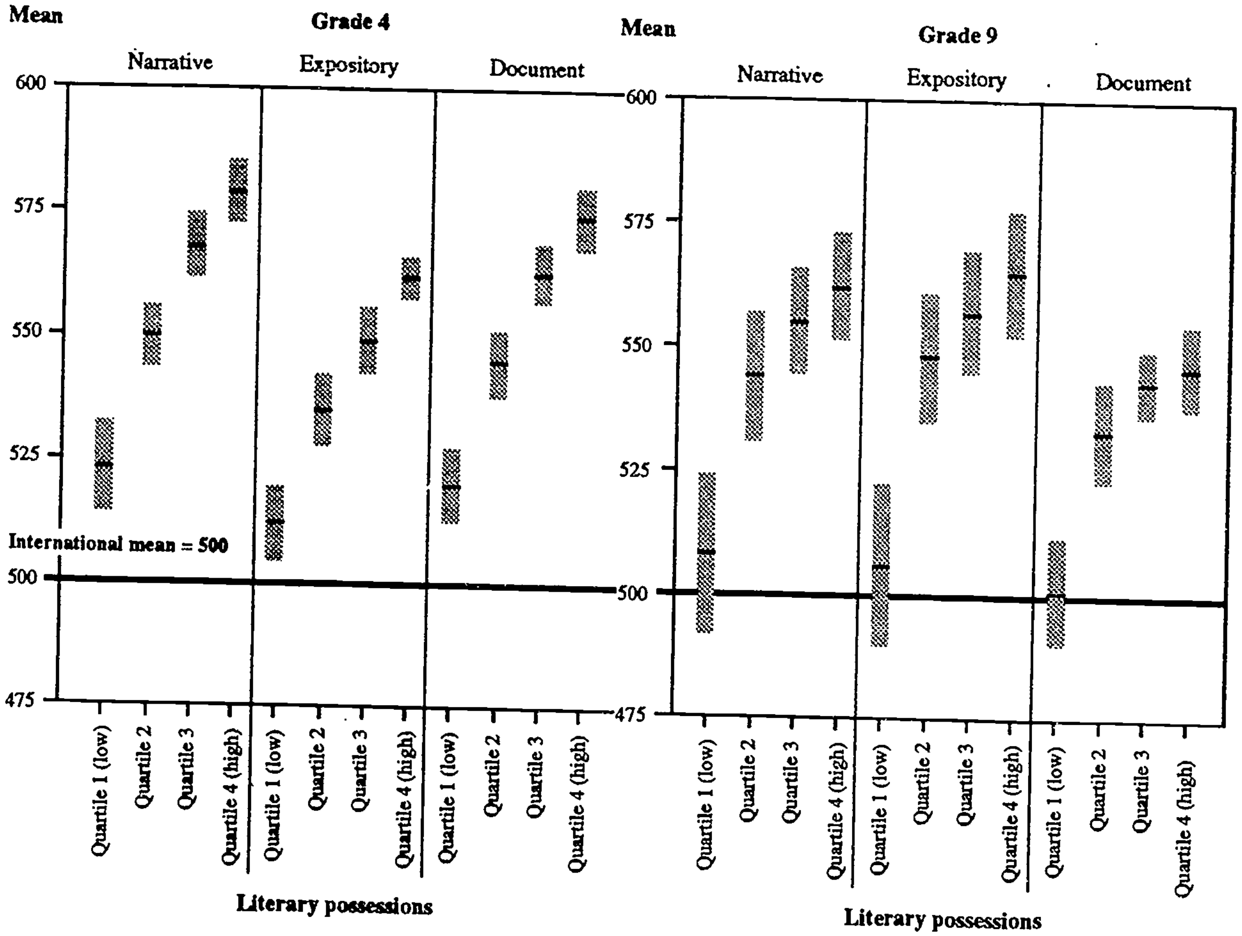
Table 13-14. Mean reading proficiency scores, by literary possessions: Grades 4 and 9

Literary possessions	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Quartile 1 (low) .	24.4	523 (4.6)	512 (3.7)	520 (3.6)	26.0	508 (8.1)	506 (8.2)	501 (5.4)
Quartile 2	25.3	550 (3.1)	535 (3.6)	545 (3.3)	25.2	544 (6.5)	548 (6.6)	533 (5.0)
Quartile 3	25.1	568 (3.3)	549 (3.4)	563 (2.8)	24.9	555 (5.2)	557 (6.1)	543 (3.4)
Quartile 4 (high)	25.3	579 (3.1)	562 (2.1)	574 (3.1)	23.9	562 (5.4)	565 (6.1)	546 (4.2)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-13. Mean reading proficiency scores, with 95 percent confidence intervals, by literary possessions: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Consistent with the NAEP and IEA data, numerous studies have demonstrated that the opportunity to use reading materials at home is positively related to proficiency in reading (Fielding, Wilson, and Anderson 1986; Gopinathan 1978; Spiegal 1981; Teale 1978; Walberg and Tsai 1984; Ingham 1981). This is true for preschoolers whose parents read to them (Morrow 1983; Hess and Holloway 1984), as well as for school-aged children (Sheldon and Carrillo 1952; Polhemus 1955, Napoli 1968).

Ownership of books is one of the more important factors (Briggs and Elkind 1977; Clark 1976; Durkin 1966). Polhemus (1955) reported that children in homes with 100 or more books had a slightly lower percentage of poor readers than did children in homes with fewer books. Napoli (1968) demonstrated that visiting and borrowing books from the public library was more often characteristic of honors students than low achieving students. Further, honor students' parents tended to encourage their children to read more often than low achieving students' parents. Similarly, Lamme and Olmstead (1977) found that the availability of books from different sources was positively related to grade 1 students' reading achievements.

Social class differences were interrelated with the availability of reading materials in the home. Newson and Newson (1977) found that students whose fathers were professionals tended to have more books available for their use and read more books than did students from the homes of nonprofessionals. While the social class difference seems related to book ownership (i.e., family wealth would affect the amount of disposable income available for purchasing books), it is also possible to borrow books from both the public and school libraries. However, Newson and Newson (1977) showed that nearly twice as many children from middle class families visit the public library as children from working class families. It would appear then that when cost is eliminated, working class children are not likely to borrow books even when there is an opportunity to do so.

TV Watching (S4Q9, S9Q14). Students were asked to indicate the number of hours per day spent watching TV or video outside of school hours. The distribution of responses is displayed in Table 13-15 and Figure 13-14.

For grade 9 the tables show clear evidence of a downward trend across the three scales of decreasing mean reading achievement with increasing reported hours of television viewing. One cannot discern any clear break points at which mean proficiency begins to decline dramatically. However, the general trend is unmistakable.

For grade 4 the situation is a little different. There is no significant evidence of any differences between the group that reports viewing 2 or 3 hours a day and the group reporting 1 hour a day or less. There is some evidence that a downward trend in means begins at about the 4-hours-a-day point, and, in particular, students watching 6 or more hours per day have significantly lower mean proficiencies on all three scales than do other students.

Television viewing absorbs a large portion of students' leisure time (Finn 1980). According to NELS:88 data, it in fact dominates the out-of-school time of most eighth graders, accounting for almost 4 times as many hours as they spend on homework and about 10 times the number of hours they spend on voluntary reading (Hafner et al. 1990, 47). The 1992 NAEP data corroborates the NELS findings. American children of all ages watch considerable amounts of television each day.

In general, studies indicate that television viewing is associated with lower school achievement, although Anderson, Wilson and Fielding (1988) have reported that watching up to 3 hours a day may not have a negative impact on reading proficiency. However, they point out that more than 3 hours can have a negative influence on reading achievement. The Reading Literacy Study data for grade 4 are consistent with this finding.

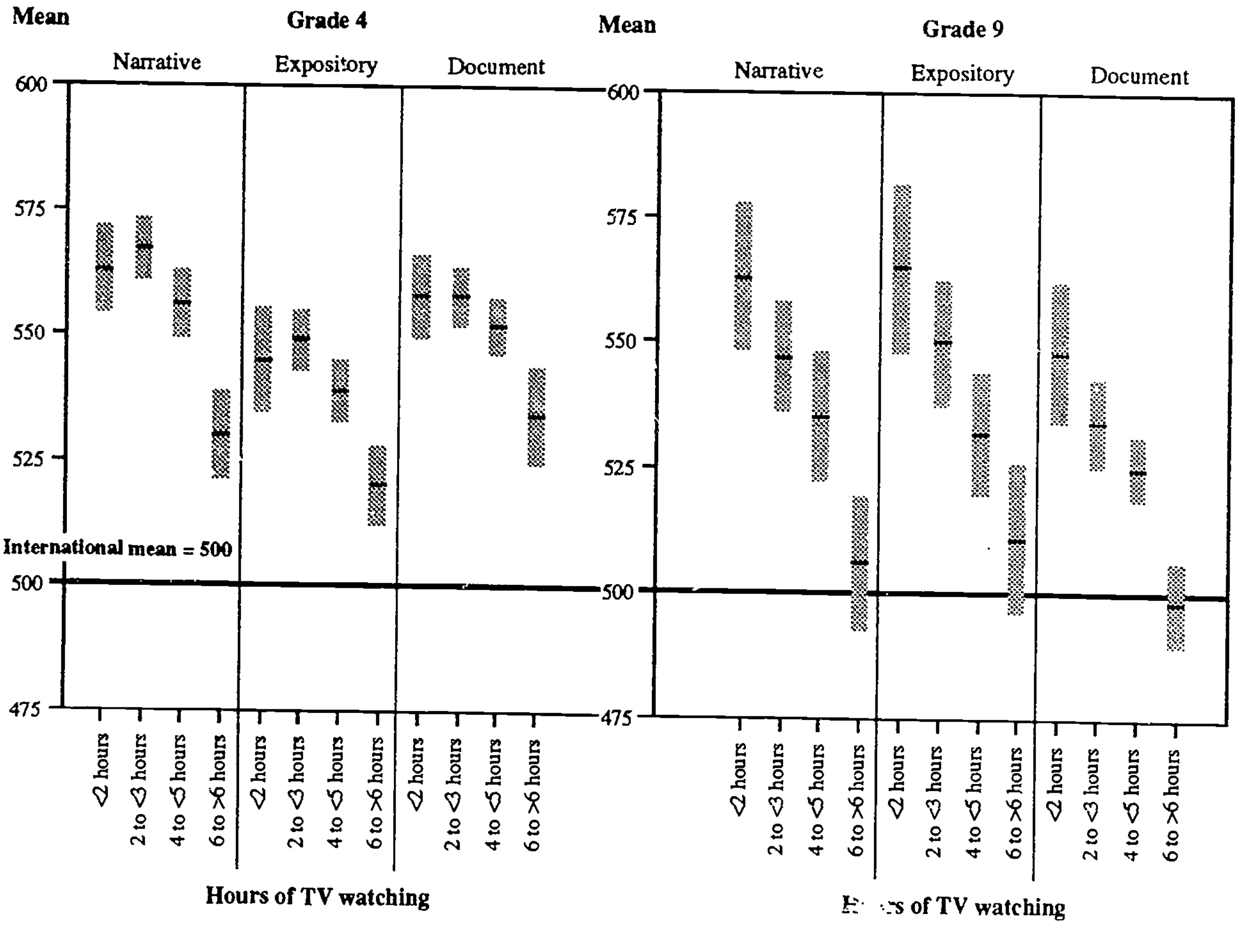
Table 13-15. Mean reading proficiency scores, by hours TV watching outside school: Grades 4 and 9

Hours	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
1 hour or less .	16.3	563 (4.3)	545 (5.1)	558 (4.1)	17.4	563 (7.4)	565 (8.3)	548 (7.0)
2 or 3 hours . .	35.8	567 (3.1)	549 (2.9)	558 (2.9)	46.6	547 (5.5)	550 (6.2)	534 (4.3)
4 or 5 hours . .	25.8	556 (3.4)	539 (3.1)	552 (2.9)	24.1	535 (6.3)	532 (6.1)	525 (3.2)
6 or more hours	22.0	530 (4.4)	520 (3.8)	534 (4.9)	12.0	506 (6.7)	511 (7.3)	498 (4.2)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-14. Mean reading proficiency scores, with 95 percent confidence intervals, by hours of TV watching outside school: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

While one might argue that if children were not watching television, they might read more, Neuman (1986) reports finding almost no relationship between the two; that is, many avid readers spend a great deal of time watching television, while other children neither read nor watch a great deal of television. In the same vein, avid readers of books have been shown to spend less time "inactive" or "lying about" than their peers (Greaney 1980).

Given these findings, one is inclined to conclude that the relationship between television viewing and reading proficiency is more complex than just the amount of time diverted from reading. What may be more important is what the child is watching and how it relates to both the home environment more globally and to the child's interests. For example, Japanese evidence suggests that children's interests are fairly consistent across media such that students interested in academic and technical topics are apt to watch educational television (Ogawa 1986). Beentjes and Van der Voort (1988) and Neuman (1986) report that within the United States, children's television viewing patterns tend to follow the example set by their parents. Children who are avid readers tend to come from families where there are rules regarding television viewing (Whitehead, Capey, and Maddren 1975).

Homework Help and Parental Interest (S4Q31, S4Q32, S4Q38, S4Q42 S4Q54; S9Q47, S9Q37, S9Q38, S9Q16, S9Q48). On these issues, the selection of items thought to tap an underlying latent variable was not flagged in any obvious way by the spatial grouping of the items in the questionnaire. Rather, given that the questionnaire contained items apparently indicative of parental interest in children's reading accomplishments, and because parental interest and encouragement have well established effects on children's learning, the approach taken was to look for such items and examine their latent structure. The items in question concerned whether parents provided help with reading homework, read to the student, encouraged the student to read aloud, and showed an interest in what the student had been reading at school.

A factor analysis of the relations among these items suggested there were two latent variables involved rather than one -- the homework items separate from the parental interest items -- as indicated in Table 13-16, which summarizes the results of these analyses. Factor scores were estimated for two constructs on this basis and the factors were named homework help and parental interest.

Table 13-16. Homework help and parental interest: Factor pattern

Factor loading	Item
Factor 1 - *Homework Help*	
0.77	Homework help, mother
0.56	Homework help, father
Factor 2 - *Parental Interest*	
0.68	Read to in English at home
0.49	Read to in English elsewhere
0.34	Asked what read at school
0.26	Reading aloud at home

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

For each grade, students were classified into quartiles based on their response to the question of how much help they receive with homework (Table 13-17 and Figure 13-15). For grade 4 students, the amount of help with reading homework appears not to be related to literacy achievement. Similarly, at grade 9, there is no evident relationship between the amount of reading homework help and the levels of reading literacy achievement.

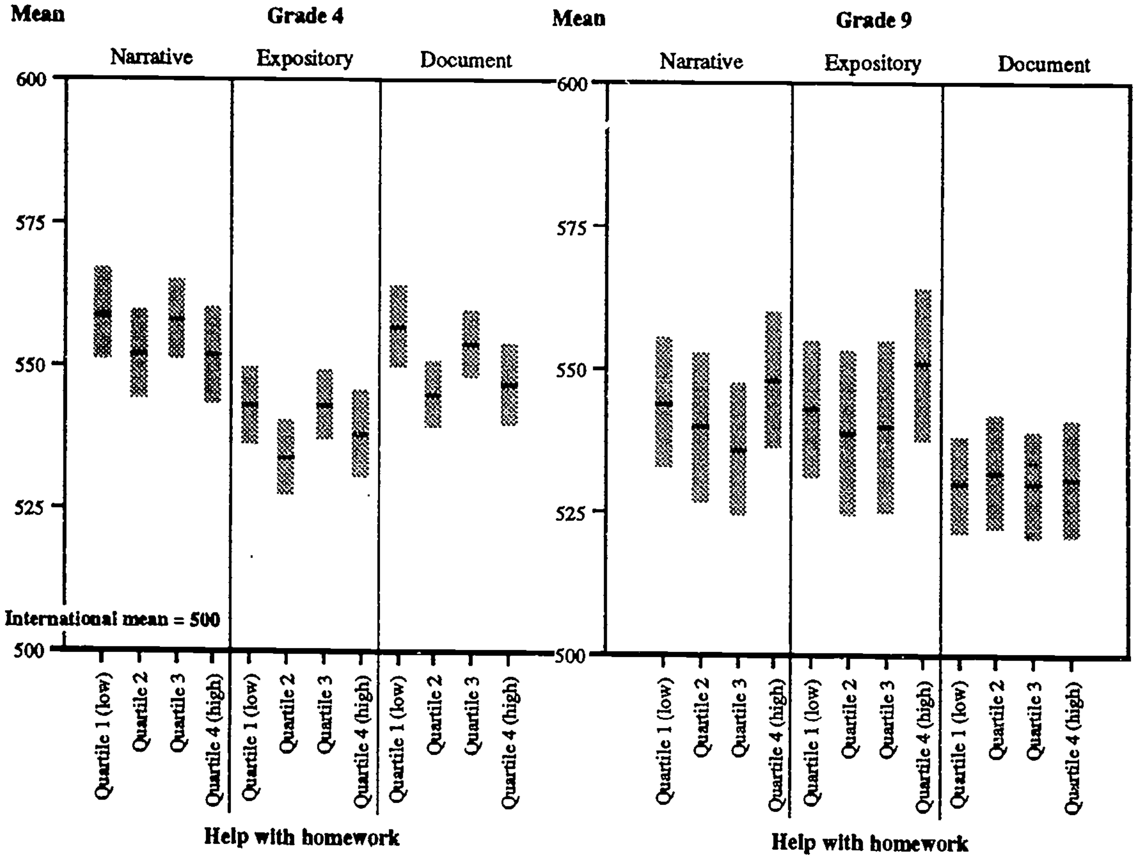
Table 13-17. Mean reading proficiency scores, by homework help: Grades 4 and 9

Homework help	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Quartile 1 (low) .	25.2	559 (4.0)	543 (3.4)	557 (3.5)	25.1	544 (5.6)	543 (5.9)	530 (4.1)
Quartile 2	25.6	552 (3.8)	534 (3.3)	545 (2.8)	25.0	540 (6.4)	539 (7.1)	532 (5.0)
Quartile 3	25.2	558 (3.5)	543 (3.1)	554 (2.9)	24.5	536 (5.8)	540 (7.4)	530 (4.6)
Quartile 4 (high)	24.0	552 (4.2)	538 (3.8)	547 (3.5)	25.5	548 (5.9)	551 (6.6)	531 (5.1)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-15. Mean reading proficiency scores, with 95 percent confidence intervals, by homework help: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In general there is very little in the research literature that specifically addresses parental help with homework. As Dornbusch and Wood (1989) suggest, "parental involvement in homework may lead to school success, but it is also possible that parental involvement in homework follows from a poor school record" (p. 69). Newson and Newson (1977) have shown that parents do help their children in the elementary grades with their homework. Further, Watson, Brown, and Swick (1983) have shown that parents' direct help with children's reading has resulted in reading gains, that this makes a positive difference in children's school progress, and that this effect appears to be irrespective of parents' level of education or income but is directly related to the kind and quality of parental support.

For each grade, students were classified into quartiles based on their response to the questions related to how much parental interest in their educational activities the students felt that they received (Table 13-18 and Figure 13-16). There is evidence in the table, especially for grade 9, of a negative relationship between the amount of interest/interaction and the mean reading achievement score. Specifically, for each grade and scale, the difference in mean reading achievement between the first and fourth quartile groups of parental interest is statistically significant.

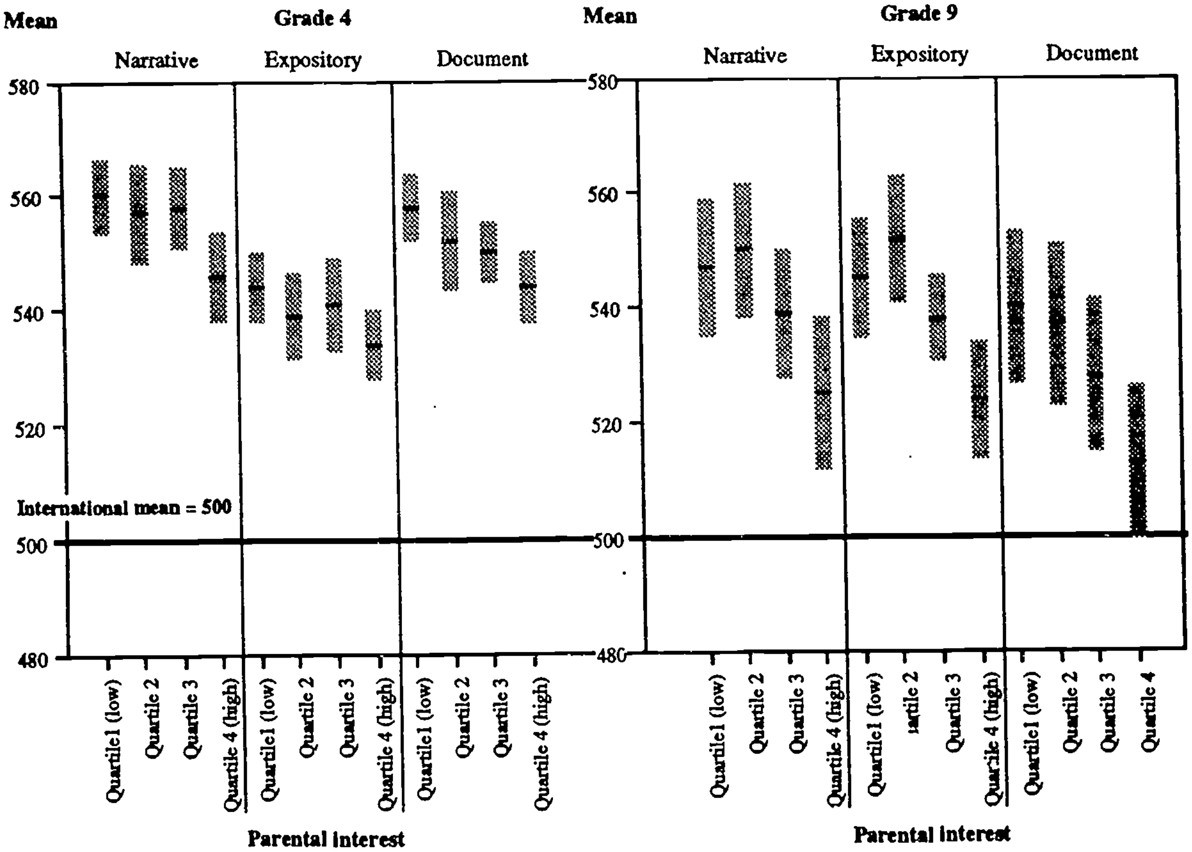
Table 13-18. Mean reading proficiency scores, by parental interest: Grades 4 and 9

Parental interest	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Quartile 1 (low) .	25.4	560 (3.1)	544 (3.1)	558 (2.9)	25.2	547 (6.1)	545 (5.2)	540 (6.7)
Quartile 2	25.0	557 (4.3)	539 (3.7)	552 (4.2)	24.5	550 (5.8)	554 (5.1)	537 (7.0)
Quartile 3	24.6	558 (3.5)	541 (4.0)	550 (2.6)	25.2	539 (5.5)	538 (3.8)	528 (6.6)
Quartile 4 (high)	25.0	546 (3.8)	534 (3.0)	544 (3.1)	25.0	525 (6.6)	524 (5.1)	513 (6.6)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-16. Mean reading proficiency scores, with 95 percent confidence intervals, by parental interest: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

This finding suggests that parents are most likely to get involved with their children's education (as perceived by the students themselves) if those students are lower achievers. An alternative explanation is that, for a given level of parental interest, lower achieving students perceive their parents to show a greater level of interest than do higher achieving students. Perhaps this might occur because their expectations for parental interest are lower.

Although we have labeled this construct "parental interest," what is being measured is very narrowly defined in terms of either reading aloud or discussions of what is read. The literature documents the importance of parents' reading to young children before they attend school (Boehm and Slater 1974; Greany 1986; Morrow 1983; Clark 1976; Durkin 1966). The effect of this parental interaction during the preschool years has been demonstrated to have a long lasting impact. For example, Polhemus (1955) reported that most children from a sixth grade sample who were successful in reading had parents who read to them before they started school, while few of the children who were having difficulty reported having been read to. Studies by Newson and Newson (1977) and Cousert (1978) support the early conclusion that parental reading to children during the preschool years appears to contribute to children's reading achievements throughout the primary and elementary years.

The available NAEP data strongly suggest that students who were read to frequently (at least weekly) when they were children tend to have a higher reading proficiency than those who were not (Langer et al. 1990). However, as Phillips and Bolt (1992) note, there are no studies that explore the relationship between being read to during the middle school years and achievement.

One would expect discussions at home of what students read to become more important in the formation of critical readers. Recent research that looks at literacy as interactive, familial, and social in nature views the home as a crucial element in providing a dynamic and literate environment (Teale 1986). Hynds (1990) argues that reading is a social activity, and that readers develop through participation in "supportive communities of readers."

NAEP results indicate that discussions of reading are not a frequent household practice. The results indicate that 45 percent of 8th and 12th graders talk to someone about reading at least weekly, while 30 percent never do. Those who never talk about reading at home had significantly lower reading proficiency (Langer et al. 1990).

In looking at the literature on parental interest more globally, we find that parental interest might take a variety of forms. These include parental monitoring of homework, discussions of what happens in school, and involvement in the school generally. The literature would support decreasing parental involvement over time. For example, Dornbusch and Wood (1989) note that over time (i.e., as the child matures), one would expect that children develop increased internal motivation for school performance. Further, they argue that while close surveillance might get a student to do an immediate task, it is impossible to maintain such a level of oversight throughout a child's school career. Consequently, too close surveillance of student performance often results in very negative consequences. Work by Lepper, Greene, and Nisbett (1973) and Lepper and Greene (1978) supports this position and further argues that those students who receive external rewards tend to perceive outside forces as controlling their behavior, thereby reducing their own internal motivation.

In considering the importance of the family environment, we note that the research literature pursues a few additional, and extremely important, lines of inquiry that are not included in the Reading Literacy Study. These include communication and decisionmaking within the family, parenting styles, parental expectations, parental and children's reading habits, and participation in cultural and extracurricular activities. An overview of this relevant literature follows.

Communication and Decisionmaking in the Home. Studies by Allen and Chaffee (1977), Chaffee, McLeod, and Atkin (1971), Hess and Holloway (1984), and Epstein (1981, 1989) all indicate the importance of verbal communication within the home. These studies have shown that there are at least two general styles of communication: one that emphasizes conformity to others, and another that emphasizes a concern for the content of ideas and a willingness to consider the diversity of ideas people might have. While the former communication style has been associated with lower grades, lower SES groups, and minority racial and ethnic groups, the latter is associated with higher grades and middle class families where parents are likely to have higher levels of education. It is postulated that children growing up in an environment where active participation in conversation in which diversity of opinion is encouraged are more likely to be well prepared for successful interactions with teachers and peers.

While this literature supports open discussions, it does not necessarily follow that children make decisions regarding school choices, curfews, spending, choice of friends, and the like autonomously. The research by Dornbusch et al. (1985) and Epstein and McPartland (1977) indicates that higher grades, more positive personality development, and better coping skills in school are consistently associated with joint decisionmaking between parents and their children. In the best situations, as the child gets older, parental domination of decisionmaking decreases and youth autonomy increases.

Parenting Styles. The research literature roughly categorizes parenting styles along a continuum from authoritarian to authoritative to permissive (Baumrind 1973; Dornbusch et al. 1987). Authoritarian parents attempt to shape, control, and evaluate the behavior and attitudes of their children in accordance with an absolute set of standards; they emphasize obedience, respect for authority, work, tradition, and the preservation of order; verbal give-and-take is discouraged. Authoritative parents expect mature behavior from the child and have a clear setting of standards that are firmly enforced using commands and sanctions only when necessary; they encourage the child's independence, encourage verbal give-and-take, and recognize the rights of both parents and children. Permissive parents are tolerant of children's impulses, use little punishment, make few demands for mature behavior, and allow considerable self-regulation by the child.

Studies by Baumrind (1983) and Dornbusch and Wood (1989) show that authoritarian and permissive parenting are negatively associated with grades, while children of authoritative parents tend to have higher grades and are more socially responsible and independent. Steinberg et al. (1988) and Pulkkinen (1982) suggest that parental consistency will also play a major role in promoting school success. Findings of this nature are supported by the work of Clark (1983), Steinberg, Elmen, and Mounts (1989), Hill (1980), and Maccoby and Martin (1983).

Parental Expectations. Among the studies of parental expectations, three that typify the work in this area are Rankin (1967), Newson and Newson (1977), and Marjoribanks (1983). In general these studies reveal that parental expectations are positively related to children's reading achievement. Rankin's (1967) findings indicate that there are differences in the degree and kinds of parental interest between the parents of high and low achievers. The parents of high achievers were more likely to discuss homework assignments with their children, to attempt to find out what problems their children had in school and help to solve these problems, to hold the expectation that their children would go on to college, and to require that their children achieve high marks.

Newson and Newson (1977) found significant social class differences in parental expectations for their children, with middle and upper class parents having expectations and working class parents having hopes with regard to their children's higher education and future jobs. Upper class mothers tended to assume that their children would have the ability to do certain jobs but may not want to, whereas lower class mothers were concerned about their children's abilities.

Marjoribanks' (1983) study of parental aspirations tended to confirm that there were social class differences related to parent-child interactions and parents' concern for their child's independence. The study indicated that in the lower classes children are relatively unaffected by their parents' aspirations, while in the upper classes parents appear to exert greater influence.

Parental and Children's Reading Habits. As Greaney (1986) has noted, having adult reading models in the home influences children's desire to read. This view is supported by the work of Clark (1976), Gibson and Levin (1975), Pickering (1977), Roberson (1970), Krus and Rubin (1974), Kontos (1986), Durkin (1966), and Freshour (1972). In general, the position these theorists have espoused is that the parents who model good reading behaviors and demonstrate not only the value but also the fun inherent in reading tend to have children who become avid readers. Based on their studies they have recommended that children have opportunities to observe their parents reading, that both children and parents discuss what they have read, and that parents create a special family time for reading. The research of Cousert (1978), Dix (1976), and Lee (1984) has established that in addition to developing positive reading habits, there is a relationship between parental reading habits and children's reading performance. However, the only authors listed above who included students at equivalent ages as those in the Reading Literacy Study were Dix (1976), who looked at 213 children in grades 1 through 6, and Lee (1984), who studied 54 students in grades 2, 4, and 6.

While the research shows that there is a relationship between voluntary reading and reading achievement (Long and Henderson 1973), not all students who demonstrate skill in reading are necessarily voluntary readers (Morrow 1983). Although most curriculum mission statements are likely to include a statement of the desire to develop lifelong reading habits, numerous studies have demonstrated that a substantial number of students do not engage in reading for their own personal interest and enjoyment. Instead, most only read that which has been assigned. Greaney (1980) established that few fifth grade students devote their leisure time to reading. While 22 percent of the students in the sample reported no voluntary reading, of the 78 percent who did read voluntarily, only 5.4 percent of their time was spent in this way. In fact, out of the nine major leisure categories identified in the study, reading was ranked seventh.

Two factors seem to influence students' voluntary reading. The first is the level of literacy in the home environment (Hansen 1969). This trait was related to such indicators as the availability of books, the amount of reading done with the child, the amount of guidance provided, and the extent to which parents model good reading habits. The second factor influencing voluntary reading is the emphasis of the reading program at school. Morrow (1985) contended that school programs emphasizing skills failed to promote voluntary reading.

For the most part, children who came from families where parents had a college education or a graduate degree were more likely to have an interest in reading (Morrow 1983). Similarly, the higher the level of social class status of the family, the greater the amount of time students were likely to spend reading (Long and Henderson 1973).

The few studies in this area studying the relationship between reading achievement and participation in cultural and extracurricular activities are characterized by somewhat similar findings. Students who are achieving above grade level tend to be involved in more cultural and extracurricular activities than those who achieve at or below grade level (Berry et al. 1971). In contrast, Olson (1984) found that the number of sports activities in which students participated was negatively related to the prediction of reading achievement for all students, while the number of club activities was positively correlated with girls' reading achievement. However, social class and family size also appear to be significantly related to participation in cultural activities. Newson and Newson (1977) found that about twice as many middle class families as working class families took part in a variety of cultural activities. Similarly, children from families with three or fewer children were more likely to take part in a variety of cultural activities than children from larger families.

These studies suggest that exposure to a wide variety of activities tends to provide children with a background of knowledge, skills, and experiences that are likely to be useful to their educational development. However, factors of family wealth and size appear to have a direct impact upon the likelihood of participation, thereby biasing the probability of a broad cultural exposure to only those who have access. In addition, these findings are based on studies with very small sample sizes.

13.3.4. Teacher Attributes

Analysis of the research on effective schools and effective teaching shows that within-school differences are much greater than between-school differences. Although it would seem that teacher attributes are highly important in accounting for the within-school differences observed, numerous studies have shown that few characteristics of teachers or schools are related to the average achievement of students within a school (Hanushek 1986). The models developed in the Reading Literacy Study include a number of teacher-related variables that have previously shown some relationship to academic achievement, although these relationships may have been weak or insignificant.

One of the difficulties in examining data on the relationship between teacher characteristics and student performance is that the data are only modestly stable across years (Brophy and Good 1986). Stability coefficients are significant, but only between .2 and .4 in the studies Brophy and Good describe. In addition, there may be systematic bias in the way that students are assigned to classes and teachers. Such practices may account for the "considerable" within-school variation among classrooms reported, although the few studies that permit analyses relating individual teacher characteristics to students show that none of the measured teacher characteristics show a strong, consistent relationship to student achievement (Smith and O'Day 1988, 7).

The teacher attributes considered in the conceptual structure of the Reading Literacy Study that may contribute to achievement are

- Gender,
- Race/ethnicity,
- Formal education,
- Training, and
- Teaching experience.

Teacher's Gender. The Teacher Questionnaire for both grades included the question "Are you a male or female?" Table 13-19 displays the relevant data.

The Reading Literacy Study included 42 male (14 percent) and 261 female (86 percent) fourth grade teachers, and 49 male (30 percent) and 116 female (70 percent) ninth grade teachers. This distribution is only slightly different from the 1990-91 School and Staffing Survey (SASS), which reported that 17 percent of public elementary school teachers were male (as were 13 percent of private elementary school teachers), and 46 percent of public secondary school teachers were male (as were 44 percent of private secondary school teachers; Choy et al. 1993). The differences in distribution between the IEA and SASS data may be attributed to differences in the sample designs, as well as to sampling error. While SASS includes a nationally representative sample of teachers, the IEA sample was representative of classes, not teachers, and as such may not provide as accurate a picture of the teaching population as the SASS sample. Neither NELS:88 nor 1992 NAEP reported data on teacher gender to use as comparisons for the current study.

Table 13-19. Class mean reading proficiency scores, by teacher gender: Grades 4 and 9

Gender	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Male	13.9	551 (8.8)	531 (7.7)	546 (7.0)	29.7	542 (10.4)	546 (11.9)	531 (8.1)
Female	86.1	558 (3.6)	544 (3.5)	554 (2.8)	70.3	535 (5.8)	535 (7.1)	525 (4.4)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Although there appears to be an association between teacher gender and achievement that changes between grades 4 and 9 (with children in classes with female teachers in grade 4 having higher mean scores and students in classes with male teachers in grade 9 having higher mean scores), these differences are not statistically significant.

There is little research including analysis of student achievement based on teacher's gender. Smith and O'Day (1988) include teacher gender as one of the characteristics of schools that is neutral with respect to student achievement. They point out that few studies enable analyses where student performance can be linked to individual teachers, so that the direct effect of teacher characteristics can be measured (p.7).

Teacher's Race/Ethnicity. The teacher's race/ethnicity was determined by self-identification on the Teacher Questionnaire in response to the question, "What is your ethnicity/race?" The data from the Reading Literacy Study indicate that the mean reading proficiencies of students of black teachers are significantly below those of students of white teachers. It is impossible to draw any conclusions from the data on other minority teachers because of the limited sample sizes (Table 13-20). Before drawing any inferences with regard to the relative quality of black and white teachers based on these data, it is important to recognize that the data do not take into account class assignment or teacher training variables. Both are likely to play an important role in this relationship.

Data from other sources indicate that there are very few minority teachers. For example, according to SASS, the teaching population in 1990-91 was overwhelmingly white (87 percent; approximately 8 percent were black, 3 percent were Hispanic, less than 1 percent were American Indians, and only slightly more than 1 percent were Asian or Pacific Islanders; Choy et al. 1993, 34). Note that whites made up 87 percent of the teaching staff of public schools and 92 percent of the teaching staff of private schools. NELS:88 data show that white teachers made up the majority of the teaching faculty in public (88.4 percent), private (93.5 percent), and private, nonreligious schools (96.1 percent). In all categories the percentage of minority faculty was lower than the percentage of minority students in the schools. Consequently, very few students are likely to have an opportunity to see minority teachers in this position as a positive model.

Table 13-20. Class mean reading proficiency scores, by teacher's race/ethnicity: Grades 4 and 9

Race/ethnicity	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Asian	0.2	466 (-)	457 (-)	463 (-)	0.0	0 (-)	0 (-)	0 (-)
American Indian	0.8	495 (31.3)	493 (18.8)	520 (18.8)	0.0	0 (-)	0 (-)	0 (-)
Hispanic	1.2	544 (19.5)	523 (23.2)	536 (23.2)	1.4	502 (8.9)	487 (10.8)	496 (5.0)
White	90.8	562 (3.0)	547 (2.5)	557 (2.5)	93.4	542 (4.9)	543 (5.9)	530 (3.8)
Black	6.9	500 (6.1)	488 (4.0)	499 (4.0)	5.2	462 (11.9)	469 (18.0)	471 (11.5)

KEY: (-) = Sample size is too small to compute standard error.

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

While some argue that minority children should have teachers within their own racial/ethnic group as social role models, there may also be a specific benefit in having such teachers as literacy models. Some groups, particularly Native Americans, feel that teachers outside their own group lack the cultural and linguistic knowledge necessary to effectively teach their children, whether they attend schools that are public, private, or operated by the Bureau of Indian Affairs (Cahape and Howley 1992).

Similarly, many argue that the need for black educators is particularly acute (Graham 1987). Among black teachers, there are relatively few members of the profession who are younger than 30 years; 8 percent of black teachers were under 30 years old compared to 14 percent of white, non-Hispanic teachers, and 15 percent of Hispanic and other racial/ethnic groups (NCES 1993). King (1993) discusses the need for African American teachers and other "teachers of color." She summarizes research indicating that the race/ethnicity of the teacher may be far more important for minority children than for white children. Black teachers are more important as role models and mentors for black children than are white teachers for white, majority children with more opportunities to observe role models like themselves. King also suggests that there are differences in teaching philosophy and pedagogy among African American teachers that may benefit black children.

Teachers' Training, Formal Education, and Years of Experience. The variables associated with teachers' training, formal education, and experience all relate to the knowledge and skill that teachers have about teaching and learning, the content of the curriculum they teach, and how these relate to reading achievement outcomes for students. Data from earlier studies indicate that the amount of formal training completed by a teacher is not consistently related to student achievement; similarly, easily measurable differences among schools (such as class size variation, teacher preservice training, teacher experience) have little consistent relationship to student achievement (Purkey and Smith 1983). These variables are like many other individual teaching variables in that their statistical correlations with measures of student achievement are low (Davis and Thomas 1989). Likewise, the Reading Literacy Study data also show little effect for these variables (Tables 13-21, 13-22, and 13-23). However, even weak relationships and small effects, taken together, can add up to significant achievement differences (Davis and Thomas 1989; Gage 1978, 1985). It is notable that teachers with 8 or fewer years of experience have significantly lower class mean proficiencies at grade 4 for both the narrative and document scales than do those teachers with 9 to 16 years experience.

Stringfield and Teddie (1987) find that principals in effective schools select new teachers for their staff "with great care [looking] for 'spark' or 'energy' [rather than] years of teaching experience or advanced degrees." This position might indicate that effective principals recognize that factors other than education, training, and experience are more important to effective teaching. This view might imply that giving teachers knowledge or information about teaching -- whether through preservice or inservice programs -- is likely to have little effect, unless the individuals care about the outcome of their teaching for their students.

Table 13-21. Class mean reading proficiency scores, by teacher's education: Grades 4 and 9

Education	Grade 4							Grade 9						
	Percent	Narrative		Expository		Document		Percent	Narrative		Expository		Document	
Less than BA . .	1.1	507	(43.4)	509	(36.0)	542	(42.2)	0.0	-	-	-	-	-	-
BA	21.8	547	(7.9)	533	(10.1)	544	(6.4)	15.9	533	(9.1)	532	(10.6)	523	(7.0)
Higher than BA .	77.1	561	(3.1)	545	(3.0)	555	(3.1)	84.1	538	(5.8)	539	(6.6)	527	(4.4)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-22. Class mean reading proficiency scores, by level of teacher training: Grades 4 and 9

Teacher training	Grade 4							Grade 9						
	Percent	Narrative		Expository		Document		Percent	Narrative		Expository		Document	
Quartile 1 (low) .	28.7	549	(6.2)	534	(5.7)	542	(4.6)	24.2	533	(11.9)	534	(14.4)	524	(8.4)
Quartile 2	23.1	558	(6.4)	540	(5.5)	553	(5.4)	21.4	533	(11.0)	534	(12.1)	521	(8.4)
Quartile 3	25.8	564	(4.5)	553	(5.2)	565	(2.9)	25.2	542	(7.8)	543	(8.2)	533	(6.3)
Quartile 4 (high)	22.3	558	(6.9)	542	(6.9)	552	(6.7)	29.2	539	(9.3)	540	(11.1)	528	(7.3)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-23. Class mean reading proficiency scores, by teacher's years of experience: Grades 4 and 9

Years of experience	Grade 4							Grade 9						
	Percent	Narrative		Expository		Document		Percent	Narrative		Expository		Document	
8 or fewer	34.5	546	(6.2)	532	(5.5)	540	(4.3)	33.1	530	(9.0)	529	(10.5)	522	(7.2)
9 to 16	23.1	563	(3.9)	544	(4.7)	560	(5.5)	30.2	545	(9.8)	549	(10.8)	531	(6.8)
17 to 24	30.8	556	(4.6)	544	(3.8)	553	(4.7)	24.4	540	(12.9)	535	(14.5)	529	(9.8)
25 to 32	7.6	569	(9.7)	553	(9.1)	568	(9.9)	10.5	532	(12.2)	541	(13.0)	526	(10.0)
33 or more	4.0	580	(23.4)	583	(44.4)	580	(29.9)	1.7	522	(31.0)	550	(36.0)	520	(12.7)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In reviewing effective schools research, Edmonds (1981) attributes much of the characteristics of effective schools to teachers, but recognizes that it is not really known how to help teachers develop the desired characteristics through any structured means. The lack of effect for formal education may be due to the nature of preservice teacher training. Experienced teachers seem to view their formal preservice training as too theoretical, not sufficiently practical, and lacking in opportunities for additional professional input (Dreeben 1970; Lortie 1975; Rosenholtz 1985). Inservice education, however, may serve teachers' needs better. Effective schools research suggests a need for opportunities for continuing education for teachers. Effective schools have teachers who continually expand their repertoire of teaching methods and use innovative curricula and teaching methods (Davis and Thomas 1989).

Years of experience is an interesting variable to consider at either the individual or aggregate level. With regard to the former, researchers have found a curvilinear relationship between teacher experience and student achievement (Katzman 1971; McLaughlin and Marsh 1978; Murnane 1975), with effectiveness beginning to decline after 5 years. With regard to the aggregate level, effective schools have been found to have a lower rate of teacher turnover and, consequently, fewer new teachers than less effective schools (Brookover and Lezzotte 1979; Sizemore et al. 1983; Venezky and Winfield 1979). This effect may be related to other features of effective schools, such as the degree of collaboration between teachers in the school. Rosenholtz (1985) reports that the amount of task-related discussion between teachers is associated with more effective schools. However, in isolated (i.e., noncollaborative) settings, there was a decline in task-related discussions with increasing years of teaching experience.

National studies indicate that, for most teachers, initial preparation for teaching culminates in an undergraduate degree in education for most elementary school teachers and in the academic discipline that they plan to teach for most secondary teachers. The content of teacher preparation programs are often tied to state certification requirements for teachers, which differ for the elementary and secondary levels, as well as for such areas as special education or reading specialist. At the elementary school level, private schools have more teachers with degrees *not* in education than public schools, perhaps because their teachers are not held to the same certification requirements as public school teachers (Choy et al. 1993).

Most teachers have some postgraduate training, either through college programs leading toward advanced degrees or inservice programs. Approximately 50 percent of all teachers obtain an advanced degree, most frequently in education, with public school teachers more likely than their private school counterparts to hold an advanced degree. The 1987-88 School and Staffing Survey (SASS) found that 33.8 percent of all teachers had engaged in inservice training or college courses within the past 2 years. (Only substantial programs consisting of at least 30 hours of instruction were included in this survey, although many inservice programs are offered to teachers that require much less than 30 hours.) In a study by Lyons (1991) comparing the effectiveness of a 2-week inservice program with a year-long program in developing teachers' skills in delivering a complex reading program, Lyons finds that the year-long program was substantially more effective than the brief program. Perhaps research on the effects of teachers' backgrounds should focus more on the type of courses and/or programs taken.

13.3.5. Teacher-Created Environment—Reading Instruction

Although the instructional variables included in this study have no demonstrable effect on achievement, the description of what goes on within a classroom is a very important aspect of the study. That these variables have little effect is not surprising given the cross-sectional design of the study. It would be unreasonable to expect that instruction during less than 1 year might alter the cumulative effects of all the years of schooling that preceded it, or the cumulative effects of the family. Despite this limitation, the description of instruction in classrooms provides insight into what teachers believe, do, have children do, and test. This information is valuable in its own right and provides information important to policy decisions related to reforming American education.

In Section 13.3.5.1, we begin looking at the data by focusing on items. We have selected 3 of the 190 items to illustrate the type of available information. The 3 items selected closely correspond to items included in NAEP and demonstrate how similar the data are. But organizing 190 items into usable chunks when there is so much information to balance is difficult. We, therefore, devised reasonable rules for data reduction. In Section 13.3.5.2., we look carefully at the procedures used in data reduction. We describe how we grouped items within four question blocks together to form factors so that we might develop a macroperspective about instruction. We then describe instruction in terms of how these factors fit together empirically and as they relate to general theories of reading and learning. A more detailed analysis of this data and the literature is included in the separate report *Methodological Issues in Comparative Education Studies.*

13.3.5.1. Item-Level Analysis

For the purposes of providing examples of instructional data available in the Reading Literacy Study, we have chosen three items that correspond closely to items included in NAEP. These items ask about phonics instruction, the use of writing in response to reading, and the provision of time to read silently in class. The criteria for the selection of these items was based solely on the correspondence to NAEP.

The items were drawn from question T4Q30, which asked teachers to state how often their students were typically involved in specified reading activities. The activities ranged from learning letter sound relationships to making generalizations and inferences to reading in other subject areas.

Frequency of Phonics Instruction. As concisely stated by Lundberg and Linnakyla (1992, 2), phonics most often refers to a stage-wise, objective-based strategy where specific decoding skills are taught with the aim of full mastery within the first two school years. The purpose of this instruction is to make certain that children understand the fundamental nature of the alphabetic principle and that they acquire ready familiarity with frequent words and with spelling patterns and their mapping to sounds (also see Adams 1990; Chall 1967; and Anderson et al. 1985).

Within the context of the Reading Literacy Study Teacher Questionnaire, teachers were asked not about their adherence to the more global theoretic positions most often associated with phonics instruction, but rather how frequently they included instruction in letter sound relationships in their classroom activities (Table 13-24).

Table 13-24. Class mean reading proficiency scores, by frequency of phonics instruction: Grade 4

Frequency	Percent	Narrative	Expository	Document
Almost never	22.5	564 (7.0)	550 (6.7)	562 (5.9)
About once or twice a month	15.5	556 (6.4)	538 (5.8)	553 (6.8)
About once or twice a week	35.0	557 (5.4)	542 (4.4)	551 (3.4)
Almost every day	27.0	550 (6.4)	535 (5.3)	545 (4.4)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Although the data suggest that for each of the three scales, the more frequently the teacher reports the use of phonics instruction, the lower the mean score of students, there are no statistically significant differences among the means of groups of students whose teachers reported using this practice to varying extents for any scale. The level of statistical significance must be controlled because many comparisons are involved (six for each of the three scales), with the result that even the apparent differences between the extreme groups in each case are not significant.

This contrasts with the results of the 1992 NAEP reading assessment, in which fourth grade students of teachers who reported a heavy emphasis on the use of phonics had considerably lower mean reading achievement than those whose teachers reported a moderate emphasis. This second group in turn had slightly lower mean achievement than students of teachers who reported little or no emphasis on phonics.

While the Reading Literacy Study item asks about a specific subset of activities, the question in NAEP places phonics in a context where it represents an entire instructional approach. In contrast to the other instructional approaches (literature-based reading, integration of reading and writing, and whole language) considered in NAEP, phonics alone would tend to be limited only to beginning reading instruction in its intent. It is the only approach included in that list that focuses solely on the decoding aspect of reading instruction.[14] Therefore, we might expect that the NAEP data would show larger effects associated with this instructional stance than the IEA data. In fact, the two studies are quite consistent in their findings that greater use of phonics instruction is associated with lower student achievement.

In interpreting the data from both the Reading Literacy Study and NAEP, we note that where fourth grade teachers report high levels of phonics instruction, students tend to have lower achievement scores. Given that this activity or instructional approach is recommended for beginning or delayed readers (Stahl 1992), and that it should be suspended after the second grade if students demonstrate adequate abilities (Anderson and Pearson 1984), it seems reasonable to surmise that the students who are receiving phonics instruction have entered the grade with lower reading abilities. Consequently, we can draw no conclusions about the efficacy of phonics instruction based on these data.

Writing in Response to Reading. When reading is considered in the larger context of language usage or communication, the interrelationships between speaking and listening and reading and writing become more prominent. There has been a growing emphasis on more closely relating reading and writing because of the natural ways in which they complement each other and call upon related cognitive capacities (Loban 1963; Durkin 1988; Moffett and Wagner 1983; Lewin 1992; Farr 1990; Reid 1990; Clay 1985). Strategies for having students respond in ways that more closely emulate what people more generally do when reading are multiplying. Consequently, more and more children are being asked to write summaries, to keep a personal reading journal, or to write to a friend about a book and their reactions to it (McGinley and Madigan 1990).

The group of students whose teachers report that students almost never write in response to something that they have read is too small for drawing meaningful conclusions. There are no significant differences in mean proficiency among students with teachers in the other three groups, who report with varying frequency that they use this practice (Table 13-25). Similarly, the 1992 NAEP reading assessment showed that there is no significant difference in mean reading achievement among fourth grade students whose teachers report that they write almost every day in response to something they have read, write at least once a week, or write less than weekly.

---

[14] This does not mean that advocates of phonics instruction believe that once children have learned phonics, reading instruction is complete. Rather, they then advocate continued reading instruction to facilitate comprehension.

Table 13-25. Class mean reading proficiency scores, by teacher-reported frequency of written responses to readings: Grade 4

Frequency	Percent	Narrative		Expository		Document	
Almost never	0.1	497	(34.2)	500	(27.3)	521	(20.4)
About once or twice a month	18.2	559	(7.5)	543	(7.0)	559	(7.2)
About once or twice a week	48.8	558	(4.2)	543	(5.7)	551	(4.0)
About once a day	32.4	557	(5.9)	541	(5.2)	551	(5.4)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Can we conclude that writing in response to reading has no effect on reading achievement? Tierney and Shanahan (1991), in their extensive review of the literature related to the integration of reading and writing, would argue that the opposite is true. They point out that, despite methodological advances in exploring this area, "the research on reading-writing relationships should be viewed as still in its infancy." Consequently, those instructional strategies that have been implemented to date might, in fact, be misguided or misused. In addition, as compared to other instructional strategies, teachers at this grade level may not be comfortable with activities of the kind that emphasize work in progress and multiplicities of right answers. Further, it is striking that an overwhelming majority of the research in this area has focused on students of high school or college age. Therefore, one would wonder about the ability of fourth graders to successfully use similar approaches. Alternatively, we have no measure of whether beginning this type of instruction at this age results in higher achievement in successive grades. Once again, there is little indication from these data of an appropriate policy decision.

Silent Reading in Class. Numerous researchers have found that how much a child reads is highly associated with various measures of reading achievement (Anderson, Wilson, and Fielding 1988; Greaney 1980; Greany and Hegarty 1984; Kirsch and Guthrie 1984; Krashen 1988; Heyns 1978). However, Thurlow et al. (1984) have pointed out that perhaps too little sustained reading may be occurring in school as a part of reading instruction. In the Reading Literacy Study teachers were asked how frequently their students silently read in class (Table 13-26).

Table 13-26. Class mean reading proficiency scores, by frequency of silent reading in class: Grade 4

Frequency	Percent	Narrative		Expository		Document	
Almost never	0.1	484	(-)	469	(-)	498	(-)
About once or twice a month	0.1	561	(14.5)	559	(14.3)	556	(15.9)
About once or twice a week	12.8	547	(12.7)	535	(11.0)	546	(9.7)
Almost every day	85.4	560	(3.2)	544	(3.4)	554	(3.1)

KEY: (-) Sample size is too small to compute standard error.

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The groups of students whose teachers reported that they read silently almost never or once or twice a month are too small for drawing meaningful conclusions. There are no significant differences across the scales in the mean proficiency of the students of teachers who report silent reading once or twice a week and those with teachers who report silent reading almost every day.

In comparison, the 1992 NAEP reading assessment showed the following results. Grade 4 students whose teachers reported that their students read silently almost every day had somewhat higher mean achievement than did students of teachers who reported that their students read silently at least once a week. The group of students whose teachers reported that they read silently less than weekly was very small.

There are two things that stand out when considering these findings. First, in contrast to, or perhaps as a consequence of, the warning that not enough sustained reading was going on in classrooms during the early 1980s, we note that it is a fairly common practice as measured by either the IEA survey or NAEP in the 1990s. About 98 percent of the teachers in both surveys reported silent reading in class at least once a week. Second, there does not seem to be much difference in performance associated with teacher reports of whether students read at least once or twice a week or almost every day.

The real question, however, is whether this use of instructional time has improved reading achievement. Pearson and Fielding (1991) were surprised that methods designed to increase the amount of uninterrupted reading children do in class had met with limited success. They reported on three methods -- book floods, use of classroom library, and sustained silent reading. The findings related to book floods indicated that these were most successful in settings where few books were available prior to the intervention (Elley and Mangubhai 1983; Ingham 1982; Holdaway 1979). The work of Morrow and Weinstein (1986) indicated that the voluntary free-time reading of second graders could be increased if the classroom library was well stocked and if there were related book enjoyment activities. However, this did not seem to transfer to increased out-of-school reading. Sustained silent reading seemed most successful when it was accompanied by peer and teacher interaction about books (McCracken 1971; Cline and Kretke 1980; Collins 1980; Manning and Manning 1984). Once again, there is no clear indication of an appropriate policy direction.

Although we could continue to review each of the remaining 187 instructional variables in the same way, we are not certain how it would help instructional practice. How would a curriculum specialist use the information, particularly in light of the fact that these data seem to contradict the research literature reported in the journals? How would a policymaker choose one instructional activity over another?

A further consideration is what the teacher would do in response to a mandate to use a particular instructional activity. When presented with a particular activity as a model of good instruction, the teacher is likely to evaluate it against what he or she is already doing and against the theory of reading he or she has espoused. Although teachers use most activities listed in the questionnaire, how they put them together and what aspect they emphasize is dependent upon their implicit theory of reading.

The kinds of questions one would want to pose are more concerned with the mix of instructional practices -- the combinations that work. Therefore, it might be more useful to step back and to group these items into more meaningful units for analysis. These units should, in principle, correspond to theories of reading.

13.3.5.2. Turning Groups of Items Related to Instruction Into Meaningful Constructs

While in principle the 190 items related to instruction might be grouped on the basis of a theoretical stance, the questionnaire was not explicitly designed to reflect theoretic stances. Instead, blocks of items that had the same response scales were grouped together as a single question because they had a common theme that tied the block of items to the literature on reading. Therefore, to establish both a theoretical (i.e., based on reading theory) and an empirical basis for our groupings, we engaged in exploratory factor analyses to examine the latent structure of responses to these items. We focused on four of these omnibus questions: what teachers believe about reading instruction; what they do when teaching reading; what they have students do; and, what they assess.

What Teachers Believe (T4Q43). Teachers were asked to indicate their level of agreement with statements about issues in reading instruction. This question provides a glimpse into teachers' beliefs about reading theory and how instruction should be organized.

Table 13-27. What teachers believe about reading instruction: Factor pattern

Factor loading	Item	Disagree	Agree
Factor 1 - *Sequenced Instruction*		Percent*	
0.58	Reading learning materials should be carefully sequenced in terms of language structures and vocabulary	44	41
0.56	Most of what a student reads should be assessed	60	22
0.56	Every mistake a student makes in reading aloud should be corrected at once	82	12
0.55	Teachers should carefully follow the sequence of the textbook	72	14
0.55	Teachers should always group students according to their reading ability	84	13
0.54	All students' comprehension assignments should be carefully marked to provide them with feedback	23	67
0.52	Students should not start a new book until they have finished the last	69	17
0.46	When my students read to me, I expect them to read every word accurately	65	27
0.46	Class sets of graded reading material should be used as the basis for the reading program	37	32
0.45	Students should learn most of their new words from lessons designed to enhance their vocabulary	57	27
0.39	Teachers should keep careful records of every student's reading progress	7	84
0.32	A word recognition test is sufficient for assessing students' reading levels	90	3
0.31	Students who can't understand what they read haven't been taught proper comprehension skills	66	10
0.24	9-year-olds should not have access to books they will read in the next year at school	76	10
Factor 2 - *Extensive Exposure to Reading*			
0.51	Students should take a book home to read every day	13	76
0.41	Every day students should be read to by the teacher from a story book	11	86
0.40	Students should always understand what they are reading	21	58
0.39	All students should enjoy reading	10	82
0.38	Students should be encouraged to read texts they have written	10	95
0.32	Students should always understand why they are reading	12	74
0.30	Most students improve their reading best by extensive reading on their own	11	75

*Percents do not add to 100 because response category "uncertain" has not been included.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Two factors were identified (Table 13-27). However, neither factor falls neatly under a particular reading theory. Based both on an empirical rule of thumb under which we considered factors with eigenvalues greater than one, and a theoretical stance in which the group of items in the first factor contrast with those in the second, we defined two factors from this question block. The first factor, labeled *sequenced instruction*, is characterized by sequencing, mastery of prior levels before moving on, accuracy, and heavy teacher direction. While this stance is likely to be consistent with what phonics advocates might suggest, it is broader than just phonics. Although never specifically stated, one might read into this factor a belief in developmental stages that are carefully orchestrated by either the materials or the teacher. In considering the distribution of teachers' responses to the items in this factor, the general picture that emerges is that, at a minimum, 60 percent of the teachers appear to disagree with beliefs that are consistent with this factor. However, there are four items where this pattern is not as strong. Two items are related to the use of sequenced materials in class. Here teachers seem to be more evenly divided in their beliefs. Teachers also seem to be strongly supportive of providing feedback and monitoring student progress.

In contrast, the second factor, *extensive exposure to reading*, is characterized by students' active involvement in frequent extended reading both at school and at home. There is little mention of teacher direction in this factor. It is characterized most by its focus on what the student does. Here are elements of whole language approaches, with students being given a more central role in constructing meaning. Similarly, there is mention of the integration of reading and writing where students are encouraged to read texts they themselves have written. Teachers appear to strongly support the beliefs espoused in this factor. More than 74 percent of the teachers agree with all but one of the items. In that item, *students should always understand what they are reading*, teachers seem to be permitting students a bit more latitude, and perhaps leaving more room for students to be challenged by working at constructing meaning more interactively.

What Teachers Do (T4Q53). Teachers were asked how often they used specified teaching practices in their classes. The items reflect a teacher's views and behavior with regard to who controls learning (Table 13-28). What is at issue across these questions is the degree of autonomy that students are given.

Across all the items in each of the three factors, there is an underlying assumption that the teacher is orchestrating instruction. The teacher is creating an environment in which students are expected to learn certain things -- both content and process. Within this structured environment there is, however, a broad range in which instruction and learning can flourish. Based both on the empirical test and the theoretical contrasts across sets of items, three distinct patterns emerge.

Factor one, *student centered*, stresses student independence. Students are asked and encouraged to consider and decide how they are doing, what they are doing, and how they will do it. This does not imply anarchy. Rather, within a structured environment, students are given every opportunity to organize themselves and the materials they use to construct meaning. An inspection of the distribution of teachers' responses seems to indicate that at best only about a third of the teachers surveyed are likely to strongly support extensive use of student-centered teaching strategies. It seems reasonable to conclude that, for the most part, teachers are still likely to be making most of the decisions regarding instruction and are probably providing direct instruction.

The items in factor two, *materials directed*, represent the other end of the continuum. Here students are directed as to what to do in a specified sequence. The teacher carefully maps out what will be done in accordance with a highly structured and ordered sense of progression. As indicated by responses to the first three items in this factor, approximately two-thirds of the teachers surveyed indicate that students are expected to frequently work on activities that are skills oriented and orchestrated in specific ways by the teacher or by the materials they have been assigned.

The items in the third factor, *shared direction*, represent a give and take between teachers and students. Teachers still provide a high level of direction and feedback, but students are expected to generate ideas, to share with one another, and to relate what they are learning to their own experiences. What underlies this collection of items is the sense that students are given a great deal of latitude while they work within in a prescribed structure. Although the distribution of teacher responses seems to vary a great deal across the items within this factor, close inspection of the items reveals an inherent logic consistent with the notion of shared direction. Teachers who believe in and practice behaviors that are consistent with an authoritative, facilitating approach are likely to provide students with feedback and less likely to assign specific topics. In principle, there seems to be reasonably high acceptance of this perspective among teachers.

Table 13-28. Teaching practices: Factor pattern

Factor loading	Item	Rarely	Frequently
Factor 1 -- *Student Centered*		Percent*	
0.72	Students are given the opportunity to consider what they think they have learned, as well as their perception of their strengths and weaknesses	70	31
0.72	Students are given the opportunity to assess their own progress	76	24
0.70	Students are encouraged to compare their written texts with the reading selection	88	12
0.69	Students are encouraged to use the reading selection as a source for ideas when writing their texts	61	39
0.65	Students are given the opportunity to provide input on how they will be assessed	92	8
0.60	Students are given the opportunity to work on a variety of different projects	67	33
0.59	Students establish their own purposes and goals	85	15
0.54	Students are given the opportunity to discuss various possible themes for the selection	72	28
0.54	Students are encouraged to compare their written texts with other students' written texts	81	19
0.50	Students decide how they will approach their texts	90	10
0.40	Students have a choice in what they will do	84	16
0.34	Students are given feedback by the teacher on the themes or main ideas of the selections they read	64	46
Factor 2 -- *Materials Directed*			
0.73	Students are given guided practice with skills	34	66
0.53	Specific skills are taught at certain times	35	65
0.53	Students are expected to follow the activities outlined in the lesson the teacher has planned	15	85
0.45	Students are invited to consider how skills apply to what they have written	59	41
0.39	Students are told what they have learned and have yet to learn	54	46
0.38	Students are directed to answer a set of the teacher's questions	55	45
0.29	Students are given teacher feedback on how they compare with other students	87	14
Factor 3 -- *Shared Direction*			
0.61	Students receive feedback from the teacher on their ideas	17	83
0.57	Students are informed as to the purposes of lessons	15	85
0.51	Students deal with issues and topics related to their own experiences	52	48
0.43	Students are directed to proceed based upon set guidelines	23	77
0.43	Students share their ideas with each other	43	57
0.41	Students are told how what they know relates to a topic	49	51
0.27	Students are assigned specific topics to study	62	38

*Percents do not add to 100 because response category "uncertain" has not been included.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In general, it seems safe to conclude that the majority of teachers do not regularly use practices that put the student at the center and with the most control. Rather, the teachers surveyed seem to most favor teaching practices associated with either shared direction or teacher direction.

The three positions represented in these items are sometimes, but not necessarily, associated with particular theories of reading. For example, phonics programs are most often materials or teacher directed. On the other hand, whole language programs are built on a student-centered approach. But programs based on the integration of reading and writing or literature could exist in conjunction with any of these positions.

What Teachers Have Students Do (T4Q30). Teachers were asked how frequently they have students do certain reading activities. In contrast to the last question, where the focus was on descriptions of teacher behaviors, this question looks at the kinds of assignments and activities teachers expect students to complete (Table 13-29).

Table 13-29. What teachers have students do: Factor pattern

Factor loading	Item	Rarely	Frequently
Factor 1 -- *Schema-based activities*		Percent*	
0.76	Making predictions during reading	16	84
0.71	Making generalizations and inferences	15	54
0.67	Relating experiences to reading	21	79
0.65	Orally summarizing their reading	31	69
0.63	Looking for the theme or message	25	75
0.42	Studying the style or structure of a text	60	40
Factor 2 -- *Integrated language arts activities*			
0.63	Listening to students reading aloud to small groups or pairs	32	66
0.59	Discussion of books read by students	73	37
0.56	Dramatizing stories	95	5
0.46	Drawing in response to reading	72	28
0.45	Diagramming story content	82	18
0.43	Writing in response to reading	23	77
0.42	Reading other students' writing	70	40
0.38	Student leading discussion about passage	70	30
0.35	Reading plays or dramas	97	3
0.30	Comparing pictures and stories	45	55
Factor 3 -- *Skills-based activities*			
0.81	Learning letter-sound relationships	41	59
0.65	Word attack skills	23	77
0.37	Learning new vocabulary from texts	8	92
0.35	Answering reading comprehension exercises in writing	9	91
0.35	Playing reading games (e.g., forming sentences from jumbled words)	82	18

*Percents do not add to 100 because response category "uncertain" not included.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Again based both on an empirical rule of thumb, using only factors with an eigenvalue greater than one, and on theoretical contrasts, three factors emerge from this question. In each factor the emphasis reflects a philosophic position. In factor one, *schema-based activities*, students focus on the organization and interrelated aspects of text. They move back and forth from the detail to the overarching theme to make predictions and generalizations. They use what they know from experience and about the structure of text. On all but two of the items included in this factor, over 70 percent of the teachers report frequently having students do these things. In looking at the items, it is clear that they represent very common practices associated with a directed reading lesson and have been suggested and included in teaching manuals for years. With regard to the two remaining items, *making generalizations and inferences* and *studying the style or structure of a text*, if one believed in a hierarchy of skills these would likely be considered beyond the range of a fourth grader. Therefore, it is not surprising that fewer teachers reported frequent use of these activities.

In factor two, *integrated language arts activities*, the emphasis is on bringing all communication modes together. Students listen and discuss; they read and write as well as respond through other symbolic modes (drama, art). That there is a great deal of variability in the frequency with which teachers report using the instructional activities in this group is to be expected given the nature of these items. Having students dramatize stories, or read plays or dramas, is quite time consuming and possibly results in little added benefit. Even if the teacher were highly committed to this type of approach, we would expect such differences among the items. However, in looking at those items that teachers report using frequently, we note that they need not be associated only with this type of program. Students are often asked to read aloud for diagnostic purposes. Students in any class frequently write something in response to reading. And, it is not uncommon to have teachers in any subject area draw students' attention to the accompanying pictures or diagrams in order to make comparisons with the text. Given the dispersion of response rates, one would be very hard pressed to make any statement about teachers' commitment to this approach as a whole.

In the third factor, *skills-based activities*, the emphasis is on what is literally in the text. It is a very bottom-up orientation focusing on letters, words, sentences, and text-based understanding. The teachers surveyed seem to most frequently use the instructional activities included in this factor. There is an almost overwhelming pattern here, with teachers using these activities a great deal. That only 58.5 percent report frequently teaching letter-sound relationships is not surprising, because these are teachers of fourth graders who in principle should have moved beyond this particular type of activity. Similarly, playing reading games would also be most likely to be associated with earlier grades -- preschool, kindergarten, and first or second grade.

What Teachers Assess (T4Q46). Teachers were asked how frequently they assessed certain aspects of reading (Table 13-30). In their assessments, teachers appear to emphasize three different things. As seen in factor one, *contextualized reading*, teachers are testing the entire process. The basics of decoding and vocabulary are given as much emphasis as is relating reading to what the student knows. The second factor, *reading skills*, focuses entirely on the basic subskills of reading -- decoding, phonics. The third factor, *text-based understanding*, maintains a heavy text-based, bottom-up orientation. Teachers focus on what is specifically in the text.

What is most striking about this group of factors and the distribution of teacher responses to the items in each factor is that the teachers surveyed report frequently assessing everything, irrespective of the content implied in the factor, and perhaps irrespective of what they might be teaching.

Table 13-30. **Assessment emphasis: Factor pattern**

Factor loading	Item	Rarely	Frequently
Factor 1 -- *Contextualized Reading*		Percent*	
0.85	Use of background knowledge	11	89
0.72	Literary appreciation	21	80
0.62	Amount of reading	16	84
0.51	Vocabulary	5	95
0.51	Decoding	16	84
Factor 2 -- *Reading Skills*			
0.99	Phonic skills	21	79
0.42	Reading study skills	10	90
Factor 3 -- *Text-Based Understanding*			
0.64	Word recognition	12	88
0.56	Text comprehension	1	99
0.50	Sentence understanding	2	98

*Percents do not add to 100 because response category "uncertain" has not been included.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

In ways that are similar to each of the factors previously discussed in this section, we note that there is a correspondence between each of the assessment emphases and particular theories of reading. However, none of the factors would constitute a complete description of any single theory. There is an overlapping of theoretic stances across the question blocks, and to a lesser degree across the factors that arise within a particular question block.

13.3.5.3. Do Teachers Organize Instruction and Assessment According to an Implicit Theory of Reading?

In principle, teachers might be expected to align their beliefs about instruction, their actions, what they have students do, and what they test according to a consistent theory of either reading or learning. As we noted in the factors that emerged in the preceding sections, no clear theory emerged. Instead in each of the four sets of items there were factors that related to groups of theories. In addition, across the questions there were factors that seemed to be related. A second order factor analysis was undertaken to look at broader second order constructs that might underlie the first order factors.

This second order factor analysis resulted in three relatively meaningful factors -- one that captures all testing and two that distinguish between two schools of thought in instruction, as described below. Table 13-31 shows the second order factor loadings.

Assessment Emphasis. Second order factor 3, the easiest factor to describe, brings all the questions on assessment back together. It also included *skills-based activities.* These activities most often might be workbook or worksheet pages that a teacher would be likely to grade and are not too different in kind from what teachers would use for a skills assessment. Despite the fact that there are three possible emphases, assessment seems to run together. A teacher who tests a great deal is likely to test everything frequently.

The other two factors that emerge are derived from the item blocks on beliefs, practices, and activities. One is associated with reading and learning theories based on a notion of transmission of knowledge from the teacher or author to the student, and the other seems to be most related to notions of reading and learning as an interaction between the teacher or author and the student.

Transmission Emphasis. Instruction and reading theories that can be grouped under the heading of transmission may be characterized as placing the meaning of the text outside the reader who is expected to reproduce it (Straw 1990), organizing teaching according to a prescriptive view of language (Balmuth 1982), providing instruction that is hierarchical and subskill in nature (Barrett 1968; Gray 1960), and processing that is done in a linear fashion (Gough 1985). The first order factors that empirically fell into this category are strongly prescriptive, demand a high level of accuracy consistent with a view of language usage that is correct and that is known by the teacher and the authors of texts and materials.

Table 13-31. Theory and practice combined

Factor loading	Primary factor name
Second Order Factor 1 -- *Interactive Emphasis*	
0.73	Integrated language arts activities
0.72	Schema-based activities
0.72	Student-centered teacher behaviors
0.64	Shared-direction teacher behaviors
Second Order Factor 2 -- *Transmission Emphasis*	
0.69	Material-directed teacher behaviors
0.37	Sequenced instruction teacher beliefs
Second Order Factor 3 -- *Assessment*	
0.76	Contextualized reading assessment
0.75	Text-based understanding assessments
0.63	Skills assessments
0.52	Skills-based activities

What is interesting to note is that, while the surveyed teachers' responses to the items in sequenced instruction beliefs tended to disagree with this position, the teachers frequently reported using material-directed teaching practices.

Interactive Emphasis. An interactive emphasis may be characterized as having the meaning of the text reside with both the text and the reader who is expected to have some background knowledge that fits the text (Straw 1990). There is also an expectation that there will be an interaction between the vernacular language of the student and the more formal language of school and text.

The first order factors in this construct indicate a high level of interaction. There is the integration of reading and writing, of student and author knowledge, and of the shared decisionmaking between the student and the teacher. While student-centered teacher behaviors load heavily on this factor, it is interesting to note that teachers do not report frequent use of these behaviors. One might interpret this to mean that while teachers are beginning to move into this more open interactive environment, they have yet to develop a great deal of comfort and are thus maintaining the control of instruction.

While these analyses are of some interest in themselves as reflections of reading theories and practices, they are not especially useful in explaining differences between classrooms in the average reading comprehension of students. One year's exposure to these instructional emphases is unlikely to alter the existing distribution of reading comprehension among classes that has been formed over the past several years. This means that the study as designed (i.e., a cross-sectional design with no pre- and posttesting) does not allow us to look at the comparative effectiveness of a particular approach.

On the other hand, these data do provide us with a reasonable glimpse at the current state of teachers' instructional practices and beliefs. We might point out that teacher beliefs do not seem to line up with their teaching behaviors or with the instructional activities they are likely to assign. And, it is most evident that their assessment practices bear little relationship to their beliefs, teaching behaviors, or instructional activities.

At best these conclusions are very tentative. This caution is based on two concerns. First, the items were not specifically designed to measure the implementation of the described theories. Second, the data are based on teacher self-report and are likely to be colored to a large degree by notions of socially appropriate responses. Despite these concerns, we believe that future research could effectively be designed to explore this avenue of thought.

13.3.6. Class Attributes

Data on class size were derived from the Teacher Questionnaire item, "In *this* class, what is the total number of students?" As seen in Table 13-32 and Figure 13-17, despite the seeming fluctuation in associated proficency levels, there are no significant differences in achievement correlated with variation in class size.

Table 13-32. Class mean reading proficiency scores, by class size: Grades 4 and 9

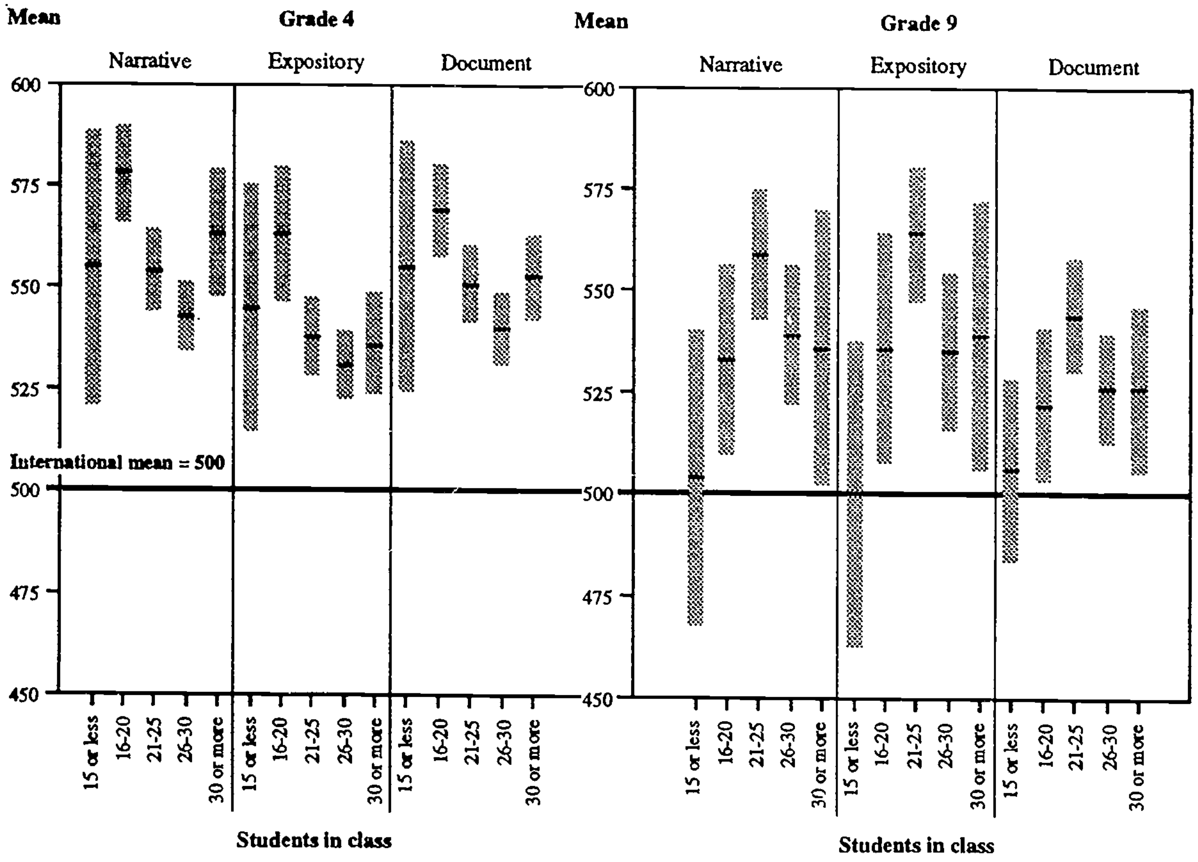
Number of students	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
15 or less . .	4.6	555 (16.8)	545 (15.2)	555 (15.3)	12.1	504 (18.0)	500 (18.6)	506 (11.2)
16-20	19.5	578 (5.9)	563 (8.3)	569 (5.6)	24.2	533 (11.7)	536 (14.2)	522 (9.2)
21-25	33.3	554 (5.1)	538 (4.9)	551 (4.7)	27.9	559 (8.1)	564 (8.1)	544 (6.9)
26-30	32.0	543 (4.1)	531 (4.1)	540 (4.4)	23.0	539 (8.6)	535 (9.7)	526 (6.7)
30 or more .	10.6	563 (7.8)	536 (6.3)	553 (5.1)	12.7	536 (16.9)	539 (16.5)	526 (10.1)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The relationship between class size and achievement has had a lengthy and contentious history in educational thought. The issue at the center of this debate is whether there is, in fact, a sufficient improvement in outcome over time that will justify the immediate and tangible costs of supplying teachers and classrooms.

Figure 13-17. **Mean reading proficiency scores, with 95 percent confidence intervals, by class size: Grades 4 and 9**



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Despite the dozens of studies examining the effects of class size conducted over the past 50 years, there is still no conclusive proof that smaller class sizes result in higher achievement (Tomlinson 1989). Three major works put the issues in perspective. The most often cited work is that of Glass and Smith (1979) who used a meta-analysis of the 80 or so studies on the issue conducted since 1900, with most emphasis on the 14 best studies after 1950. Although their findings seemed to indicate that smaller classes resulted in higher achievements, this conclusion was based on their development of an "effects curve," which indicated two underlying truths:

- As class size increases from 20 to 40 students, there seemed to be nearly no gain or loss in achievement (other things equal, 40 students taught together will learn about 5 percent less than 20 students will); and
- The class size had to be 15 students or less to produce a significant improvement in achievement.

Two critical reviews of this work severely challenged the Glass and Smith findings. In the first, Robinson and Wittebols (1986) concluded that "the relationships between class size and pupil achievement vary greatly across grade levels, among subject areas, and by methods of instruction." In the second, Slavin (1984) concluded that estimates of the benefits from classes of 15 or fewer pupils were also far too optimistic. After carefully reviewing all the studies under consideration, Slavin made three observations. First, some of the studies had no relevance for conventional classroom instruction. Second, because almost no research on class size exists for high schools and only slightly more exists for middle schools, there is no research justification for claiming that smaller classes contribute to higher student achievement at the middle and secondary levels. Third, the evidence of genuinely significant achievement gains at the elementary level does not materialize until class size falls to from 3 to 5 pupils -- and typically only when classes of 1 or 2 students are compared to classes of 20 or more students. As Tomlinson (1989) points out, "missing from the evidence was a study of the scale and design necessary to inform practice as well as remove uncertainty as to the effect on class size on learning."

Two major state level studies have addressed these concerns. Indiana's Project Primetime (Cavin, Murnane, and Brown 1985; Gilman et al. 1987) was based on a sample of Indiana school districts that reduced class size in K-1 to 18 pupils and in grades 2 and/or 3 to 22. At the end of the first year, 50 percent of the sampled Indiana school districts with smaller classes showed *significant* improvement in reading scores and 30 percent had higher math scores. At the end of the second year, the number of districts with measurable differences due to smaller classes had shrunk to 30 percent in reading and 10 percent in math. By the third year, the benefits for smaller classes had vanished entirely.

Tennessee's Project STAR (Finn and Achilles 1990) studied 6,946 students in 350 classrooms in 75 schools over a 3-year period in the most comprehensive controlled experiment to test whether reductions in class size would improve achievement. The project showed that for a period of 1 year, classes of 15 children learned more than classes of 23. The net benefit to achievement was a one-time, one-quarter standard deviation improvement in test scores for these kindergarten or first grade children in small classes. Children in regular classes with an aide did not benefit commensurately, although they tended to exceed the scores of children in regular classes without an aide. Although the initial gain was maintained, scores did not continue to improve in the subsequent 2 years of the project.

Although many believed that Project STAR had provided a more definitive answer to the question of the effects of class size, these findings continued to raise concern. As Tomlinson (1990) points out, ability grouping across classes may have initially benefited the small classes. In fact, Project STAR teachers had noted that the range of ability within the smaller classes was less than in their larger classes. Subsequently, this may have produced an easier teaching task (Johnston 1990). A reanalysis of the Project STAR data by Mitchell, Beach, and Badarak (1991) pointed to the conclusion that "the impact of class size on student achievement is the *indirect* result of differences in the *pattern* of student achievement in each classroom, rather than a direct effect of the number of children in each classroom." "A simple reduction in the number of students is not a sufficient practical or theoretical explanation, ...what teachers do and what students experience in smaller classes are the appropriate topics" (McGiverin, Glass, and Tillitski 1989, 55). In Project STAR, instruction may not have varied because of the change in class size. Teachers were not necessarily taking advantage of the reduced class size to introduce different and more innovative instructional practices (Shapson et al. 1980). Consequently, despite the rather positive findings associated with Project STAR, there appears to be no definitive end to the debate, particularly if one considers the question of cost effectiveness. Further research should ensue.

13.3.7. Principal Attributes

Research on effective schools is consistent in finding the role of the principal important in student achievement outcomes. Therefore, this study has focused attention on principals by collecting data about their

- Gender,
- Race/ethnicity,
- Training,
- Experience, and
- Role as instructional leader and staff developer.

Principal's Gender and Race/Ethnicity. Data relating either the gender or race/ethnicity of the school principal to student achievement are sparse, as reflected by a search of educational research data bases that resulted in finding no body of relevant research. The best inference that might possibly be made is that the principal's gender and race/ethnicity affect student achievement in the same ways that teachers' gender and race/ethnicity would. The data from the Reading Literacy Study indicate that there are no strong, significant correlations between these variables and reading achievement (Tables 13-33 and 13-34), with the exception that the mean proficiency for classes from schools with black principals is substantially and statistically significantly below that of classes with white principals for all three scales at both grades. There are very few conclusions or policy implications that might be drawn from these findings, because we have no information relative to the assignment policies within the district.

Table 13-33. Class mean reading proficiency scores, by principal's gender: Grades 4 and 9

Gender	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Male	60.5	566 (6.1)	552 (6.1)	560 (3.6)	91.0	539 (5.7)	540 (6.8)	528 (4.1)
Female	39.5	554 (4.5)	538 (4.1)	550 (3.6)	9.0	525 (10.0)	527 (11.6)	518 (8.9)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-34. Class mean reading proficiency scores, by principal's race/ethnicity: Grades 4 and 9

Race/ethnicity	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Asian	0.0	*	*	*	0.2	494 (-)	477 (-)	492 (-)
American Indian	0.0	*	*	*	1.8	447 (-)	415 (-)	507 (-)
Hispanic	1.2	503 (14.1)	477 (11.6)	494 (8.0)	0.2	502 (0.0)	508 (0.0)	533 (0.0)
White	92.3	566 (4.1)	551 (4.2)	561 (2.8)	92.0	545 (5.5)	547 (6.5)	527 (4.3)
Black	6.5	491 (8.7)	488 (8.1)	500 (6.6)	5.8	484 (16.4)	489 (16.8)	481 (10.5)

KEY: - Sample size is too small to compute standard error; * - No cases.

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Principal's Training and Experience. Although the literature does not suggest that principal's training or experience is related to student performance, the possibility that some of these factors might be important in some combination was considered. It was anticipated that perhaps the training of principals and/or their experience as administrators might have an effect on the reading performance of students in the school. Accordingly, the School Questionnaire included items on the principal's training in administration and reading, total experience as a principal, and years at the present school. As the data shown in Tables 13-35 through 13-38 indicate, principal's training, experience, and years at the school were not related substantially to reading scores

Table 13-35. Class mean reading proficiency scores, by principal's training (administration): Grades 4 and 9

Degree of training	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Low	30.4	566 (12.6)	554 (15.3)	556 (11.1)	26.7	537 (13.5)	536 (17.0)	526 (11.3)
High	69.6	558 (3.2)	543 (3.5)	556 (4.7)	73.3	537 (5.4)	538 (6.1)	527 (3.8)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-36. Class mean reading proficiency scores, by principal's training in reading: Grades 4 and 9

Degree of training	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Low	56.4	562 (6.8)	552 (6.1)	561 (4.0)	42.6	535 (7.6)	537 (9.0)	526 (6.0)
High	43.6	557 (4.0)	539 (3.7)	549 (3.4)	57.4	539 (7.1)	539 (7.9)	527 (5.2)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, Technical Report, National Center for Education Statistics, 1991.

Table 13-37. Class mean reading proficiency scores, by number of years in principal's career: Grades 4 and 9

Years	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
5 or less .	34.1	552 (7.6)	538 (8.4)	551 (10.2)	20.6	537 (10.2)	543 (12.7)	528 (9.5)
6 - 10 ...	18.2	572 (11.4)	560 (22.8)	567 (12.2)	15.6	531 (9.0)	531 (11.0)	525 (6.5)
11 - 15 ..	15.7	552 (7.4)	538 (6.1)	546 (5.9)	36.3	546 (12.4)	547 (14.7)	532 (8.9)
16 or more	32.0	566 (7.1)	552 (6.3)	560 (3.2)	27.5	535 (9.7)	532 (10.5)	522 (6.5)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-38. Class mean reading proficency scores, by number of years as school principal in present school: Grades 4 and 9

Years	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
3 or less	33.3	544 (5.8)	528 (4.9)	539 (5.3)	22.2	545 (7.6)	553 (8.5)	536 (6.4)
4 - 6	26.1	569 (3.7)	554 (6.1)	569 (8.7)	38.7	525 (9.1)	522 (11.1)	517 (7.1)
7 - 12......	24.2	571 (12.0)	561 (17.1)	563 (8.9)	20.2	534 (11.0)	533 (12.4)	521 (7.9)
13 or more ..	16.4	562 (8.1)	550 (7.0)	561 (6.7)	18.9	544 (5.0)	543 (13.7)	532 (7.5)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Principal's Role. Leithwood (1990) identified four different functions of a school principal: adminstration/plant management, interpersonal relations/climate, program, and student development. The first two are directed at maintaining the school and are the primary focus of a majority of principals (Morris et al. 1986; Trider and Leithwood 1988). The second two foci, which Leithwood groups into a single category called instructional leadership, have been associated with improvements in student outcomes. In effective schools, principals are seen as central to the formulation of well-articulated school goals related to instruction and student achievement (Bossert et al. 1982; Edmonds 1979; Hoy and Ferguson 1985).

Given the importance of the principal as an instructional leader, we examined the Reading Literacy Study data describing the principal's job emphasis. These data were obtained from teachers' responses to a list of items describing principal/teacher interactions. An exploratory factor analysis resulted in two factors, which we named principal's emphasis on instruction and principal's emphasis on staff development (Table 13-39).

Table 13-39. The principal's role: Factor pattern

Factor loading	Item
Factor 1 -- *Instruction*	
0.78	Discuss explicit achievement standards for the subject that you teach
0.71	Make suggestions about the content that must be covered in reading
0.67	Make suggestions about the choice of instructional methods in reading
0.56	Ask for evaluation results or progress of your students in reading
Factor 2 -- *Staff Development*	
0.8	Initiate activities directed at the professional development of teachers
0.6	Encourage contacts among teachers

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 13-40 shows that the fourth and ninth grade class mean proficiency associated with principals with the least emphasis on instruction (quartile 1) is significantly higher than that associated with principals with the most (quartile 4) or with moderately great (quartile 3) emphasis on instruction across all scales for both grades.

Table 13-40. Class mean reading proficiency scores, by principal leadership: Grades 4 and 9

	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Quartile 1 (low) .	27.5	570 (4.3)	557 (5.3)	567 (4.2)	21.9	552 (9.2)	559 (10.7)	542 (7.5)
Quartile 2	28.8	563 (6.6)	543 (5.7)	551 (3.8)	25.7	548 (13.9)	548 (15.7)	531 (10.6)
Quartile 3	17.2	545 (7.4)	533 (7.3)	543 (7.5)	22.8	520 (9.2)	522 (9.5)	519 (6.5)
Quartile 4 (high)	26.4	546 (6.8)	530 (5.1)	545 (5.6)	29.6	529 (10.0)	527 (10.9)	518 (7.8)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The findings in Table 13-41 suggest that the relationship between principal's emphasis on staff development and mean reading proficiency is negligible. There are no statistically significant differences between the groups for any scale at either grade. Both sets of findings appear to be counterintuitive relative to the literature cited above. To understand what might be going on, we turned to the research literature for guidance.

Table 13-41. Class mean reading proficiency scores, by principal's emphasis on staff development: Grades 4 and 9

Quartile	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Low	22.8	548 (6.9)	533 (6.1)	545 (7.1)	60.0	540 (6.3)	542 (7.8)	531 (5.2)
Medium	24.5	561 (4.8)	546 (4.8)	555 (4.9)				
High	52.7	559 (4.1)	544 (3.9)	554 (3.1)	40.0	533 (8.4)	532 (9.4)	521 (6.7)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

According to McLeary and Thompson (1979), despite the importance of the principal's leadership, not all principals engage in instructional leadership activities. They found that while many principals consider instructional leadership to be their most important function, they report having little time or opportunity to be engaged in this type of activity. Cohen (1983) noted that time spent in this role by effective principals resulted in their taking increased responsibility for instruction. They observed teachers regularly, discussed teachers' work problems, supported teachers' efforts to improve by providing appropriate staff development activities, and helped those teachers who try new instructional approaches by providing feedback and assistance. Brookover et al. (1979) found that principals in effective low-SES schools dropped in on classes frequently (approximately 30 times per year for each class).

While these are the observable aspects of an effective principal's behaviors, these descriptions may not necessarily capture the nuance of the effective principal's interaction with the faculty. According to Rosenholtz (1985), part of the principal's leadership role is to establish high academic standards and expectations for all students and to convey that attitude throughout the school and community. The effective principal would act in accordance with the belief that these high expectations can be realized because the teachers are effective and the students themselves are capable of learning (Rosenholtz 1985).

To support and cultivate the effectiveness of the faculty, principals in "effective schools" give a high priority to formal staff development programs, encouraging the ongoing training and retraining of teachers and supervising teacher monitoring programs (Lee, Bryk, and Smith 1993; Reynolds and Cuttance 1992). These principals actively construct time for joint planning and problem solving, within the context of inservice programs involving a large portion of the staff (Armor et al. 1976; Phi Delta Kappa 1980; Rosenholtz 1985; Cohen, 1983). These meetings may serve a mediating role related to increased student achievement as they result in increases in teachers' knowledge about teaching and about subject matter (Clark, Lotto, and Astuto 1984; Lee, 1993; Lee, Bryk, and Smith 1993). The underlying and unifying characteristic of these actions may be the facilitating nature of the principal's behavior as opposed to an evaluative or controlling role that would be associated with the principal's supervisory functions.

13.3.8. School Attributes

The school itself acts as a context for the classroom in which reading comprehension skills are learned and refined. There are at least three ways in which schools may differ that are likely to have an impact on what goes on within classrooms. The first is related to the organization of authority or its governance structure (Coleman, Hoffer, and Kilgore 1981; McPartland and McDill 1982; Murnane 1981; Bryk and Driscoll 1988; Chubb and Moe 1990). In this study, as in the literature, this refers most often to the school sector--be it public or private. The second perspective is related to attributes of the school such as size, the proportion of specialist teachers in the school, and the proportion of minority teachers. The third perspective is related to the actions of the administration, most specifically as they relate to the allocation of resources (Murphy 1988; Bridges 1982; Fuller and Izu 1986; Newmann, Rutter, and Smith 1989; Lee, Dedrick, and Smith 1991; Bidwell 1965; Little 1982; Rosenholtz 1985; Meyer and Rowan 1983; Gersten, Carnine, and Green 1982). In this study, instructional time and library resources, the available resources a principal might allocate, appear to have had an effect.

The school-level factors included in this study are

- School sector (public/private),
- School size,
- Proportion of specialist teachers,
- Proportion of minority teachers,
- Instructional hours, and
- Library resources.

Sector. Schools are identified as either public or private based on a dichotomous choice to a single question in the School Questionnaire for each grade. While much of the related research on differences between public and private schools creates a separate category for Catholic schools, the Reading Literacy Study does not.

The sample of schools in the United States included 142 public and 25 private schools for grade 4 and 147 public and 18 private schools for grade 9. The pattern of student performance observed in the Reading Literacy Study was similar to patterns appearing in other large studies such as NELS:88

and NAEP. However, there were only significant differences in reading performance favoring students attending private schools on the narrative scale at grade 4 (Table 13-42 and Figure 13-18).

Coleman, Hoffer, and Kilgore (1981) and Coleman and Hoffer (1987) sparked a great deal of interest in the effect of differences in school governance when they reported that private secondary schools produced superior academic achievement compared with public schools (Lee, Bryk, and Smith 1993). Many researchers criticized the initial research on methodological grounds.[15] However, after a number of reanalyses and subsequent investigations, researchers have concluded that average achievement is somewhat higher in Catholic high schools (Lee, Bryk, and Smith 1993; Jencks 1985).

To explain this difference, researchers looked closely at the factors that might be influencing this apparent advantage of private and Catholic high schools. To illustrate, we draw your attention to two -- selectivity and course offerings. It was suggested that when students choose a school or are chosen by a school, there are associated differences in the organizational life of the school (McPartland and McDill 1982; Murnane 1981; Salganik and Karweit 1982). For example, the selectivity and the costs involved necessitate a greater involvement by parents such that there would be a stronger consensus among the school, parents, and students thus strengthening and reinforcing the school's value systems and in general creating a more motivated student body (Bryk and Driscoll 1988). Additionally, the private high schools not only had greater opportunities in selection of the student body but could also more easily remove students who did not conform (Grant 1988). Similarly, private high schools exercised greater authority in controlling faculty membership (Chubb and Moe 1988, 1990; Bridges 1986). Clearly, it was easier to replace or remove faculty who were considered unacceptable.

Another contributing factor, after controlling for student characteristics, may be the curricular organization of the school -- including differentiated access to academic opportunities (Lee 1988). Course-taking patterns seem to differ. Because private high schools seem to generally offer fewer nonacademic courses than large, comprehensive public high schools, students enrolled in private high schools follow a more academic course of studies (U.S. Department of Education 1993).

The Reading Literacy Study data are not sufficient to add much more to this debate.

Table 13-42. Class mean reading proficency scores, by school sector: Grades 4 and 9

Sector	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Public	85.1	553 (3.8)	539 (3.7)	551 (5.0)	89.1	534 (4.0)	534 (6.0)	526 (3.8)
Private	14.9	577 (10.4)	565 (14.3)	568 (10.7)	10.9	544 (19.2)	542 (22.0)	529 (14.8)

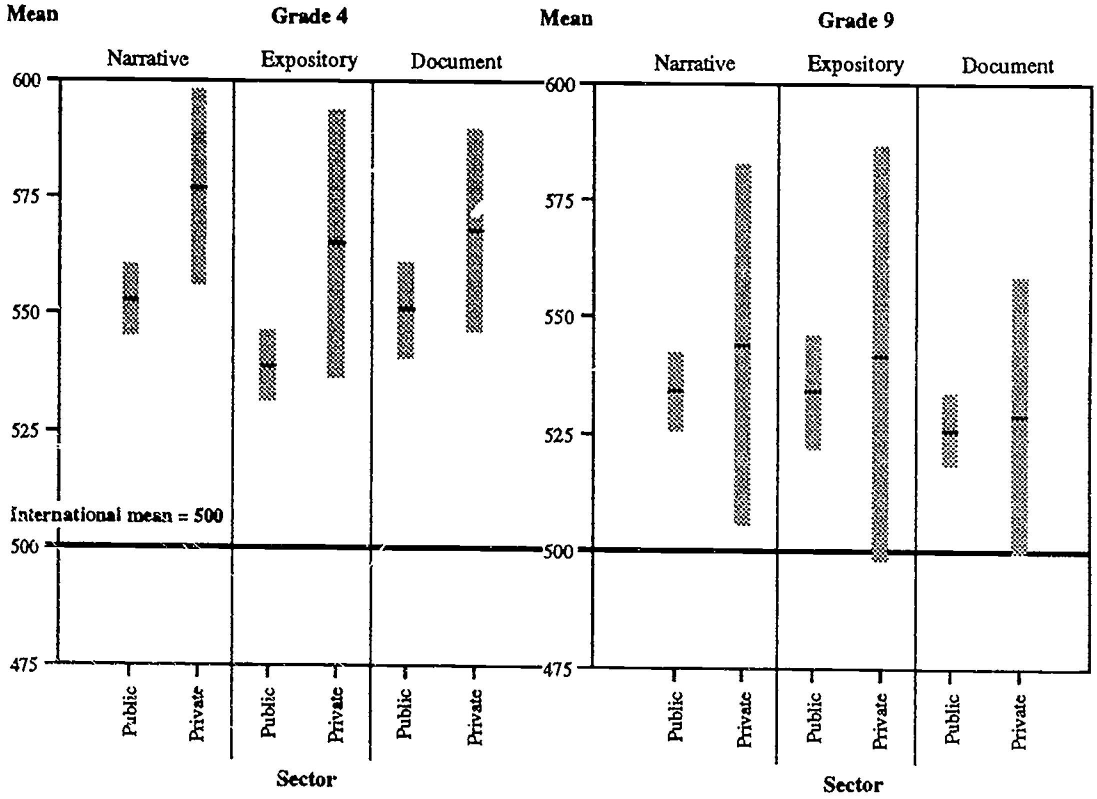
NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

---

[15] See *Harvard Educational Review*, November 1981, and *Sociology of Education*, Spring 1982, for discussions of the issues.

Figure 13-18. **Mean reading proficiency scores, with 95 percent confidence intervals, by school sector: Grades 4 and 9**



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

School Size. School size (i.e., the total enrollment of a school) was based on the first question in the School Questionnaire (Table 13-43). Although there has been considerable debate over the most advantageous size for schools at both elementary and secondary levels, the current data suggest that there is no consistent effect for school size on reading achievement at either the elementary or the secondary level. It is important to note, however, that school size is not independent from other variables, such as urbanicity, region, or proportion of specialists.

Table 13-43. **Class mean reading proficiency scores, by school size: Grades 4 and 9**

School enrollment	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Less than 250	34.8	568 (10.5)	557 (9.8)	566 (7.4)	32.6	529 (7.2)	527 (11.9)	522 (7.1)
250 - 499 . . .	41.3	560 (4.9)	544 (4.4)	555 (4.1)	24.9	532 (11.5)	532 (13.5)	526 (9.1)
500 - 999 . . .	21.2	551 (4.6)	537 (4.4)	545 (4.1)	24.0	535 (8.1)	536 (9.1)	524 (5.9)
1,000 or more	2.7	536 (24.0)	526 (21.2)	538 (24.3)	18.5	553 (9.5)	556 (12.2)	540 (6.9)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Debate about benefits of larger or smaller schools focuses mainly on two primary issues: economic benefits derived from economy of scale, and the degree of formality of social interactions required by a larger organization (Lee 1993). Advocates of larger schools suggest that economy of scale provides marginal residual resources that may be applied to strengthening a school's academic offerings. Another potential benefit is that larger schools seem to hire better trained teachers. One of the underlying assumptions of the economy-of-scale argument is that there is a decrease in the proportion of administrative to instructional staff. In fact, Lee (1993) finds that there is an *increase* in the number of noninstructional staff required to handle the greater bureaucratic needs of larger schools. Friedkin and Necochea (1988) report similar findings, noting that in large schools, energy is shifted away from teaching and toward administrative activities.

One of the difficulties in examining effects of school size is that size may be related to urbanicity, school district size, students' SES, and school sector differences. Also, since larger schools tend to have more programs for their more diverse student populations, including those with special needs, schoolwide achievement scores may be misleading. For example, Friedkin and Necochea (1988) find that increasing school size has a negative impact on low-SES populations, but not on high-SES populations.

Lee (1993) finds that school system size is a more important variable than the size of the individual school. This finding is also presented by Friedkin and Necochea (1988), who explain that community SES is a major factor that influences both the constraints and opportunities of a school system. Larger systems tend to serve more exceptional students, leaving fewer economic resources available for the majority of their students. Another consideration is the diversity of curricular offerings available in schools of different sizes, especially at the secondary level. Lee and Bryk (1989) find that larger secondary schools with more diverse curricular offerings increase the initial differences in relating student background to achievement. These affect the distribution but not the mean of achievement scores.

Proportion of Specialist Teachers. Data on the proportion of specialist teachers within a school were obtained from items on the School Questionnaire asking for the number of full-time (or full-time-equivalent) regular classroom teachers in the school, and the number of full-time (or full-time-equivalent) employees who were special education teachers, guidance counselors, reading specialists, and librarians and other professional media staff. For grade 4, no statistically significant differences were found. For grade 9, in general, having fewer specialists on the staff was associated with higher average achievement for the school. As noted in Table 13-44 and Figure 13-19, there were significant large differences between means for ninth grade classes from schools with few specialist teachers (less than 10 percent) and those schools with many (over 25 percent), for all three domains.

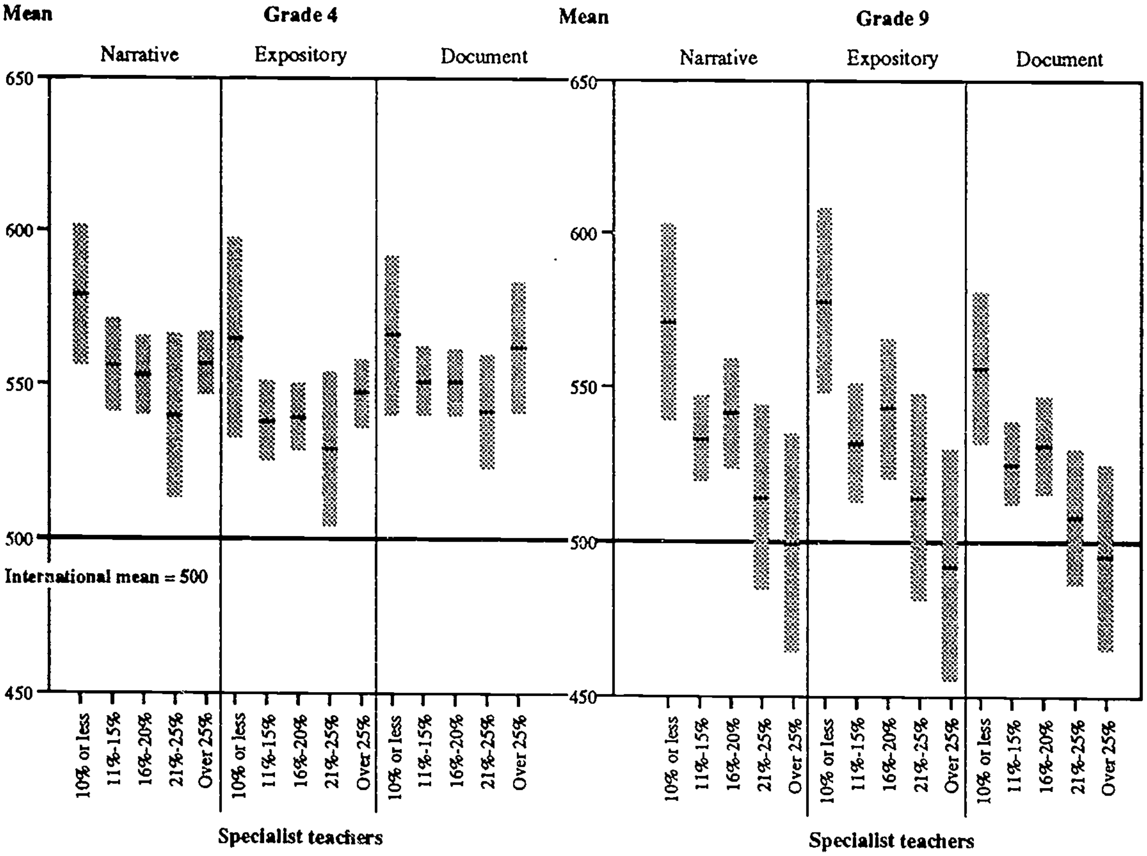
Table 13-44. Class mean reading proficiency scores, by percentage of faculty who are specialist teachers: Grades 4 and 9

Percent specialist teachers	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
10% or less	21.9	579 (11.4)	565 (16.3)	566 (13.1)	10.7	571 (15.8)	578 (15.0)	556 (12.1)
11%-15%	27.9	556 (7.5)	538 (6.2)	551 (5.6)	38.2	533 (6.9)	532 (9.4)	525 (6.6)
16%-20%	18.7	553 (6.2)	539 (5.4)	551 (5.3)	31.6	542 (8.8)	543 (11.3)	531 (7.7)
21%-25%	10.7	540 (13.3)	529 (12.4)	541 (9.2)	14.3	514 (15.0)	514 (16.5)	508 (11.0)
25% or more	20.8	557 (5.1)	547 (5.7)	562 (10.5)	5.3	499 (17.7)	492 (18.7)	495 (14.9)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-19. Class mean reading proficiency, with 95 percent confidence intervals, scores, by percentage of faculty who are specialist teachers: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

These findings are influenced by a number of factors. Most importantly, schools with large numbers of specialists are likely to be those with extensive programs for students with special needs, such as special education, ESL, or remedial reading. Since the achievement scores of these students are included in the achievement data, they depress the achievement scores for the school as a whole. Second, schools with significant problems relating to student behavior (including violence, drug abuse) are likely to employ more counselors and other types of professional specialists to help students deal with these problems. At the secondary level, low-SES students, minority students, and students with lower aspirations are most likely to need counseling regarding curricular decisions, but are least likely to actually get help from counselors (Lee and Ekstrom 1987).

There are other reasons, however, why reading achievement might be influenced by a school having a large proportion of specialists. From an organizational perspective, Lee et al. (1993, 216) suggest that staff specialization is problematic as it fosters transient interactions between teachers and students and creates barriers to more generalized affiliative adult-student relationships. Other researchers express concern for fragmentation of student's experiences (Newmann 1981). Other research, consisting of ethnographic studies, indicates that a more diffuse, less specialized role of teachers facilitates classroom instruction (Schwartz, Merten, and Bursik 1987). Finally, specialization creates the situation where more complex structures are required to maintain the enterprise (Lee 1993).

However, some aspects of using specialists and special programs are effective in promoting student achievement. For example, Doss and Holly's (1982) evaluation of Title I programs indicates that schoolwide programs are more effective than pull-out programs. In schools with schoolwide programs, class size was smaller because additional personnel were used to reduce class size and to improve instruction for lower achieving students within regular classrooms. Also, bureaucratic organization theory holds that staff specialization enhances a school system's efficiency in delivering educational services. However, there is little empirical scrutiny on whether benefits accrue to students as a result of this specialization. Such research requires assessment of direct effects of each special program or activity as well as indirect effects on overall school organization (Lee 1993). With increased specialization, teachers have greater freedom to determine courses they will offer and activities in which they will engage (Powell, Farrar, and Cohen 1985; Cusik 1983).

Proportion of Minority Teachers. As previously described, a considerable majority of teachers belonging to minority racial/ethnic groups teach in schools with high proportions of minority students, most often in urban settings. While the data show a strong and significant relationship to student achievement (Table 13-45), there are numerous other factors contributing to this effect that have been described previously.

Table 13-45. Class mean reading proficiency scores, by percentage of minority teachers in the school: Grades 4 and 9

Percent of minority teachers	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
5% or less . . .	65.4	575 (4.8)	561 (5.5)	569 (4.0)	59.6	549 (6.5)	550 (7.4)	538 (4.9)
6 - 10%	10.6	566 (5.8)	547 (5.4)	561 (5.1)	21.9	531 (18.2)	525 (19.0)	521 (14.9)
11 - 33%	12.4	539 (8.0)	529 (6.0)	540 (7.4)	13.6	509 (14.0)	514 (19.6)	505 (11.2)
33% or more . .	11.6	496 (6.4)	487 (4.3)	500 (3.3)	4.9	471 (13.1)	476 (11.5)	473 (6.4)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Instructional Time Per Week. Time provided for instruction is well documented as a significant factor in promoting academic achievement (Fisher et al. 1980; Brookover et al. 1979). While many studies of effective schools vary on their findings with respect to the relative importance of a number of variables, some variables are found to characterize effective schools in nearly all studies, regardless of methodology or the interaction of other school characteristics. One such factor is high academic engagement, which requires providing large amounts of class time for direct instruction in reading (Austin and Garber 1985; Blumberg and Greenfield 1986; Lipsitz 1983; Peterson 1988; Sergiovanni 1987; Squires, Huitt, and Segars 1983; Ubben and Hughes 1987; U.S. Department of Education 1986, 1987).

For grade 4 and grade 9, the variable "instructional time per week" is determined by responses to the School Questionnaire question, "What is the *total instruction time* (in hours and minutes), excluding breaks, for this class in a typical week (for all subject areas)?"

For grade 4, schools offering 31 or more hours per week had significantly higher class means for the expository and document scales, but not the narrative scale, than those offering 26 to 30 hours per week. For grade 9, the means for schools providing 31 or more hours per week instruction were significantly higher than those of schools offering 25 or fewer hours per week, for all three scales (Table 13-46 and Figure 13-20). It is important to note that this variable does not account for the length of the school year, the length of a class period, or the proportion of time devoted specifically to literacy-related instruction (such as reading or English classes). These other factors, in addition to total instructional time per week, are used in much of the related research.

Table 13-46. Class mean reading proficiency scores, by school instructional time: Grades 4 and 9

Hours	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
25 hours/week or less . .	32.3	554 (8.9)	542 (14.5)	548 (10.7)	16.4	515 (10.2)	514 (10.7)	512 (7.1)
26-30 hours/week	54.8	560 (6.8)	544 (5.4)	554 (3.7)	63.1	538 (9.9)	536 (12.6)	526 (8.3)
31 hours/week or more .	12.9	573 (4.5)	568 (4.2)	585 (10.6)	20.4	545 (7.6)	554 (9.9)	539 (6.1)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

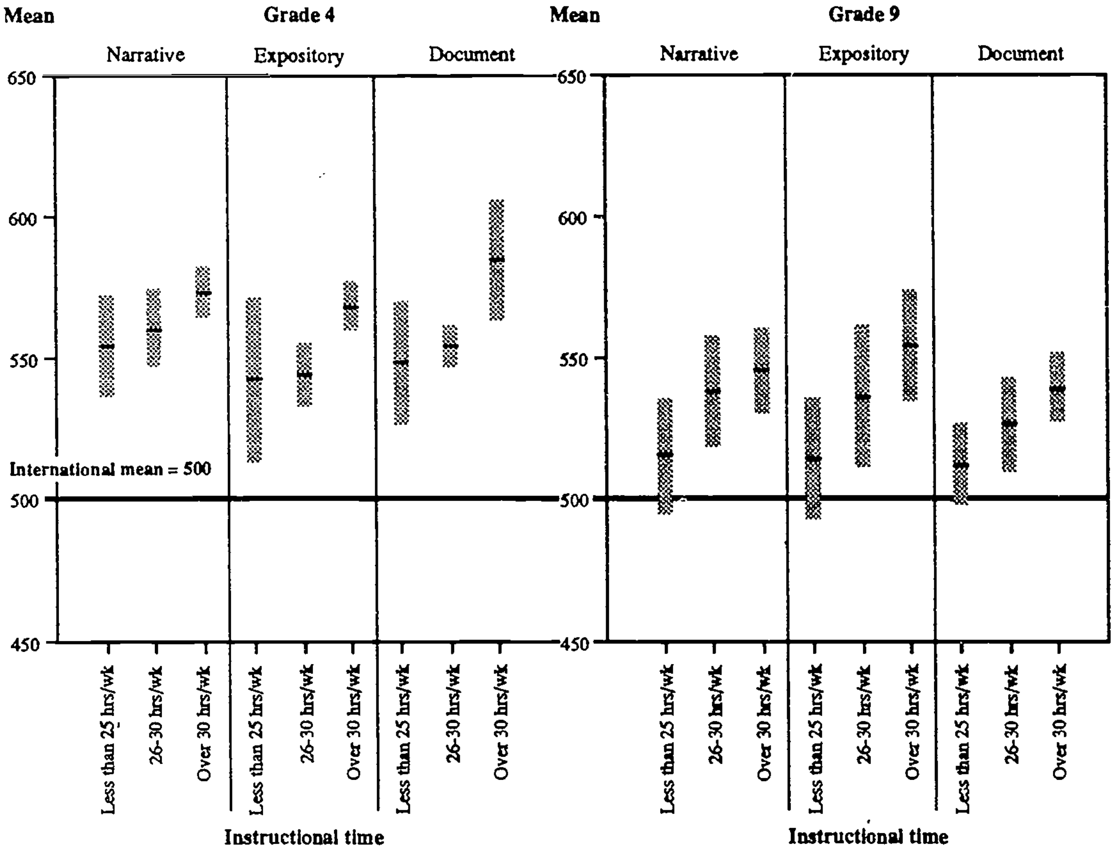
The NELS:88 study reports that most public schools hold 180 days of school per year, with each day lasting 6.5 hours and consisting of 7.1 classes per day of 48.3 minutes each. This data translates into approximately 28.6 hours per week of class time [(48.3 minutes x 7.1 x 5 days) / 60 minutes per hour]. Urban schools, however, are more likely than suburban or rural schools to have a longer school year (i.e., more than 180 school days per year), but a shorter school day, shorter class periods, and fewer classes per day than rural schools. Thus, the total time available for instruction in urban and rural schools may be equalized by these factors. In the Beginning Teacher Evaluation Study (Fisher et al. 1980), approximately 58 percent of the school day was allocated to academic subjects, and 24 percent to nonacademic instructional areas such as music, art, and physical education. Another 18 percent was spent on noninstructional activities such as transitions, housekeeping, and waiting between activities.

The School and Staffing Survey (Choy et al. 1993) findings are consistent with NELS:88. Elementary teachers in self-contained classrooms spent an average of 20 hours per week in academic instruction, according to the survey, of which 48 percent was spent on literacy-related instruction. These amounts represent a statistically significant decrease from 1987-88, when teachers spent 21 hours per week on instruction, allocating 49 percent of their time for English and language arts. Although public school teachers spent somewhat more time per week than private school teachers on reading instruction, there were no differences by school control in the distribution of instruction across academic disciplines.

The 1992 NAEP data show that for grade 4, schools among the lowest scoring third of the participating schools were more likely than schools in the highest scoring third to allocate 90 minutes or more per day (7.5 hours per week) to reading instruction. As the authors of the NAEP report point out, this data may reflect that staff at schools with poorer performance scores recognize their students' need for more reading instruction and consequently provide it.

The relationship of time to academic performance includes a number of factors. The Beginning Teacher Evaluation Study (BTES; Fisher et al. 1980) describes a construct the authors call

Figure 13-20. Class mean reading proficiency scores, with 95 percent confidence intervals, by school instructional time: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

"academic learning time," which includes several time-related factors associated with student performance. One of the key components is *allocated time*, or to the amount of time actually available for instruction after subtracting out time for such things as lunch, physical education, movement between classes, opening exercises, recess, and dealing with disruptive behavior from the time included in the overall school day. The BTES found that length of school day was not directly proportional to allocated time, since there was more allocated time in some classrooms than others in the same school, based on teachers' ability to minimize time spent in certain noneducational activities. Similarly, many other researchers (Berliner 1984; Rosenshine and Berliner 1978; Denham and Lieberman 1980; Rosenshine 1979; Rosenshine and Stevens 1984; Bickel 1983) indicate that an increase in allocated time is only effective in improving learning outcomes if the time is used specifically toward a learning outcome. A review by Rosenshine and Stevens (1984) indicates that "there was no nonacademic activity which yielded positive correlations with reading and mathematics achievement" (p.754). Brookover et al. (1979) report that observational study data indicate that in high-achieving schools, more time is devoted to teaching (minimizing time for study halls and doing homework assignments) than in low-achieving schools. Some between-school differences can be attributed to specific actions taken by the principal of a school. Studies have found that teachers in effective schools have more instructional time per week because principals buffer them from losing time to interruptions, school assemblies, and other low-priority intrusive events (Stallings 1980; Fisher et al. 1980; Rutter et al. 1979; Glenn and McLean 1981).

Stallings (1980) describes time as the critical difference in achievement, because it is a metric for the *quantity* of potential instruction; that is, it limits the amount of content teachers can cover. Rosenshine and Stevens (1984) find a significant positive relationship, ranging from correlations at 0.2 to 0.7 between content covered and student achievement, in nearly all studies they review. Rosenshine and Stevens suggest that Durkin's findings (1978-79) that reading comprehension is best taught by successive practice suggests that increased time for practice would be likely to improve reading comprehension.

One source of variability in data on the importance of allocated time may result from the source of the data being used. Brookover et al. (1979) found that teachers' reports of the amount of time devoted to instruction in a written survey were not confirmed by direct observations in their classrooms, nor were their reports consistent with the relative amounts of time reported. However, they did find that principals' reports of teachers' time allocated to instruction and their own time allocated to supervision were more reliable. The data used here in the Reading Literacy Study are from the School Questionnaire, which was usually completed by the school principal.

Library Resources. The data on library books per student serves as an indicator of the financial resources of schools, and the extent to which those resources are allocated to literacy-related products. The assumption is that schools with greater financial resources will have proportionally more books in their libraries for the size of their total enrollment. The data for this variable are ratios that come from responses to two items on the School Questionnaire: "What is the total enrollment of full-time students in your school?" and "Approximately how many books *with different titles* does your school library contain?"

For grade 4, schools with more than 20 books per student performed significantly better on narrative reading than schools with 10-20 library books per student. Although some other possible relationships between reading proficiency and number of library books per student are suggested by the results, no others are statistically significant.

The literature suggests that there is little consistent relationship between variables related to spending money and student achievement. Such variables include not only library resources but also class size, teacher salaries, and compensatory education programs, and these other variables may have a stronger effect (Hanushek 1981; Mullin and Summers 1981; Murnane 1980; Purkey and Smith 1983). Use of the school library may also be an important factor to consider (Table 13-47). While schools may have large libraries, if access is restricted, the number of books is unlikely to have an impact on reading proficiency.

Table 13-47. Student mean reading proficiency scores, by school library resources: Grades 4 and 9

Books/Students	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
10 or less	22.9	553 (15.3)	545 (22.2)	552 (15.7)	9.9	542 (13.3)	544 (14.1)	530 (10.2)
11 - 20	48.2	556 (4.7)	542 (4.9)	557 (6.9)	32.5	538 (9.6)	539 (9.3)	528 (7.2)
21 or more	28.9	573 (6.3)	555 (6.4)	558 (4.2)	57.6	533 (6.2)	533 (9.4)	525 (5.8)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

13.3.9. Community Attributes

In considering school effects, there is strong reason to believe that the community is likely to influence the character, mission, and ethnicity of the school. Four community attributes were included in the Reading Literacy Study:

- Urbanicity,
- Resources,
- Region, and
- Parental cooperation.

Urbanicity. Urbanicity refers to the population density of the area where a school is located. Responses to the item on the School Questionnaire used to determine this attribute were divided into five categories (Table 13-48). Data from the Reading Literacy Study on urbanicity of schools could not be compared directly to data from NAEP or other national studies because different categories were used in other research. The table shows no significant differences among class-level means by urbanicity for grade 4. At grade 9, large cities and their suburbs show significantly lower classroom means than medium size cities for all three domains.

Findings of the Reading Literacy Study regarding urbanicity alone cannot be interpreted because there are many factors associated with urbanicity that could account for observed achievement differences. Such factors include school size (rural schools tend to be small and urban schools are often large), as well as SES, race/ethnicity of students, race/ethnicity of teachers, and community resources. Urbanicity will also be reflected in data on regional differences in achievement, since the South tends to have more rural schools than other areas of the nation.

Table 13-48. Class mean reading proficiency scores, by urbanicity: Grades 4 and 9

Urbanicity	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Rural	22.5	557 (9.0)	549 (7.8)	564 (11.1)	48.5	532 (7.3)	532 (9.7)	522 (6.8)
Small town	27.2	563 (9.6)	550 (16.2)	558 (11.7)	18.1	549 (6.7)	549 (7.6)	534 (5.9)
Medium size city ...	14.0	565 (7.2)	547 (6.9)	557 (6.4)	9.4	569 (15.7)	573 (17.3)	545 (10.4)
Large city/suburb ...	19.8	562 (16.4)	546 (12.8)	548 (7.9)	10.9	501 (13.7)	501 (17.5)	501 (10.3)
Very large city/suburb	16.5	554 (9.0)	536 (8.1)	552 (6.5)	13.1	538 (14.7)	541 (18.0)	533 (11.9)

NOTE: Numbers in parentheses are standard errors.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Like the current study, the comparisons made in many other studies are not valid (Purkey and Smith 1983). A considerable amount of the research on effective schools was conducted in urban schools. For example, Rosenholtz's (1985) review of findings on effective schools is limited to inner city schools serving low-SES students. This review points out many reports of a "dismal and discouraging" picture of urban schools being painted by this research. In many urban schools, dropout rates of over

50 percent are not uncommon (Natriello 1986). However, Rosenholtz also discusses research showing instances where a particular organizational structure has produced achievement test results far exceeding other schools serving similar populations. Her conclusion is that the achievement of students attending inner city schools can be improved by specific actions taken by the principal and teachers of the school. This would tend to support the idea that what works in urban schools with high minority, low-SES enrollments may not be effective in another setting (Purkey and Smith 1983).

Community Resources. The attribute of community resources was determined by a single composite representing responses on the School Questionnaire to the question, "Please indicate the availability of the following resources in relation to your school (public library; bookstore/book department in a store; other secondary level school; a higher education institution; museum)."

Because of the skewness of the data, responses are represented as a dichotomous variable, indicating schools with either "high" or "low" community resources (Table 13-49). The data show no significant relationship between community resources and proficiency. There are more schools scoring high than low on community resources, and, in fact, many schools responded that all the resources listed were available in their community.

Table 13-49. Class mean reading proficiency scores, by community resources: Grades 4 and 9

Level of resources	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Low	37.7	558 (5.9)	549 (6.5)	560 (5.5)	44.2	532 (4.8)	534 (9.0)	525 (4.5)
High	62.3	562 (6.2)	544 (5.3)	553 (3.3)	55.8	541 (8.2)	537 (9.5)	529 (6.9)

NOTE: Numbers in parentheses are standard errors

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The difficulty of examining the effects of community resources is that they include the social elements of a community as well as its facilities. "The social elements of community," as Newmann and Oliver pointed out, "should be seen not as instruments toward another end but rather as ends in themselves" (Lee, Bryk, and Smith 1993). Reliance on the local community, however, rather than on larger governmental structures, creates increased variability in school-level results (Fuller and Izu 1986). Therefore, Levine (1992) suggests that researchers and practitioners in the U.S. should question the role and meaning of parent and community involvement in schooling. As a correlate of effective schools, they are concerned that terms should be clarified so that results can be compared.

Region. Proficiency in reading has been shown to differ between regions of the United States. From the Coleman study (1966), in which Northern states showed higher scores than Southern states, to the 1992 NAEP (with similar findings), regional averages have differed significantly.

The Reading Literacy Study data (Table 13-50 and Figure 13-21) show that class means in the Southeast continue to be the lowest in the nation for the narrative scale at grade 4. The estimated classroom means for the other five grade/scale combinations were lower for the Southeast than for other regions, but here the difference between the Southeast and any other region was not statistically significant. It is important to consider, however, that this result may be influenced by the number of cities with large populations of low-SES and race/ethnic minority students in schools.

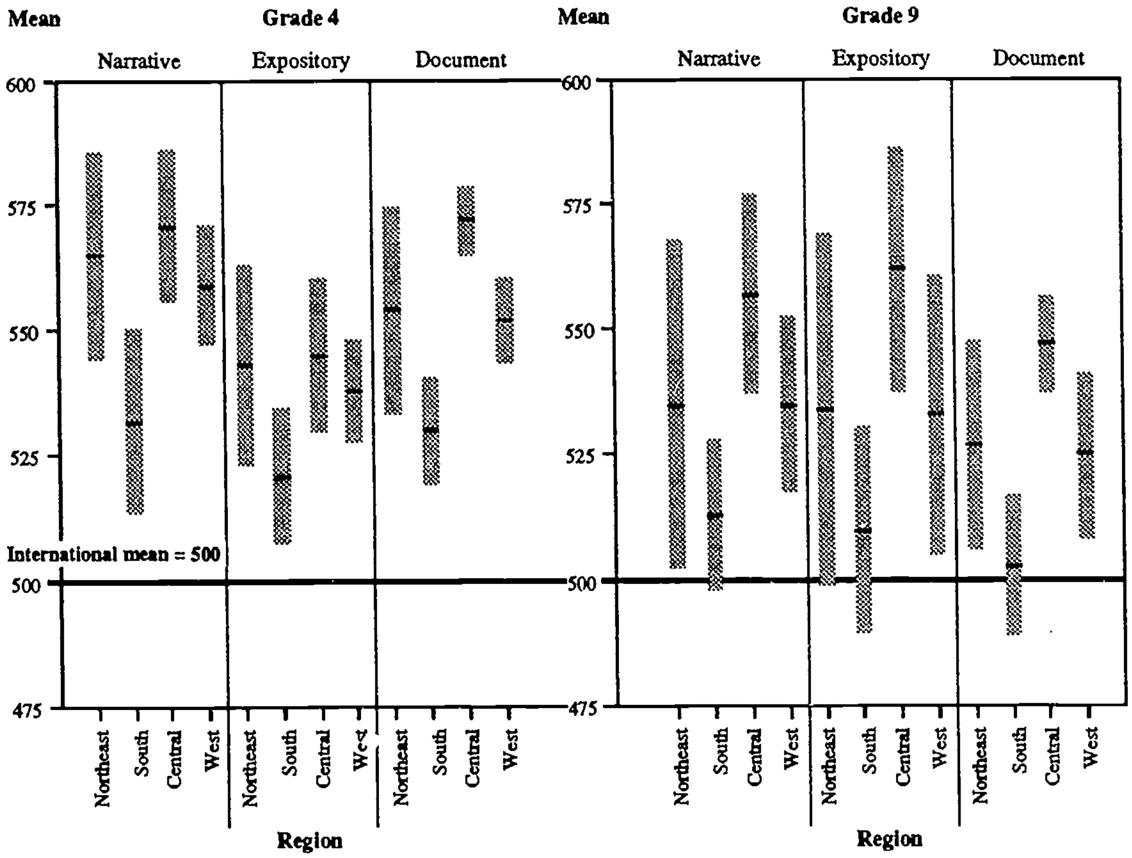
Table 13-50. Class mean reading proficiency scores, by region: Grades 4 and 9

Region	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Northeast	21.1	565 (10.3)	543 (10.0)	554 (10.2)	17.1	535 (16.3)	534 (17.5)	527 (10.4)
Southeast	17.7	532 (9.2)	521 (6.7)	530 (5.2)	20.0	513 (7.4)	510 (10.2)	503 (6.9)
Central . .	38.2	571 (7.5)	545 (7.6)	572 (3.4)	24.6	557 (10.0)	562 (12.2)	547 (4.7)
West . . .	23.1	559 (6.0)	538 (5.0)	552 (4.2)	38.4	535 (8.7)	533 (13.9)	525 (8.3)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-21. Class mean reading proficiency scores, with 95 percent confidence intervals, by region: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Parental Cooperation

Parental cooperation was determined from a single question on the School Questionnaire, "What is the degree of parent cooperation with the school in terms of support for the school's educational principles or goals (compared with other schools you know)?"

Data from the Reading Literacy Study (Table 13-51 and Figure 13-22) are consistent with other research findings suggesting that parental cooperation is related to student achievement on a schoolwide basis (e.g., Keith et al. 1986; Phillips, Smith, and Witte 1985; Wagenaar 1977). Most studies have found a positive relationship between parental involvement and children's academic success. The data show that the highest degree of parental cooperation with the school is associated with the highest achievement levels. However, as with many other factors associated with school effectiveness, parental cooperation is also related to community SES factors and parents' educational level. This reflects the strong interrelationship between schools and communities, showing that a school is "more than interrelated internal elements. It affects and is affected by the outside world, especially parents and the community" (Davis 1989).

According to Lee, Bryk, and Smith (1993), parental involvement has three broad aspects: 1) aiding children's learning with help at home; 2) forming a functional community around school between school staff and parents; and 3) including parents in decisionmaking.

Studies have found strong relationships between parents' educational levels and children's home experiences. Parental expectations and the level of importance placed on education is strongly related to academic performance. The more active parents are in planning children's education, the better the educational progress of children. According to Chubb (1988), "Schools in which parents are highly involved, cooperative and well-informed are more likely to develop effective organizations than schools in which parents do not possess these qualities."

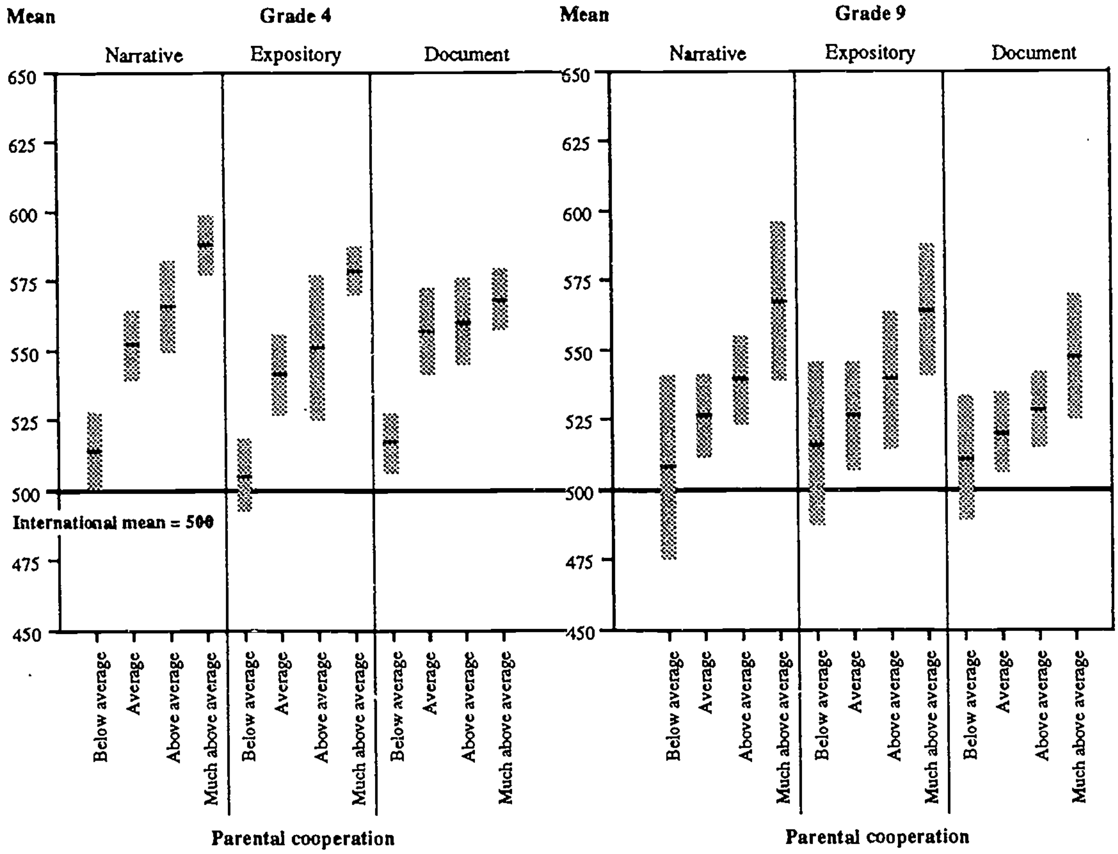
Table 13-51. Class mean reading proficiency scores, by parental cooperation: Grades 4 and 9

Parental cooperation	Grade 4				Grade 9			
	Percent	Narrative	Expository	Document	Percent	Narrative	Expository	Document
Below average	10.3	514 (6.9)	505 (6.3)	517 (5.3)	7.0	508 (16.1)	516 (14.6)	511 (11.0)
Average	32.9	552 (6.3)	541 (7.2)	557 (7.8)	32.0	526 (7.3)	526 (9.5)	520 (6.9)
Above average ...	37.7	566 (8.4)	551 (13.0)	560 (7.9)	49.3	539 (7.9)	539 (12.2)	528 (6.6)
Much above average	19.1	588 (5.1)	579 (4.3)	568 (5.3)	11.5	567 (14.2)	564 (11.8)	547 (11.3)

NOTE: Numbers in parentheses are standard errors. Percentages may not add to 100 due to rounding.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 13-22. Class mean reading proficiency scores, with 95 percent confidence intervals, by parental cooperation: Grades 4 and 9



NOTE: Shaded bands indicate the 95 confidence interval for the corresponding mean. Each confidence interval is constructed as the mean, plus and minus twice the standard error.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

A shortcoming of the Reading Literacy Study data on parental cooperation is that the type of cooperation or specific parental activities are not explored. In fact, it is possible that the type of involvement may be most important. In Brookover and Lezotte's (1979) case study of eight elementary schools in Michigan, schools with increasing rather than decreasing scores included less overall parent involvement but more parent-initiated involvement. While direct involvement in the child's education may not make a difference, schools with numerous daily communications with parents had higher achievers. The students in the best schools were most likely to have parents who encourage their children to learn, monitor their homework, and maintain higher expectations (Chubb, p. 109). These parental actions in the home may also be related to cooperation with school and school activities.

13.4. Concluding Thoughts

This chapter serves three purposes. First, it serves as a review of the available findings and, as such, allows a comparison of the observed relationship between well-established predictors of reading achievement and that outcome according to the findings of the Reading Literacy Study. Second, it provides an opportunity to consider those findings in light of the larger context of the research literature. Finally, it delineates the criteria for the selection of variables to be included in the model of reading that will be developed in Chapter 14.

We grouped the available variables into categories consistent with the literature, looking first at standard demographic variables used to describe students. Our data were consistent with the research literature and other surveys such as NAEP and NELS:88. Minority students tended to have lower reading achievement than whites. The oldest students in the grade, who were most likely to have been previously retained, also tended to have the lowest scores. And students who had exposure to languages other than English tended to do less well than those who spoke only English. But because these attributes are known to be strongly associated with family attributes, we also considered family demographic variables.

As would be expected, we found that children in conventional nuclear families, where parents were well-educated and were in the higher quartiles of wealth, were most likely to achieve at higher levels. However, as indicated in the research literature, there is strong reason to believe that the differences in social status are carried through differences in the aspirations of parents, the types of interactions that occur in the family, and the interaction between the family and the school. Consequently, we turned our attention to intervening variables associated with the family environment.

For the most part, the relationships between the family environment variables included in the Reading Literacy Study and reading achievement also looked much as would be expected. The more literacy resources available within the home, the more likely it was that the child would attain higher levels of reading proficiency. Children who watch more than 6 hours of television a day had significantly lower mean proficiency. There was, however, one seemingly counterintuitive finding: the more parents read with their 9- or 14-year-old child, the lower the child's achievement was likely to be. In considering this finding, we looked closely at the research literature on parent/child interactions. As our questions dealt specifically with reading aloud, we focused our attention on that. The research in this area centered on interactions between parents and preschoolers. Consequently, one might reasonably conclude that if parents are still engaging in this type of activity, the children in question may, in fact, be having difficulties with reading. Perhaps these parents were more likely to get involved in their children's education because their children are low achievers.

In turning our attention to what might be happening within the school, we looked first at the teacher. Again, as expected in cross-sectional data, we found no significant associations between teacher demographic variables, including measures of experience and training, and reading achievement. But intuition and the research literature might lead one to believe that even very weak relationships and small effects, when taken together, can add up to significant achievement differences. However, at this point, we considered each variable independent of all others.

We looked next at the data related to instruction. Because of the abundance of data, we had to determine how best to organize the 190 items into meaningful units. We pursued this task by conducting exploratory factor analyses on item clusters and looking for associations between empirically defined factors and reading theory. While the questionnaires had not explicitly been designed to tap specific positions on reading theories, what emerged in the U.S. was a number of factors that could be loosely associated with particular perspectives. For example, it was apparent that teachers tended to disagree with statements that might be associated with a very highly structured hierarchical approach to reading instruction. Yet they frequently use materials that would be representative of that view. They seemed to strongly agree with statements that might be consistent with elements of a whole language approach to reading instruction, but more frequently reported using teacher behaviors that were much more directive than that approach would advocate. Although three different factors emerged from the items on assessment, teachers tended to test everything frequently.

Given the nature of the study design, it was not surprising that we were unable to detect any association between particular instructional stances and reading achievement. Children in the fourth and ninth grades have had many different opportunities to interact with text and have been exposed to many different teachers and approaches. Consequently, there is no way to measure the comparative effectiveness of a particular approach in this data set. Despite this limitation, the resulting description of instructional practice does provide an interesting reflection of what is going on in American classrooms.

When we looked at class and school configurations, we found no discernable association between class size and achievement as is consistent with research using data of this kind and simple cross-tabulations between selected variables and achievement. Similarly, we found no significant difference between public and private schools. Instructional time, however, seemed to make a difference. Students who received more than 30 hours of instruction per week outperformed those who received less than 25 hours per week. In considering the allocation of resources, we noted that the proportion of specialist teachers in elementary schools was not associated with differences in achievement. However, in secondary schools low achieving students were more likely to be in schools where more than 25 percent of the teachers were specialists compared to schools where specialists accounted for less than 10 percent of the faculty.

While the literature argues for principals who are well trained and strong instructional leaders, our findings indicated no association between these attributes and achievement. Given that we have no information regarding principals' assignment to particular schools, there are no conclusions to be drawn.

In considering the school within the community, our data were consistent with the literature in suggesting that parental cooperation is related to student achievement on a schoolwide basis. But to understand this phenomenon, one must go beyond this particular association and look at the impact of other attributes of the community as well. The research literature pointed toward community SES factors, parental education, and the integration between the school and the community.

What becomes overwhelmingly clear throughout this chapter is that analyses of data, variable by variable, do not capture the full richness of what is creating differences in reading proficiency across the population. Although data presentations of this kind are common and do describe certain associations, they do not enhance our understanding of the relative importance of each variable in producing the desired outcome. In addition, the variables, while indicating an associated effect, may not be describing what should be done to produce that effect. In the next chapter, we develop a model that allows us to unravel why students come to differ in their ability to extract meaning from text.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: Massachusetts Institute of Technology.

Alexander, K., and Pallas, A. (1985). School sector and cognitive performance: When is a little a little? *Sociology of Education,* 58 (April).

Allen, R.L., and Chaffee, S.H. (1977). Racial differences in family communication patterns. *Journalism Quarterly,* 54 (1), 8-13, 57.

Alsalam, N., Fischer, G.E., Ogle, L.T., Rogers, G.T., and Smith, T.M. (1993). *The condition of education.* (U.S. Department of Education, National Center for Education Statistics) Washington, DC: U.S. Government Printing Office.

Anderson, R.C., and Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson (ed.), *Handbook of reading research.* Vol. 1, 255-291. New York: Longman.

Anderson, R.C., Hiebert, E.H., Scott, J.A., and Wilkinson, I.A.G. (1985). *Becoming a nation of readers: The report of the Commission on Reading.* Champaign, IL: Center for the Study of Reading.

Anderson, R.C., Wilson, P.T., and Fielding, L.G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly,* 23, 285-303, 389-413.

Applebee, A.N. (1981). *Writing in the secondary school: English and the content areas.* Urbana, IL: National Council of Teachers of English.

Applebee, A.N., and Langer, J.A. (1983). Instructional scaffolding: Reading and writing as natural language abilities. *Language Arts,* 60, 168-175.

Armor, D.J., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., and Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools.* Santa Monica, CA: Rand Corp.

Ashley, L.F. (1970). Children's reading interests and individualized reading. *Elementary English,* 47, 1088-1096.

Ashov, E.N., and Fishbach, T.J. (1973). An investigation of primary pupils' attitudes toward reading. *Journal of Experimental Education,* 41, 1-7.

August, D., and Garcia, E. (1988). *Language minority education in the United States: Research, policy and practice.* Springfield, IL: Charles C. Thomas.

Austin, G.R., and Garber, H. (eds.). (1985). *Research on exemplary schools.* New York: Academic Press.

Baden, B., and Maehr, M. (1986). Conforming culture with culture: A perspective for designing schools for children of diverse sociocultural backgrounds. In R. Feldman (ed.), *The social psychology of education,* 289-309. Cambridge, England: Cambridge University Press.

Bain, B., and Yu, A. (1980). Cognitive consequences of raising children bilingually: One parent, one language. *Canadian Journal of Psychology,* 34, 304-313;

Baker K., and deKanter, A. (1983). An answer from research on bilingual education. *American Education,* 19, 48-88.

Baker, K. (1990). Bilingual education's 20 year failure to provide rights protection for language minority students. In A. Barona and E. Garcia (eds.), *Children at risk: Poverty, minority status and other issues in educational equity,* 29-52. Washington DC: National Association of School Psychologists.

Balmuth, M. (1982). *The roots of phonics.* New York: Teachers College Press.

Barr, R., and Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Barrett, T.A. (1968). A taxonomy of cognitive and affective dimensions of reading comprehension. Outlined by Clymer, T.C. What is reading? Some current concepts. In H.M. Robinson (ed.), *Innovation and change in reading instruction*, 252-258. NSSE 67th Yearbook, Part II. Chicago: University of Chicago Press.

Bateman, B. (1979). Teaching reading to learning disabled and other hard to teach children. In L.A. Resnick and P.A. Weaver (eds.), *Theory and practice of early reading*. Vol. 1, 227-259. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baumrind, D. (1973). The development of instrumental competence through socialization. In A. Pick (ed.), *Child psychology*. Minnesota Symposium on Child Psychology, Vol. 7, 3-46, Hillsdale, NJ: Lawrence Erlbaum Associates.

Baumrind, D. (1983). Rejoinder to Lewis' reinterpretation of parental firm control effects: Are authoritative families really harmonious? *Psychological Bulletin*, 94, 132-142.

Beach, R., and Hynds, S. (1991). Research on response to literature. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 453-489. New York: Longman.

Beentjes, J.W.J., and Van der Voort, T.H.A. (1988). Television's impact on children's reading skills: A review of the research. *Reading Research Quarterly*, 23, 389-413.

Berliner, D.C. (1984). The half-full glass: A review of research on teaching. In P.L. Hosford (ed.), *Using what we know about teaching*. Reston, VA: Virginia Association for Supervision and Curriculum Development.

Bernstein, B. (1971). A sociolinguistic approach to socialization with some reference to educability. In B. Bernstein (ed.), *Class, codes and control: Theoretical studies towards a sociology of language*, (146-171). London: Routledge and Kegan Paul.

Berry, M.T., Fischer, S.L., Parker, F.S., and Zwier, M.D. (1971). Average and superior readers in lab school. *The Reading Teacher*, 25, 271-275.

Bickel, W.E. (1983). Effective schools: Knowledge, dissemination, inquiry. *Educational Researcher*, 12 (4), 3-5.

Bidwell, C.E. (1965). The school as a formal organization. In J.G. March (ed.), *Handbook of organizations*. Chicago: Rand McNally.

Bloom, B. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership*, 41 (8), 4-17.

Blumberg, A., and Greenfield, W.D. (1986). *The effective principal*. 2nd ed. Newton, MA: Allyn and Bacon.

Boehm, A., and Slater, B. (1974). *Cognitive skills assessment battery*. New York: Teachers College Press.

Boger, R.P., Richter, R.A., and Paolucci, B. (1986). Parent as teacher: What do we know? In R.J. Griffore and R.P. Boger (eds.), *Child rearing in the home and school*. New York: Plenum.

Boggs, S.T. (1972). The meaning of questions and narratives to Hawaiian children. In C. Cazden, V. John, and D. Hymes (eds.), *Foundations of language in the classroom*. New York: Teachers College Press.

Bond, G.L., and Dykstra, R. (1967). The cooperative research program in first grade reading instruction. *Reading Research Quarterly*, 2, 5-142.

Bossert, S., Dwyer, D., Rowan, B., and Lee, G. (1982). The instructional management role of the principal. *Education Administration Quarterly*, 18, 34-64.

Boyd, D.A. (1984). Children from divorced and reconstituted families: An examination of academic achievement in light of familial configuration. Ph.D. diss., Kansas State University: Dissertation Abstracts International 46/01B.

Boykin, A. (1986). The triple quandary and the schooling of Afro-American children. In U. Neisser (ed.), *The school achievement of minority children*, 57-92. New York: New Perspectives.

Bridges, E.M. (1982). Research on the school administrator: The state of the art, 1967-1980. *Education Administration Quarterly*, 18, 12-33.

Bridges, E.M. (1986). *The incompetent teacher: The challenge and the response*. Philadelphia: The Falmer Press.

Briggs, C., and Elkind, D. (1977). Characteristics of early readers. *Perceptual and Motor Skills*, 44, 1231-1237.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brook, J.S., Whiteman, M., Peisach, E., and Deutsch, M. (1974). Aspiration levels of and for children: Age, sex, race and socioeconomic correlates. *Journal of Genetic Psychology*, 124, 3-16.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., and Wisenbaker, J. (1979). *School social systems and student achievement: Schools can make a difference*. Brooklyn, NY: Praeger Publishers.

Brookover, W.B., and Lezotte, L.W. (1979). *Changes in school characteristics coincident with changes in student achievement*. ERIC, ED 181 005.

Brophy, J., and Good, T.L. (1986). Teacher behavior and student achievement. In M. Wittrock (ed.), *Handbook of research on teaching*, 3rd ed. New York: Macmillian.

Cahape, P., and Howley, C.B. (1992). *Indian nations at risk: Listening to the people*. Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

California State Department of Education. (1980). *Report on the special studies of selected ECE schools with increasing and decreasing reading scores*. Sacramento, CA: Office of Program Evaluation and Research.

Canale, M. (1983). From communicative complete to communicative pedagogy. In J. Richards and R. Schmidt (eds.), *Language and communication*. London: Langman.

Carnegie Forum on Education and the Economy (1986). *A nation prepared: Teachers for the 21st century*. The report of the task force on teaching as a profession. New York: Carnegie Corporation.

Cavin, E.S., Mumane, R.J., and Brown, R.S. (1985). School district response to enrollment changes: The direction of change matters! *Journal of Education Finance*, 10 (4), 426-440.

Cazden, C.B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chaffee, S.H., McLeod, J.M., and Atkin, C.K. (1971). Parental influences on adolescent media use. *American Behavioral Scientist*, 14, 323-340.

Chall, J.S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.

Chall, J.S. (1983). *Stages of reading development*. New York: McGraw-Hill.

Chiu, L. (1984). Children's attitudes toward reading and reading interest. *Perceptual and Motor Skills*, 58 (3), 960-962.

Choy, S.P., Bobbit, S.A., Henke, R.R., Medrich, E.A., Horn, L.J., and Lieberman, J. (1993). *America's teachers: Profile of a profession*. (U.S. Department of Education, National Center for Education Statistics). NCES 93-025. Washington, DC: U.S. Government Printing Office.

Choy, S.P., Henke, R.R., Alt, M.N., Medrich, E.A., and Bobbit, S.A. (1993). *Schools and staffing in the United States: A statistical profile, 1990-91*. (U.S. Department of Education, National Center for Education Statistics). NCES 93-146. Washington, DC: U.S. Government Printing Office.

Chubb, J.E. (1988). Why the current wave of school reform will fail. *The Public Interest, 90*, 28-49.

Chubb, J.E., and Moc, T.M. (1988). Politics, markets, and the organization of schools. Unpublished manuscript.

Chubb, J.E., and Moe, T.M. (1990). *Politics, markets, and America's schools.* Washington, DC: Brookings Institution.

Clark, D.L., Lotto, L.S., and Astuto, T.A. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly*, 20 (3), 41-68.

Clark, M.M. (1976). *Young fluent readers: What can they teach us?* London: Heinemann.

Clark, R.M. (1983). *Family life and school achievement: Why poor black children succeed or fail.* Chicago: University of Chicago Press.

Clay, M.M. (1985). *The early detection of reading difficulties.* 3rd ed. Portsmouth, NH: Heinemann.

Cline, R.K., and Kretke, G.L. (1980). An evaluation of long-term sustained silent reading in the junior high school. *Journal of Reading*, 23, 503-506.

Cohen, M. (1983). Instructional management and social conditions in effective schools. In A. Odden and L. D. Webb (eds.), *School finance and school improvement: Linkages for the 1980's*, 17-50. Cambridge, MA: Ballinger Publishing Co.

Coleman, J.S., and Hoffer, T. (1987). *Public and private high schools: The impact of communities.* New York: Basic Books.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., and York, R.L. (1966). *Equality of educational opportunity* (Department of Health, Education and Welfare). Washington, DC: U.S. Government Printing Office.

Coleman, J.S., Hoffer, T., and Kilgore, S. (1981). *School and beyond: A national longitudinal study for the 1980's, public and private schools.* Report prepared for National Center for Education Statistics. Chicago: National Opinion Research Center.

Collins, C. (1980). Sustained silent reading periods: Effect on teachers' behaviors and students' achievement. *Elementary School Journal*, 81, 108-114.

Collins, L. (1981). A study of locus of control of children from intact, single-parent, and reconstituted families. Ph.D. diss., Georgia State University: Dissertation Abstract International 42, 3030B.

Council of Chief State School Officers (1990). *School success for limited English-proficient students: The challenges and state response.* Washington, DC: Council of Chief State School Officers.

Cousert, G. C. (1978). Six selected home reading environment factors and their relationship to reading achievement at the third grade. Ph.D. diss., Indiana University, Bloomington.

Crawford, J. (1983). A study of instructional processes in Title I classes: 1981-82. *Journal of Research and Evaluation of the Oklahoma City Public Schools*, 13 (1), 1-25.

Cullinan, B.E. (1989). Literature for young children. In D.S. Strickland and L.M. Morrow (eds.), *Emerging literacy: Young children learn to read and write*, 35-51. Newark, DE: International Reading Association.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (eds.), *Schooling and language minority students: A theoretical framework.* Los Angeles: California State University.

Cummins, J. (1984). Linguistic interdependence among Japanese and Vietnamese immigrant children. In C. Rivera (ed.), *Communicative competence approaches to language proficiency assessment: Research and applications.* Avon, England: Multilingual Matters.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Education Review, 56, 18-36.*

Cusik, P.A. (1983). *The egalitarian ideal and the American high school.* New York: Longman.

Davis, F.B. (1944). Fundamental factors of comprehension in reading. *Psychometrika*, 9, 185-197.

Davis, G.A., and Thomas, M.A. (1989). *Effective schools and effective teachers.* Boston: Allyn & Bacon.

Delgado-Gaitan, C. (1987). Traditions and transitions in the learning process of Mexican children: An ethnographic view. In G. Spindler and L. Spindler (eds.), *Interpretative ethnography of education: At home and abroad.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Denham, C., and Lieberman, A. (eds.). (1980). *Time to learn.* Washington, DC: National Institute of Education.

Dix, M. (1976). Are reading habits of parents related to reading performance of their children? Paper presented at the annual meeting of the National Council of Teachers of English, Chicago, IL.

Dolch, E., and Bloomster, M. (1937). Phonic readiness. *Elementary School Journal,* 38, 201-205.

Dornbusch, S.M., and Wood, K.D. (1989). Family process and educational achievement. In W.J. Weston (ed.), *Education and the American family: A research synthesis.* New York: New York University Press.

Dornbusch, S.M., Carlsmith, J.M., Bushwall, S.J., Ritter, P.L., Leiderman, P.H., Hastorf, A.H., and Gross, R.T. (1985). Single parents, extended households, and the control of adolescents. *Child Development,* 56, 326-341.

Dornbusch, S.M., Ritter, P.L., Leiderman, P.H., Roberts, D.F., and Fraleigh, M.J. (1987). The relation of parenting style to adolescent school performance. *Child Development,* 58, 1244-1257.

Doss, D., and Holley, F. (1982). *A cause for national pause: Title I schoolwide projects.* ORE Publication No. 81.55. Austin, TX: Office of Research and Evaluation, Austin Independent School District.

Dreeben, R. (1970). *The nature of teaching.* Glenview, IL: Scott Foresman and Co.

Dreger, R.M., and Miller, K.S. (1968). Comparative psychological studies of Negroes and Whites in the US: 1959-1965. *Psychological Bulletin,* 70 (3, Pt. 2), 1-58.

Durkin, D. (1966). *Children who read early.* New York: Teachers College Press.

Durkin, D. (1974-75). A six-year study of children who learned to read in school at the age of four. *Reading Research Quarterly,* 10, 9-61.

Durkin, D. (1988). *Teaching them to read.* 5th ed. Boston: Allyn & Bacon.

Durkin, D. (1978-79). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly,* 14, 481-553.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership,* 37, 15-24.

Edmonds, R. R. (1981). Making public schools effective. *School Policy.* 12, (2), 56-60.

Edwards, P.A. (1992). Supporting lower-SES mothers' attempts to provide scaffolding for bookreading. In J.B. Allen and J. Mason (eds.), *Reading the risks for young learners: Literacy practices and policies.* Portsmouth, NH: Heinemann.

Ehri, L.C. (1987). Learning to read and spell words. *Journal of Reading Behavior,* 19, 5-31.

Elley, W., and Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly,* 19, 53-67.

Epstein, J.L. (1981). Patterns of classroom participation, student attitudes, and achievements. In J.L. Epstein (ed.), *The quality of school life.* Lexington, MA: Lexington Books.

Epstein, J.L. (1985). Homework practices, achievements, and behaviors of elementary school students. Paper presented at the annual meeting of the American Sociological Association, Washington, DC.

Epstein, J.L. (1989). Family structures that promote student motivation and achievement: A developmental perspective. In C. Ames and R. Ames (eds.), *Research on motivation in education*. Vol. 3. Orlando, FL: Academic Press.

Epstein, J.L., and McPartland, J.M. (1977). *Family and school interactions and main effects on affective outcomes*. ERIC, ED 151 713.

Farr, R., et al. (1991). Writing in response to reading. *Educational Leadership*, 66-69.

Fielding, L., Wilson, P., and Anderson, R. (1986). A new focus on free reading: The role of trade books in reading instruction. In T. Raphael and R. Reynolds (eds.), *The context of school-based literacy*, 149-160. New York: Random House.

Finn, J., and Achilles, C. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27 (3), 557-577.

Finn, P. (1980). Developing critical television viewing skills. *Educational Forum*, 44, 473-482.

Fisher, C.W., Berliner, D., Filby, N., Marliave, R., Cahan I., and Dishaw, M. (1980). Teaching behaviors, academic learning time and student achievement. In C. Denham and A. Lieverman (eds.), *Time to learn*. Washington, DC: National Institute of Education.

Flesch, R. (1955). *Why Johnny can't read*. New York: Harper & Row.

Flodin, B., Hedinsson, E., and Roe, K. (1982). *Primary school panel: A descriptive report*. Lunds: Sociologiska Institutionen, Lunds Universitet.

Freshour, F.W. (1972). Beginning reading: Parents can help. *The Reading Teacher*, 25, 513-516.

Friedkin, N.E., and Necochea, J. (1988). School size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis*, 10, 237-249.

Friere, P. (1973). *Education for critical consciousness*. New York: Seabury Press.

Frye, N. (1957). *Anatomy of criticism*. Princeton: Princeton University Press.

Fuller, B., and Izu, J.A. (1986). Explaining school cohesion: What shapes the organizational beliefs of teachers? *American Journal of Education*, 94, 501-535.

Gage, N.L. (1985). *Hard gains in the soft sciences*. Bloomington, IN: Phi Delta Kappa.

Gage, N.L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.

Galambos, J., and Hakuta, K. (1988). Subject specific and task specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics*, 9, 141-162.

Gallimore, R., and Au, K.H. (1979). The competence/incompetence paradox in the education of minority culture children. *Quarterly Newsletter of the Laboratory for Comparative Human Cognition*, 11, 32-37.

Ganong, L., and Coleman, M. (1984). The effects of remarriage on children: A review of the empirical literature. *Family Relations*, 33, 425-432.

Garcia, E. (1989). Instructional discourse in "effective" Hispanic classrooms. In R. Jacobson and C. Faltis (eds.), *Language distribution issues in bilingual schooling*, 104-120. Clevedon, England: Multilingual Matters, Ltd.

Garcia, E. (1991). *The education of linguistically and culturally diverse students: Effective instructional practices*. Santa Cruz, CA: Center for Research on Cultural Diversity and Second Language Learning.

Garcia, E. (1992). Effective instruction for language minority students: The teacher. *Journal of Education*, 173, 130-141.

Garcia, E. (1993). Language, culture, and education. In L.D. Hammond (ed.), *Review of Research in Education*, 19, 51-98. Washington, DC: American Educational Research Association.

Garcia, E. (1993). Language, culture, and education. *Review of Research in Education*, 19.

Gardner, R.C., and Lambert, E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Genesee, F., Polich, E., and Stanley, M.H. (1977). An experimental French immersion program at the secondary school level--1969 to 1974. *The Canadian Modern Language Review*, 33, 318-332.

Gersten, R., Carnine, D., and Green, S. (1982). The principal as instructional leader: A second look. *Educational Leadership*, 40 (3): 47-49.

Gibson, E.J., and Levin, H. (1975). *The psychology of reading*. Boston: MIT Press.

Gilman, D., Tillitski, C., Swan, E., and Stone, W. (1987). *PRIME TIME at North Gibson school corporation: A three year study*. Technical report prepared by Professional School Services. Terre Haute, IN: Indiana State University.

Glass, G.V., and Smith, M.L. (1978). *Meta-analysis of the search on the relationship of class size and achievement*. San Francisco, CA: Far West Laboratory of Educational Research and Development.

Glass, G.V., and Smith, M.L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1 (2), 2-16.

Glenn, B.C., and McLean, T. (1981). *What works? An examination of effective schools for poor, black children*. Cambridge, MA: Harvard University, Center for Law and Education.

Goelman, H., Oberg, A., and Smith, F. (eds.) (1984). *Awakening to literacy*. London: Heinemann.

Goldenberg, C., and Gallimore, R. (1991). Local knowledge, research knowledge and educational change: A core study of early Spanish reading improvement. *Educational Researcher*, 20 (8), 2-14.

Goldman, S., and Trueba, H. (eds.), (1987). *Becoming literate in English as a second language: Advances in research and theory*. Norwood, NJ: Ablex.

Goldman, S.R., Reyes, R.M, and Varnhagen, C.K. (1984). Understanding fables in first and second languages. *Journal of the National Association for Bilingual Education*, 3, 35-66.

Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.

Goodman, K., Goodman, Y., and Flores, B. (1989). *Reading in the bilingual classroom: Literacy and biliteracy*. Silver Spring, MD: National Clearinghouse for Bilingual Education.

Goodman, K.S. (1970). Behind the eye: What happens in reading. In K.S. Goodman and O.S. Niles (eds.), *Reading, process and program*. Urbana, IL: National Council of Teachers of English.

Goodman, K.S. (1985). Unity in reading. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 813-840. Newark, DE: International Reading Association.

Goodman, K.S., and Goodman, Y.M. (1979). Learning to read is natural. In L.B. Resnick and P.A. Weaver (eds.), *Theory and practice of early reading*. Vol. 1, 137-154. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goodman, Y.M. (1986). Children coming to know literacy. In W.H. Teale, and E. Sulzby (eds.), *Emergent literacy: Writing and reading*, 1-14. Norwood, NJ: Ablex.

Gopinathan, S. (1978). *A measure of reading: IE survey of reading interests and habits*. Singapore: Institute of Education.

Gorman, T.P., White, J., Orchard, I., and Tate, A. (1983). *Language performance in schools: Secondary survey report No.2*. London: Her Majesty's Stationary's Office.

Gough, P.B. (1972). One second of reading. In J.F. Kavanagh and I.G. Mattingly (eds.), *Language by ear and by eye*, 331-368. Cambridge, MA: MIT Press.

Gough, P.B. (1985). One second of reading: Postscript. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 687-688. Newark, DE: International Reading Association.

Gough, P.B., and Hillinger, M.L. (1980). Learning to read: An unnatural act. *Bulletin of the Orlon Society*, 30, 179-196.

Graham, P.A. (1987). Black teachers: A drastically scarce resource. *Phi Delta Kappan* 68 (8), 598-605.

Grant, G. (1988). *The world we created at Hamilton High*. Cambridge, MA: Harvard University Press.

Gray, E. G. (1980). School performance and perception of maternal behavior of father-absent boys following divorce. Ph.D. diss., Kent State University: Dissertation Abstracts International 41/05-B.

Gray, W.S. (1960). *The major aspects of reading. Sequential development of reading abilities*. Supplementary Educational Monographs, No. 90. Chicago: University of Chicago Press.

Greaney, V. (1980). Factors related to amount and type of leisure time reading. *Reading Research Quarterly*, 15, 337-357.

Greaney, V. (1986). Parental influences on reading. *The Reading Teacher*, 39, 813-818.

Greaney, V., and Hegarty, M. (1984). Correlates of leisure-time reading. Unpublished manuscript. Dublin: Educational Research Center, St. Patricks College.

Greaney, V., and Hegarty, M. (1987). Correlates of leisure-time reading. *Journal of Research in Reading*, 10, 3-20.

Guthrie, J.T., and Greaney, V. (1991). Literacy acts. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 68-96. New York: Longman.

Hafner, A., Ingels, S., Schneider, B., Stevenson, D. (1990). *A profile of the American eighth grader: NELS:88 student descriptive summary*. (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Hakuta, K. (1986). *Mirrors of language: The debate on bilingualism*. New York: Basic Books.

Hakuta, K., and D'Andrea, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high school students. *Applied Linguistics*, 13, 72-99.

Hakuta, K., and Diaz, R. (1984). The relationship between bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K.E. Nelson (ed.), *Children's Language*. Vol. 5. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hakuta, K., and Gould, L.J. (1987). Synthesis of research on bilingual education. *Educational Leadership*, 44 (6), 39-45.

Halliday, M. (1975). *Learning how to mean: Explorations in the development of language*. London: Dover.

Hansen, H.S. (1969). The impact of home literacy environment on reading attitude. *Elementary English*, 46, 17-24.

Hanushek, E.A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management*, 1, 19-41.

Hanushek, E.A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.

Harste, J.C., and Woodward, V.A. (1989). Fostering needed change in early literacy programs. In D.S. Strickland and L.M. Morrow (eds.), *Emerging literacy: Young children learn to read and write*, 147-159. Newark, DE: International Reading Association.

Harste, J.C., Burke, C.L., and Woodward, V.A. (1982). Children's language and world: Initial encounters with print. In J.A. Langer and M.T. Smith-Burke (eds.), *Reader meets the author: Bridging the gap*, 105-131. Newark, DE: International Reading Association.

Harste, J.C., Woodward, V.A., and Burke, C.L. (1984). *Language stories and literacy lessons.* Portsmouth, NH: Heinemann.

Heath, S.B. (1982). What no bedtime story means: Narrative skills at home and school. *Language in Society*, 11, 49-76.

Heath, S.B. (1983). *Way with words: Language, life, and work in communities and classrooms.* Cambridge, MA: Harvard University Press.

Heath, S.B. (1985). Language policies: Patterns of retention and maintenance. In W. Connor (ed.), *Mexican Americans in comparative perspective.* Washington, DC: Urban Institute.

Heath, S.B. (1986). Sociocultural contexts of language development. In *Beyond language: Social and cultural factors in schooling language minority students.* Sacramento: California State Department of Education, Bilingual Education Office.

Heath, S.B. (1991). The sense of being literate. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research.* Vol. 2, 3-25. New York: Longman.

Heath, S.B., and Branscombe, A. (1985). Intelligent writing in an audience community: Teachers, students and researcher. In S.W. Freedman (ed.), *The acquisition of literacy: Revision and response.* Norwood, NJ: Ablex.

Heath, S.B., and Thomas, C. (1984). The achievement of preschool literacy for mother and child. In H. Goelman, A. Oberg, and F. Smith (eds.), *Awakening to literacy*, 51-72. Exeter, NH: Heinemann.

Heather, P. (1981). *Young people's reading. A study of the leisure reading of 13-15 year olds.* Sheffield, England: University of Sheffield, Centre for Research on User Studies.

Herzog, E., and Saudia, C.E. (1973). Children in fatherless families. In B. Caldwell and H. Ricciuti (eds.), *Review of child development research.* Vol. 3. Chicago: University of Chicago Press.

Hess, R.D., and Holloway, S.D. (1984). Family and school as educational institutions. In R.D. Parke (ed.), *Review of child development research.* Vol. 7, 179-222. Chicago: University of Chicago Press.

Hetherington, E.M., Featherman, D.L., and Camara, K.A. (1981). *Intellectual functioning and achievement of children in one-parent households.* Washington DC: National Institute of Education.

Hetherington, E.M., Featherman, D.L., and Camara, K.A. (1982). *Cognitive performance, school behavior, and achievement of children from one-parent households.* Washington, DC: National Institute of Education.

Hett, G.G. (1983). *Canadian children from single parent families: Are they an overlooked minority?* ERIC, 262 349.

Heyns, B. (1978). *Summer learning and the effects of schooling.* New York: Academic Press.

Hill, J.P. (1980). The family. In M. Johnson (ed.), *Toward adolescence, the middle school years*, 32-55. Chicago: University of Chicago Press.

Holdaway, D. (1979). *The foundations of literacy.* Portsmouth, NH: Heinemann.

Horn, L., and West, J. (1992). *A profile of parents of eighth graders.* (U.S. Department of Education, National Center for Education Statistics). NCES 92-488. Washington, DC: U.S. Government Printing Office.

Hoy, W.K., and Ferguson, J. (1985). A theoretical framework and exploration of organizational effectiveness of schools. *Educational Administration Quarterly*, 21 (2), 117-134.

Hudelson, S. (1987). The role of native language literacy in the education of language minority children. *Language Arts*, 64, 827-841.

Hunt, R.A. (1990). The parallel socialization of reading research and literary theory. In S.B. Straw and D. Bogdan (eds.), *Beyond communication: Reading comprehension and criticism*. Portsmouth, NH: Boynton/Cook-Heinemann.

Huus, H. (1979). A new look at children's interests. In J.E. Shapiro (ed.), *Using literature and poetry effectively*, 37-45. Newark, DE: International Reading Association.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania.

Hynds, S. (1990). Reading as a social event: Comprehension and response in the text, classroom, and world. In D. Bogdan and S. Straw (eds.), *Beyond communication, reading comprehension and criticism*. Portsmouth, NH: Boyton/Cook Publishers, Heinemann.

Ingham, J. (1981). *Books and reading development: The Bradford book flood experiment*. London: Heinemann Educational.

Ingham, J. (1982). *Books and reading development: The Bradford book flood experiment*. 2nd ed. Exeter, NH: Heinemann.

James, C. (1987). Book buying for and by children. *Proceedings of the children's market symposium*. London: Publishers Association.

Jencks, C., Smith, M., Acland, H., Bane, M.J., Chen, D., Gintis, H., Heyns, B., and Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Jenkinson, A.J. (1940). *What do boys and girls read?* 2nd ed. London: Methuen.

Johnson-Laird, P. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Johnston, J. (1990). What are teacher perceptions of teaching in different classroom contexts? Project STAR: Tennessee State Department of Education. Paper presented at the meeting of the American Educational Research Association, Boston, MA.

Juel, C (1991). Beginning reading. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 759-788. New York: Longman.

Just, M.A., and Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.

Just, M.A., and Carpenter, P.A. (1985). A theory of reading: From eye fixations to comprehension. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 174-208. Newark, DE: International Reading Association.

Kastler, L.A., Roser, N.L., and Hoffman, J.V. (1987). Understanding of the forms and functions of written language: Insights from children and parents. In J.E. Readance and R.S. Baldwin (eds.), *Research in literacy: Merging perspectives*, 85-92. Rochester, NY: National Reading Conference.

Katzman, M.T. (1971). *The political economy of urban schools*. Cambridge, MA: Harvard University Press.

Keith, T.Z., Reimers, T.M., Fehrmann, P.G., Pottebaum, S.M., and Aubey, L.W. (1986). Parental involvement, homework, and TV time: Direct and indirect effects on high school achievement. *Journal of Educational Psychology*, 78, 373-380.

Kessler, C., and Quinn, M.E. (1985). Positive effects of bilingualism on science problem-solving abilities. In J.E. Alatis and J.J. Staczek (eds.), *Perspectives on bilingual education*, 289-296. Washington, DC: Georgetown University Press.

Kessler, C., and Quinn, M.E. (1987). Language minority children's linguistic and cognitive creativity. *Journal of Multilingual and Multicultural Development*, 8, 173-185.

King, S.H. (1993). The limited presence of African-American teachers. *Review of Educational Research*, 63 (2), 115-149.

Kintsch, W., and van Dijk, T. (1978). Toward a model of text comprehension. *Psychological Review*, 85, 363-394.

Kirsch, I.S., and Guthrie, J.T. (1984). Prose comprehension and text search as a function of reading volume. *Reading Research Quarterly*, 19, 331-342.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the national adult literacy survey* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socializations. In D.A. Goslin (ed.), *Handbook of socializations theory and research*, 347-480. Chicago: Rand McNally.

Kominski, R., and Adams, A. (1992). *School enrollment--social and economic characteristics of students: October 1990*. Current Population Reports. (U.S. Department of Commerce, Bureau of the Census). Series P-20, No. 460. Washington, DC: U.S. Government Printing Office.

Kontos, S. (1986). What preschool children know about reading and how they learn it. *Young Children*, 42, 58-66.

Krashen, S.D. (1988). Do we learn to read by reading? The relationship between free reading and reading ability. In D. Tannen (ed.), *Linguistics in context: Connecting observation and understanding* (Lectures from the 1985 LSA/TESOL and NEH Institutes), 269-298. Norwood, NJ: Ablex.

Krus, P.H., and Rubin, R.A. (1974). Use of family history data to predict educational functioning from ages five through seven. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

LaBerge, D., and Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

LaBerge, D., and Samuels, S.J. (1985). Toward a theory of automatic information processing in reading: Updated. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 719-721. Newark, DE: International Reading Association.

Lambert, W.E., and Tucker, G.R. (1972). *Bilingual education of children: The St. Lambert experiment*. Rowley, Mass: Newbury House.

Lamme, L., and Olmstead, O. (1977). Family reading habits and children's progress in reading. Paper presented at the annual meeting of the International Reading Association, Miami Beach, FL.

Langer, J.A., Applebee, A.N., Mullis, I.V., and Foertsch, M.A. (1990). *Learning to read in our nation's schools: Instruction and achievement in 1988 at grades 4, 8, and 12*. The National Assessment of Educational Progress. Princeton, NJ: Educational Testing Service.

Laosa, L. (1983). The relationship of mother-child interaction and educational characteristics of the mother. *Journal of Educational Psychology*, 27, 146-172.

Lee, C.S. (1984). The contribution of selected home environmental factors to the reading achievement of school aged children. Ph.D. diss., Fordham University, New York.

Lee, V.E., and Bryk, A.S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education*, 61, 78-94.

Lee, V.E., and Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172-192.

Lee, V.E., Bryk, A.S., and Smith, J.B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (ed.), *Review of Research in Education*. Vol. 19, 171-267. Washington, DC: American Educational Research Association.

Lee, V.E., Dedrick, R.F., and Smith, J.B. (1991). The effect of the social organization of schools on teachers' self-efficacy and satisfaction. *Sociology of Education*, 64, 190-208.

Lee, V.E., and Ekstrom, R.B. (1987). Student access to guidance counseling in high school. *American Educational Research Journal*, 24, 287-310.

Leichter, H.J. (1974). Some perspectives on the family as educator. *Teachers College Record*, 76, 175-217.

Leithwood, K.A. (1990). The principal's role in teacher development. In B. Joyce (ed.) *Changing school culture through staff development*, 71-89. Alexandria, VA: Association for Supervision and Curriculum Development.

Lepper, M.R., and Greene, D. (1978). *The hidden costs of reward: New perspectives on the psychology of motivation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lepper, M.R., Greene, D., and Nisbett, R.E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "over-justification" hypothesis. *Journal of Personality and Social Psychology*, 28, 129-137.

Levin, I. (1988). *Accelerated schools for at-risk students*. CPRE Research Report Series RR-010. New Brunswick, NJ: Rutgers University Center for Policy Research in Education.

Levine, D.U. (1992). An interpretive review of U.S. research and practice dealing with unusually effective schools. In D. Reynolds and P. Cuttance (eds.), *School effectiveness: Research, policy and practice*. New York: Cassell.

Lewin, L. (1992). Integrating reading and writing strategies using an alternating teacher-led/student-selected instructional pattern. *The Reading Teacher*, 45 (8), 586-591.

Lightfoot, S.L. (1983). *The good high school: Portraits of character and culture*. New York: Basic Books.

Lipsitz, J. (1983). *Successful schools for young adolescents*. New Brunswick, NJ: Transaction Press.

Little, J.W. (1982). The effect of principal. *American Education*, 19, 38-43.

Loban, W. (1963). *The language of elementary school children*. Urbana, IL: National Council of Teachers of English.

Long, B.H., and Henderson, E.H. (1973). Children's use of time: Some personal and social correlates. *The Elementary School Journal*, 73, 193-199.

Lortie, D.C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.

Lundberg, I., and Linnakyla, P. (1992). *Teaching reading around the world*. The Hague, The Netherlands: IEA.

Lyons, C.A. (1991). A comparative study of the teaching effectiveness of teachers participating in a year-long or two-week inservice program. In J. Zutell and S. McCormick (eds.), *Learner factors/teacher factors: Issues in literacy research and instruction*. Fortieth Yearbook of National Reading Conference. Chicago: National Reading Conference.

Maccoby, E.E., and Martin, J.M. (1983). Socialization in the context of the family: Parent-child interaction. In E.M. Hetherington (ed.), *Handbook of child psychology*. Vol. 4, *Socialization, personality, and social development*, edited by P.H. Mussen, 1-101. New York: John Wiley.

Mackenzie, D.E. (1983). Research for school improvement: An appraisal of some recent trends. *Educational Researcher*, 12 (4), 5-17.

Mandeville, G.K., and Anderson, L.W. (1987). The stability of school effectiveness indices across grade levels and subject areas. *Journal of Educational Measurement*, 24 (3), 203-217.

Manning, G.L., and Manning, M. (1984). What models of recreational reading make a difference? *Reading World*, 23, 375-380.

Marjoribanks, K. (1983). The evaluation of a family learning environment model. *Studies in Educational Evaluation*, 9, 343-351.

Mason, J.M. (1980). When do children begin to read: An exploration of four-year-old children's letter and word reading competencies. *Reading Research Quarterly*, 15, 203-227.

McClelland, D.C. (1961). *The achieving society*. New York: The Free Press.

McCracken, R.A. (1971). Initiating sustained silent reading. *Journal of Reading*, 14, 521-524, 582-583.

McGinley, W., and Madigan, D. (1990). The research story: A form for integrating reading, writing and learning. *Language Arts*. 67, 474-483.

McGiverin, Glass, C.V., and Tillitski, C. (1981). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal*, 90 (1), 47-56.

McLaughlin, B. (1984). *Second language acquisition in childhood*. Vol. 1. *Preschool children*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McLaughlin, B. (1985). *Second language acquisition in childhood*. Vol. 2. *School age children*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McLaughlin, M.W., and Marsh, D.D. (1978). Staff development and school change. *Teacher College Record*, 80 (1), 69-94.

McLeary, L.E., and Thompson, S.D. (1979). *The senior high school principalship*. Vol. 3. *Summary report*. Reston, VA: National Association of Secondary School Principals.

McNamara, T.P., Miller, J.D., and Bransford, D.L. (1991). Mental models and reading comprehension. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 490-511. New York: Longman.

Meisel, S., and Glass, G.G. (1970). Voluntary reading interests and the interest content of basal readers. *Reading Teacher*, 23, 655-659.

Meyer, J.W., and Rowan, B. (1983). The structure of educational organization. In J.W. Meyer and W.R. Scott (eds.), *Organizational environments: Ritual and rationality*, 71-98. Beverly Hills, CA: Sage.

Milne, A.M. (1989). Family structure and the achievement of children. In W. Weston (ed.), *Education and the American family*. New York: New York University Press.

Mitchell, D.A., Beach, S.A., and Badarak, G. (1991). *Modeling the relationship between achievement and class size: A re-analysis of the Tennessee Project STAR data*. Tennessee State Department of Education.

Moffet, J., and Wagner, B.J. (1983). *Student-centered language arts and reading K-13*. 3rd ed. Boston: Houghton Mifflin.

Moffett, J. (1983). *Teaching the universe of discourse*. 2nd ed. Boston: Houghton Mifflin.

Morphett, M.V., and Washburne, C. (1931). When should children begin to read? *Elementary School Journal*, 31, 495-503.

Morris, V.C., Crowson, R., Horowitz, E., and Porter-Gehric, C. (1986). The urban principal: Middle manager in the educational bureaucracy. *Phi Delta Kappan*, 63 (1), 689-692.

Morrow, L.M. (1983). Home and school correlates of early interest in literature. *Journal of Educational Research*, 76 (4), 221-230.

Morrow, L.M. (1985). Developing young voluntary readers: The home-the child-the school. *Reading Research and Instruction*, 25, 1-8.

Morrow, L.M. (1985). *Promoting voluntary reading in school and home*. Fastback 225. Bloomington, IN: Phi Delta Kappa.

Morrow, L.M., and Weinstein, C.S. (1986). Encouraging voluntary reading. The impact of a literature program on children's use of library corners. *Reading Research Quarterly*, 21, 330-346.

Mullin, S.P., and Summers, A.A. (1981). Is more better? A review of the evidence on the effectiveness of spending on compensatory education. Unpublished manuscript. University of Pennsylvania.

Mullis, I.V.S., and Jenkins, L.B. (1990). *The reading report card, 1971-88: Trends from the nation's report card*. The National Assessment of Educational Progress, Princeton NJ: Education Testing Service.

Mullis, I.V.S., Campbell, J.R., and Farstrup, A.E. (1993). *NAEP 1992 reading report card for the nation and the states: Data from the national and trial state assessments*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Mullis, I.V.S., Dossey, J.A., Foertsch, M.A., Jones, L.R.,and Gentile, C.A. (1991). *Trends in academic progress*. (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Murnane, R.J. (1975). Teacher mobility revisited. *Journal of Human Resources*, 16 (1), 3-19.

Murnane, R.J. (1980). Interpreting the evidence on school effectiveness. Working paper #830. New Haven, CT: Institution for Social and Policy Studies, Yale University.

Murphy, J. (1988). Methodological, measurement, and the conceptual problems in the study of instructional leadership. *Educational Evaluation and Policy Analysis*, 10 (2), 117-139.

Napoli, J. (1968.) Environmental factors and reading ability. *The Reading Teacher*, 21, 552-607.

National Education Association. (1991). *Status of the American public school teacher, 1990-91*. Washington, DC: National Education Association.

Natriello, G. (ed.) (1986). *School dropouts: Patterns and policies*. New York: Teachers College Press.

Neuman, S.B. (1986). Television, reading and the home environment. *Reading Research and Instruction*, 25, 173-183.

Neuman, S.B. (1986). The home environment and fifth grade students' leisure reading. *Elementary School Journal*, 83, 333-343.

New York Office of Education Performance Review. (1974). *School factors influencing reading achievement: A case study of two inner city schools*. ERIC, 089 211.

Newmann, F.M. (1981). Reducing student alienation in high schools: Implications of theory. *Harvard Educational Review*, 51, 546-564.

Newmann, F.M., and Oliver, D.W. (1967). Education and community. *Harvard Educational Review*, 37, 61-106.

Newmann, F.M., Rutter, R.A., and Smith, M.S. (1989). Organization factors affecting school sense of efficacy, community, and expectations. *Sociology of Education*, 62, 221-38.

Newson, J., and Newson, E. (1977). *Perspectives on school at seven years old*. London: Allen and Unwin.

Nguyen, T. (1984). Positive self-concept in the Vietnamese bilingual child. *Bilingual Journal*, 8, 9-14.

Nine Curt, C.J. (1984). *Nonverbal communication*. Cambridge, MA: Evaluation, Dissemination and Assessment Center.

Ninio, A. (1980). Picture-book reading in mother-infant dyads belonging to two subgroups in Israel. *Child Development*, 51, 587-590.

Norris, S.P., and Phillips, L.M. (1987). Explanations of reading comprehension: Schema theory and critical thinking theory. *Teachers College Record*, 89, 281-306.

Norvell, G.W. (1966). The challenge of periodicals in education. *Elementary English*, 43, 402-408.

Ogawa, T. (1986). Reading and other leisure activities of Japanese adolescents. Paper presented to the Eleventh World Congress on Reading, London, England.

Ogbu, J. (1987). *Minority education and caste: The American system in cross-cultural perspective*. New York: Academic Press.

Ogbu, J., and Matute-Bianchi, M.E. (1986). Understanding sociocultural factors: Knowledge, identity and school adjustment. In *Beyond language: Social and cultural factors in schooling language minority students*, 73-142. Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Ogle, L.T., Alsalam, N., and Thompson-Rogers, G. (1991). *The condition of education, 1991*. Vol. 1. *Elementary and secondary education*. (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Olson, C.C. (1984). Effects of parenting and home environment on student achievement in grades 3 to 5. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Olson, D.R., Torrance, N., and Hildyard, A. (eds.). (1985). *Literacy, language, and learning*. Cambridge, England: Cambridge University Press.

Pallas, N., Natriello, G., and McDill, E. (1989). The changing nature of the disadvantaged population: Current dimensions and future trends. *Educational Researcher*, 18, 16-22.

Peal, E., and Lambert, W.E. (1962). The relationship of intelligence to bilingualism. *Psychological Monographs*, 76, 27, Whole No. 546.

Pearl, A. (1991). Democratic education: Myth or reality. In R. Valencia (ed.), *Chicano school failure and success*, 101-118. New York: The Falmer Press.

Pearson, P.D. (1985). Changing the face of reading comprehension instruction. *The Reading Teacher*, 38, 724-738.

Pearson, P.D., and Fielding, L. (1991). Comprehension instruction. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2. New York: Longman.

Pellegrini, A.D., Brody, G.H., and Sigel, I.E. (1985). Parents' book reading habits with their children. *Journal of Educational Psychology*, 77 (3), 332-340.

Peterson, T. (1988). Education reform implementation and impact in South Carolina. Paper presented at annual meeting of the American Educational Research Association, New Orleans.

Phi Delta Kappa. (1980). *Why do some urban schools succeed? The Phi Delta Kappa Study of Exceptional Urban Elementary Schools*. Bloomington, IN: Phi Delta Kappa.

Philips, S.U. (1972). Participant structures and communication incompetence: Warm Springs children in community and classroom. In C. Cazden, D. Hymes, and W.J. John (eds.), *Function in the classroom*, 161-189. New York: Teachers College Press.

Phillips, L.M., and Bolt, D. (1992). Family variables which impact on reading literacy: A review. Paper prepared for Westat.

Phillips, L.M., and Walker, L. (1987). Three views of language and their influence on instruction in reading and writing. *Educational Theory*, 37, 135-144.

Phillips, S.D., Smith, M.C., and Witte, J.F. (1985). Parents and schools: Staff report to the study commission on the quality of education in metropolitan Milwaukee. In A.T. Henderson (ed.), *The evidence continues to grow: Parent involvement improves student achievement*. Washington, DC: National Committee for Citizens in Education.

Piaget, J. (1959). *The language and thought of the child*. 3rd ed. London: Routledge and Kegan Paul.

Piaget, J. (1964). Development and learning. *Journal of Research in Science Teaching*, 2, 176-186.

Pickering, C.T. (1977). *Helping children learn to read*. New York: Chesford.

Polhemus, M.E. (1955). Home school cooperation for better readers. *Elementary English*, 32, 461-465.

Powell, A.G., Farrar, E., and Cohen, D.K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.

Pulkkinen, L. (1982). Self-control and continuity from childhood to late adolescence. In P.B. Bates and O.G. Brim, Jr. (eds.), *Life-span development and behavior*. Vol. 4. New York: Academic Press.

Purkey, S.C., and Smith, M.S. (1983). Effective schools: A review. *Elementary School Journal*, 83 (1), 427-452.

Purves, A. (1981). *Reading and literature: American achievement in international perspective*. Urbana, IL: National Council of Teachers of English.

Ramirez, A. (1985). *Bilingualism through schooling*. Albany: State University of New York Press.

Ramirez, J., and Merino, B.J. (1990). Classroom talk in English immersion, early exit and late exit transitional bilingual education programs. In R. Jacobson and C. Faltis (eds.), *Language distribution issues in bilingual schooling*, 61-103. Clevedon: England: Multilingual Matters.

Ramirez, J.D., Yuen, S.D., Ramey, D.R., and Pasta, D.J. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children.* San Mateo, CA: Aguirre International.

Rankin, P.T. (1967). The relationship between parent behavior and achievement of inner city elementary school children. Paper presented at the annual meeting of the American Educational Research Association, New York City.

Reid, I. (1990). Reading as framing, writing as reframing. In M. Hayhoe and S. Parker (eds.), *Reading and response.* Philadelphia: Open University Press.

Reinking, D., and Bridwell-Bowles, L. (1991). Computers in reading and writing. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 310-340. New York: Longman.

Resnick, L.B., and Robinson, B.H. (1975). Motivational aspects of the literacy problem. In J.B. Carroll (ed.), *Towards a literate society*. New York: McGraw-Hill.

Reynolds, D., and Cuttance, P. (1992). *School effectiveness: Research, policy and practice*. New York: Cassell.

Rivera-Medina, E.J. (1984). The Puerto Rican return migrant student: A challenge to educators. *Educational Research Quarterly*, 8, 82-91.

Roberson, D.R. (1970). Parents and teachers: Partners in the teaching of reading. *The Reading Teacher*, 23, 722-726.

Robinson, G., and Wittebols, J. (1986). *Class size research: A related cluster analysis for decisionmaking*. Arlington, VA: Educational Research Service.

Robinson, H.M., and Weintraub, S. (1973). Research related to children's interests and to developmental values of reading. *Library Trends*, 22, 81-108.

Rosen, B.C. (1959). Race, ethnicity and the achievement syndrome. *American Sociological Review*, 24, 47-60.

Rosenblatt, L.M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Carbondale, IL: Southern Illinois University Press.

Rosenholtz, S.J. (1985). Effective schools: Interpreting the evidence. *American Educational Research Journal*, 93 (3), 352-388.

Rosenshine, B. (1979). Content, time, and direct instruction. In P.L. Peterson and H.L. Walberg (eds.), *Research on teaching: Concepts, findings, and implications*. Berkeley, CA: McCutchan.

Rosenshine, B. (1986). Synthesis of research on explicit teaching. *Educational Leadership*, 43 (3), 60-69.

Rosenshine, B., and Berliner, D.C. (1978). Academic engaged time. *British Journal of Teacher Education*, 4, 3-16.

Rosenshine, B., and Stevens, R. (1984). Classroom instruction in reading. In P.D. Pearson, R. Barr, M.L. Kamil, and P. Mosenthal (eds.), *Handbook of reading research*. New York: Longman.

Rossell, C., and Ross, J.M. (1986). *The social science evidence on bilingual education*. Boston: Boston University.

Rowan, B., Bossert, S.T., and Dwyer, D.C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12 (4), 24-31.

Ruddell, R.B., and Speaker, R. (1985). The interactive reading process: A model. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 751-793. Newark, DE: International Reading Association.

Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (ed.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association.

Rumelhart, D.E. (1985). Toward an interactive model of reading. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. 3rd ed., 722-750. Newark, DE: International Reading Association.

Rutter, M., Maughan, B., Mortimore, P., and Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.

Salganik, L., and Karweit, N. (1982). Voluntarism and governance in education. *Sociology of Education*, 55 (April/July).

Salzman, S.A. (1987). Meta-analysis of studies investigating the effects of father absence on children's cognitive performance. Paper presented at annual meeting of the American Educational Research Association, Washington, DC.

Schumann, J.H. (1976). Affective factors and the problem of age in second language acquisition. *Language Learning*, 25, 209-239.

Schwartz, G., Merten, D., and Bursik, R.J., Jr. (1987). Teaching styles and performance values in junior high school: The impersonal, nonpersonal, and personal. *American Journal of Education*, 95, 346-370.

Secada, W. (1990). Research, politics and bilingual education. In C.B. Cazden and C.E. Snow (eds.), *English plus: Issues in bilingual education*, 147-161, Newbury Park, CA: Sage.

Sergiovanni, T.J. (1987). *The principalship: A reflective practice perspective*. Boston: Allyn and Bacon.

Shapson, S., Wright, E., Eason, G., and Fitzgerals, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, 17, 141-152.

Sheldon, W.D., and Carrillo, L. (1952). Relation of parents, home and certain developmental characteristics to children's reading ability. *The Elementary School Journal*, 52, 262-270.

Shilling, P., and Lynch, P. (1985). Father versus mother custody and academic achievement of eighth grade children. *Journal of Research and Development in Education*, 18 (2), 7-11.

Shinn, M. (1978). Father absence and children's cognitive development. *Psychological Bulletin*, 85, (2), 295-324.

Sindell, P. (1988). Some discontinuities in the enculturation of Mistassini Cree children. In J. Wurzel (ed.), *Toward multiculturalism: A reader in multicultural education*. Yarmouth: ME: Intercultural Press.

Singer, H., Gerard, H.B., and Redfearn, D. (1975). Achievement. In H.B. Gerard and N. Miller (eds.), *School desegregation*. New York: Plenum.

Sizemore, B.A., et al. (1983). *An abashing anomaly: The high achieving predominantly black elementary school.* ERIC, 236 274.

Skrabanek, R.L. (1970). Language maintenance among Mexican Americans. *International Journal of Comparative Sociology*, 11, 272-282.

Slaughter, D.T. (ed.). (1989). *Black children and poverty.* San Francisco: Jossey-Bass.

Slavin, R. (1984). Meta-analysis in education: How has it been used? and A rejoiner to Carlberg, et al. *Educational Researcher*, October 1984, 6-15 and 24-27.

Slavin, R., Forthcoming. Achievement effects of substantial reductions in class size. In *School and classroom organization.* Hillsdale, NJ: Erlbaum Associates.

Slavin, R., Karweit, N., and Madden, N. (1989). *Effective programs for students at risk.* Needham Heights, MA: Allyn and Bacon.

Smith, F. (1973). *Psycholinguistics and reading.* New York: Holt, Rinehart, and Winston.

Smith, M. (1972). Equality of educational opportunity: The basic findings. In F. Mosteller and D.P. Moynihan (eds.), *On equality of educational opportunity.* New York: Randon House.

Smith, M., and Glass, G. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17, 419-433.

Smith, M.S., and O'Day, J. (1988). *Teaching policy and research on teaching.* Stanford, CA: Stanford University, Center for Educational Research at Stanford.

Snow, C.E. (1983). Language and literacy: Relationships during the preschool years. *Harvard Educational Review*, 53, 165-189.

Snow, C.E. (1987). Beyond conversations: Second language learners' acquisition of description and explanation. In J.P. Lantolf and A. Labarca (eds.), *Research in second language learning: Focus on the classroom*, 3-16, Norwood, NJ: Ablex.

Snow, E., and Perlman (1985). Assessing children's knowledge about book reading. In L. Galda and A. Pelligrini (eds.), *Play, language, and stories*, 165-189. Norwood, NJ: Ablex.

Snyder, T.D. (1991). *Youth indicators 1991.* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Snyder, T.D., and Fromboluti, C.S. (1993). *Youth indicators 1993.* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Snyder, T.D., and Hoffman, C.M. (1992). *Digest of education statistics 1992.* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Southgate, V., Arnold, H., and Johnson, S. (1981). *Extending beginning reading.* London: Heinemann Educational.

Spiegal, D.L. (1981). *Reading for pleasure: Guidelines.* Newark, DE: International Reading Association.

Squires, D.A., Huitt, W.G., and Segars, J.K. (1983). *Effective schools and classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Stahl, S.A. (1992). Saying the "P" word: Nine guidelines for exemplary phonics instruction. *The Reading Teacher*, 34, 618-625.

Stallings J.A., Cory, R., Fairweather, J., and Needels, M. (1977). *Early childhood education classroom evaluation.* Menlo Park, CA: Stanford Research Institute.

Stallings, J. (1980). Allocated academic learning time revisited. Or beyond time on task. *Educational Researcher*, 9 (11), 11-16.

Stanovich, K.E. (1991). Word recognition: Changing perspectives. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, 419-452. New York: Longman.

Starkey, R., and Starkey, J. (1981). *Some differences between third graders who read up to expectancy and those who read below expectancy*. ERIC, ED 212 978.

Steinberg, L., Brown, B., Cazmarek, N., Cider, M., and Lazarro, C. (1988). *Noninstructional influences on high school student achievement*. Madison, WI: University of Wisconsin, National Center on Effective Secondary Schools.

Steinberg, L., Elmen, J.D., and Mounts, N. (1989). Authoritative parenting, psychological maturity and academic success among adolescents. Unpublished paper.

Stern, H.H., Swain, M., and McLean, L.D. (1976). *Three approaches to learning French*. Toronto: Ontario Institute for Studies in Education.

Straw, S. (1989). The actualization of reading and writing: Public policy and conceptualizations of literacy. In S.P. Norris and L.M. Phillips (eds.), *Foundations of literacy policy in Canada*, 165-181. Calgary, Alberta: Detselig.

Strickland, D.S. and Morrow, L.M. (eds.). (1989). *Emerging literacy: Young children learn to read and write*. Newark, DE: International Reading Association.

Stringfield, S., and Teddie, C. (1987). A time to summarize the Louisiana School Effectiveness Study. Paper delivered at the annual meeting of the American Educational Research Association, Washington, DC.

Sulzby, E. (1985). Children's emergent reading of favorite books: A developmental study. *Reading Research Quarterly*, 20, 458-481.

Sulzby, E., and Teale, W. H., (1991). Emergent Literacy. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2, New York: Longman.

Sulzby, E., and Teale, W.H. (1987). *Young children's storybook reading: Longitudinal study of parent-child interaction and children's independent functioning*. Final report to the Spencer Foundation. Ann Arbor: University of Michigan.

Swain, M. (1978). French immersion: Early, late or partial? *Canadian Modern Language Review*, 34, 577-585.

Swain, M., and Lapkin, R. (1991). The influence of bilingualism on cognitive functioning. *Canadian Modern Language Review*, 47, 635-641.

Teale, W.H. (1978). Positive environments for learning to read: What studies of early reading tell us. *Language Arts*, 55, 922-932.

Teale, W.H. (1984). Reading to young children: Its significance in the process of literacy development. In H. Goelman, A. Oberg, and F. Smith (eds.), *Awakening to literacy*, 110-21. Exeter, NH: Heinemann.

Teale, W.H. (1986). Home background and literacy development. In W.H. Teale and E. Sulzby (eds.), *Emergent literacy: Writing and reading*. Norwood, NJ: Ablex.

Tharp, R.G. (1989). Psychocultural variables and k constants: Effects on teaching and learning in schools. *American Psychologist*, 44, 349-359.

Thompson, G.G. (1952). *Child psychology*. Boston: Houghton Mifflin.

Thurlow, M., Graden, J., Ysseldyke, J.E., and Algozzine, R. (1984). Student reading during reading class: The lost activity in reading instruction. *Journal of Educational Research*, 77 (5), 267-272.

Tierney, R.J., and Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research*. Vol. 2. New York: Longman.

Tobin, A.W., and Pikulski, J.J. (1988). A longitudinal study of the reading achievement of early and nonearly readers through sixth grade. *National Reading Conference Yearbook*, 37, 49-58.

Tomlinson, T.M. (1989). Class size and public policy: Politics and panaceas. *Educational Policy*, 3 (3) 261-273.

Tomlinson, T.M. (1990). Class size and public policy: The plot thickens. *Contemporary Education*, 62 (1).

Tompkins, J.P. (1980). An introduction to reader-response criticism. In J.P. Tompkins (ed.), *Reader-response criticism: From formalism to post-structuralism*. Baltimore: Johns Hopkins University Press.

Trider, D., and Leithwood, K.A. (1988). Exploring influences on principals' behavior. *Educational Leadership*, 18 (3), 289-312.

Troike, R.C. (1981). Synthesis of research in bilingual education. *Educational Leadership*, 38, 498-504.

U.S. Department of Commerce, Bureau of the Census (1992). *School enrollment--Social and economic characteristics of students: October 1990*. Current Population Reports, Series P-20, No. 460. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (1992). *Dropout Rates in the United States: 1991*. GPO# 065-000-00519-1. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (1992). *The condition of bilingual education in the nation: A report to Congress and the President*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (1992). *The condition of education 1992.*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (1993). *The 1990 High School Transcript Study tabulations: Comparative data on credits earned and demographics for 1990, 1987, and 1982 high school graduates*. Washington, DC: Office of Educational Research and Improvement.

U.S. Department of Education. (1986). *What works: Research about teaching and learning*. Washington, DC: U.S. Department of Education.

Ubben, G.C., and Hughes, L.W. (1987). *The principal: Creative leadership for effective schools*. Newton, MA: Allyn & Bacon.

Veltman, C. (1988). *The future of the Spanish language in the United States*. New York: Hispanic Policy Project.

Venezky, R.L., and Winfield, L. (1979). *Schools that succeed beyond expectations in teaching reading*. Technical Report No. 1. Newark, DE: University of Delaware Studies on Education.

Vogt, L., Jordan, C., and Tharp, R. (1987). Explaining school failure, producing school success: Two cases. *Anthropology and Education Quarterly*, 18, 276-286.

Vygotsky, L.S. (1978). *Mind in society* (1938). (eds. and trans. M. Cole, V. John-Steiner, S. Scribner, and E. Sonberman). Cambridge. MA: Harvard University Press.

Wagenaar, T.C. (1977). School achievement level vis-a-vis community involvement and support: An empirical assessment. Paper presented at the annual meeting of the American Sociological Association, Chicago.

Waggoner, D. (1984). The need for bilingual education estimates from the 1980 census. *Journal of the National Association for Bilingual Education (NABE)*, 8, 1-14.

Walberg, H.J. (1986). Synthesis of research on teaching. In M. Wittrock (ed.), *Handbook of research on teaching*. 3rd ed., 15-32. New York: Macmillan

Walberg, H.J. (1985). Syntheses of research on teaching. In M.C. Wittrock (ed.), *Handbook of research on teaching*. Chicago: Rand-McNally.

Walberg, H.J., and Marjoribanks, K. (1976). Family environment and cognitive development: Twelve analytic models. *Review of Educational Research*, 46, 527-51.

Walberg, H.J., and Tsai, S. (1983). Reading achievement and attitude productivity among 17-year-olds. *Journal of Reading Behavior*, 15(3), 41-53.

Walberg, H.J., and Tsai, S. (1984). Reading achievement and diminishing returns to time. *Journal of Educational Psychology*, 76, 442-451.

Wallach, M.A., and Wallach, L. (1979). Helping disadvantaged children learn to read by teaching them phoneme identification skills. In L.A. Resnick and P.A. Weaver (eds.), *Theory and practice of early reading*. Vol. 3, 227-259. Hillsdale, NJ: Lawrence Erlbaum Associates.

Watson, T., Brown, M., and Swick, K.J. (1983). The relationship of parents' support to children's school achievement. *Child Welfare*, 62, 175-180.

Wells, G. (1986). *The meaning makers*. Portsmouth, NH: Heinemann.

Weston, W. (ed.). (1989). *Education and the American family*. New York: New York University Press.

Whitehead, F., Capey, A.C., Maddren, W. (1975). *Children's reading interests*. London: Evans and Methuen.

Whitehurst, G.J., Falco, F.L., Lonigan, C.J., Fischel, J.E., DeBaryshe, B.D., Valdez-Menchaca, M.C., and Caulfield, M. (1985). Accelerating language development through picture book reading. *Developmental Psychology*, 24, 552-559.

Wiesner, T.S., Gallimore, R., and Jordan, C. (1988). Unpackaging cultural effects on classroom learning. Native Hawaiian peer assistance and child-generated activity. *Anthropology and Education Quarterly*, 19, 327-353.

Wigfield, A., and Asher, S.R. (1984). Social and motivational influences on reading. In P.D. Pearson (ed.), *Handbook of reading research*, 423-452. New York: Longman.

Williams, J.P. (1979). The A B D's of reading: A program for the learning disabled. In L.A. Resnick and P.A. Weaver (eds.), *Theory and practice of early reading*. Vol. 3, 227-259. Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, T. (1987). *Participation in education*. Research Monograph No. 30. Hawthorn: Australian Council for Educational Research.

Williams, T., Long, M., Carpenter, P. and Hayden, M. (1993b). *Entering higher education in the 1980s*. Canberra, Australia: Australian Government Publishing Service.

Williams, T., Long, M., Carpenter, P., and Hayden, M. (1993a). *Year 12 in the 1980s*. Canberra, Australia: Australian Government Publishing Service.

Willig, A. (1986). Effectiveness of bilingual education. *Review of Educational Research*, 55, 269-287.

Wilson, W.J. (1987). *The true disadvantages: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

Wolff, R. (1966). The measurement of environment. In A. Anastasai (ed.), *Testing problems in perspective*. Washington, DC: American Council on Education.

BEST COPY AVAILABLE

Wong-Fillmore, L. (1983). The language learner as an individual: Implications of research on individual differences for the ESL teacher. In M.A. Clarke and J. Handscombe (eds.), *TESOL'82: Pacific perspectives on language learning and teaching*, 181-209. Washington, DC: TESOL.

Yopp, H.K., and Singer, H. (1985). Toward an interactive reading instructional model: Explanation of acceleration of linguistic awareness and metalinguistic ability in learning to read. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*, 135-143. Newark, DE: International Reading Association.

14. MODELING THE READING LITERACY OF FOURTH AND NINTH GRADERS

14.1. Introduction

The task of these analyses is the modeling of the between-student variation in reading comprehension with a view to understanding why students differ in their ability to comprehend written text. Three kinds of text are examined: narrative, expository, and document. The model as developed accounts for between-student variation in the comprehension of these text forms in terms of the attributes of the students, their families, the schools they attend, the classrooms in which they are schooled, and the forms of instruction to which they are exposed.

14.2. Modeling Reading Literacy

The development of a conceptual model is the first step. Essentially this is a view of why students differ in their ability to extract meaning from written text. Such a model specifies in some detail the important sources of between-student differences in reading comprehension, how these sources relate to each other, and how they affect reading comprehension. Ideally, this view of the world is anchored in the literature on reading and brings this literature, or logical extensions of it, together as a justification for the model advanced. The structure of the relationships developed in this way also frames the analyses and justifies their form.

The second step involves the selection/development of a statistical model that will adequately represent these theoretical considerations and, at the same time, be consistent with the study design and the structure of the data. The development of an explicit conceptual model is especially important because the IEA Reading Literacy Study is a cross-sectional, nonexperimental design. Under these conditions there is no opportunity to randomize away known and unknown sources of confounding variation when estimating the unique effect of a particular variable. These sources of variation in students' reading comprehension must be controlled statistically. In order to do this, one needs to identify the variables in question and measure them, a process whose logic is best defined within an explicit conceptual framework.

Equally important is the need for explicit measurement models, one for each construct in the conceptual model. Measurement models define the relationships between constructs (which, in principle, are unobservable) and their indicators, the real-world, observable but fallible measures of these constructs. In the Reading Literacy Study analyses, these models would specify the linkages of achievement test items and questionnaire items, singly or in groups, to the constructs they tap.

14.2.1. The Place of Modeling in IEA Studies

To lay a foundation for the way in which we develop the conceptual and measurement models, and the analyses that flow from these, a brief note about the place of modeling in IEA studies may be in order.

The modeling of achievement is only one of three general sets of concerns that drive IEA studies. Measures of subject matter knowledge dominate the design of these studies. These achievement tests provide for an assessment of what students within nations have learned, and for comparisons of national achievement levels, the latter usually displayed as a ranking of nations. Such rankings usually capture a good deal of public attention in the short term (Medrich and Griffith 1992, 2). However, assessments alone tell only part of the story, since they do not suggest why it is that levels of learning differ between students and between nations--and what one might do about this difference.

Knowledge about levels of student learning is complemented by information about the structures and processes of schooling in the various nations. Traditionally, students, teachers, and principals respond to questionnaires to provide this information. The bulk of the statistics produced in this connection have to do with differences between nations in educational policy and provision: national assessment structures, the locus of curriculum decisions, the resources available to schools and classrooms, the attributes of teachers and principals, the instructional practices of teachers, and so on.

This information tends to be used in three main ways. First, it can be simply displayed in much the same way as the assessment results, that is, national education systems described one by one, and nations ranked on the attributes of their education systems--levels of resources expended, qualifications of teachers, instructional time, and so on. Second, when national rankings on these school system characteristics parallel national rankings for achievement, inferences can be made about the achievement-enhancing qualities of school systems -- for example, nations that track students between schools (as distinct from within schools) have higher levels of achievement (Medrich and Griffith 1992, 31). Third, within-nation relationships can be provided in descriptive displays and sometimes in between-nation comparisons.

More elaborate attempts to model achievement are attempted only occasionally. And, when attempted, this modeling takes on many of the characteristics of secondary analysis, in the sense that the models are developed a posteriori within the constraints of the data at hand. Nevertheless, the conceptual structure of the emergent models tends to be fairly uniform, reflecting the enduring concerns of IEA studies--the effects of schools, classrooms, and teachers on student achievement, after due allowance for differences in student background. For recent examples of international modeling, see Schmidt and Kifer (1989), Keeves and Saha (1992), Keeves and Morgenstern, (1992), and Postlethwaite and Wiley (1992). Rosier and Long (1991) provide an example of modeling in a national report.

14.2.2. Developing Conceptual and Measurement Models

In the IEA Reading Literacy Study the conceptual and measurement models were not well specified as part of the study design. As a consequence, the analyses took on many of the characteristics of secondary analyses undertaken within a relatively focused omnibus survey such as High School and Beyond (HS&B) or the National Education Longitudinal Study of 1988 (NELS:88). Modeling with NAEP data may be an even better example, since measures of student achievement are the driving force there as they are in IEA studies.

Under these circumstances model development followed a fairly common sense two-stage course. The first stage focused on the identification and development of measurement models within the data at hand, which allowed us to take stock of the major available constructs. The second step involved inferences about the underlying conceptual framework, followed by the organization of the constructs into this schema.

350

14.2.2.1. Measurement Models

The first step was the development of measurement models, which lay out the hypothesized relationships between constructs (unmeasured and unobserved) and their observed indicators. By defining constructs, these models identify the substantive content of the data, reduce the number of variables through aggregation of questionnaire items, and establish at a more conceptual level what the data are about. Faced with 500 or so questionnaire responses from students, teachers, and principals in each population, we had little in the way of formal guidance about the constructs being measured or about which items are thought to go with which construct. Consequently, measurement models were developed based on inferences derived from both the structure and placement of the questionnaire items, and from the assumption that the basic notions behind the items were grounded in the reading literature and/or in the more general literature on models of student achievement. This process is described in detail in Chapter 13.

14.2.2.2. Conceptual Models

Conceptual models reflect arguments, theoretical and otherwise, about the social processes one is attempting to capture in the research design. As such, they are simply a formalization of arguments about the patterns of relationships among constructs--in essence, what affects reading comprehension and the things that contribute to it. What kind of instructional methods affect reading comprehension? what kinds of teachers use these methods? what kind of schools do these teachers inhabit? what kinds of communities accommodate these schools? Arguments of this kind give form and meaning to the analyses.

At a more formal level, conceptual models can be usefully represented as structural models, which lay out the structure of the relationships among the constructs being considered. These models specify on theoretical/substantive grounds the underlying social processes that give rise to patterns of relationships among observed variables, social processes whose existence is inferred from these observed patterns. In the sociological literature, these formulations tend to be called structural equation models (see Bielby and Hauser 1977). At a more operational level, these theoretical structures specify which variables are to be controlled statistically when a certain effect is being examined, the theoretical arguments embodied in the structural relations providing the justification for these statistical controls.

In dealing with the assemblage of variables/constructs identified, the first step was to infer a structure for the data by grouping variables within relatively homogeneous and substantively meaningful categories. This view of the basic conceptual structure of the data can be represented as a matrix that defines 11 broad groups of variables. This matrix goes some way toward capturing, in a fairly common sense way, the underlying dimensions of the explanatory variables present in the Student, Teacher, and School Questionnaires. The development of this classification was described in some detail in Chapter 13. It is repeated at this point in the narrative (as Figure 14-1) to serve as a foundation for the models developed from it.

14.2.3. Hierarchical Models and Hierarchical Data

The classification of variables in this way also reflects the structure of the data. The Reading Literacy Study data have a hierarchical structure--students nested within classrooms within schools. Approximately 7,000 fourth grade students in 300 classrooms, along with almost 4,000 ninth grade students in 165 classrooms, provided the data for this study of reading comprehension. The characteristics of each student and each student's family, along with the measures of reading comprehension, are unique to each student. However, the attributes of community, school, teacher, classroom, and instructional practices are common to each of the students in a particular classroom.

Figure 14-1 indicates this distinction with the line separating the 11 categories into two groups. Above the line are the data on the students, their families, and their achievements in reading comprehension. Below the line are the measures of schools and schooling derived from their teachers and school principals. The matter to be resolved is how best to link the two. IEA studies to date have not solved this problem, in part because an appropriate statistical model was not widely available until the late 1980s, though the problem was recognized earlier (Burstein 1980).

Though we will consider both of these levels simultaneously in the actual analyses, for at least the reason that we postulate effects across the two levels, the explanation of the model development process is made clearer by a separate treatment of each level. Note, however, that this is a single model rather than two separate models.

14.2.3.1. Student-Level Model

The nature of the structural relations among constructs at each level is implicit in the layout of the blocks of variables. In the student-level model, above the line, the variables labeled as attributes of students and their families are considered as exogenous. Measures of student and parent behaviors within the family are treated as endogenous variables, at once an outcome of student and family attributes and a source of variation in reading comprehension (Figure 14-2).

The meaning of the structural relations postulated is as follows. With respect to the exogenous variables noted in the left-most block, the student-level model postulates that between-student variation in reading comprehension is due, in part, to the following factors:

1. **Age differences among students.** These arise partly as a function of developmental differences related to age (especially among fourth graders), partly from differences in exposure to reading and reading instruction, and partly as a reflection of the grade retardation of some students with learning difficulties. Since all of these effects are captured in the one variable, an unequivocal interpretation of the age effect is difficult.

2. **Gender differences**, reflecting at least the superior verbal capabilities of female students. The expectation is that gender differences themselves will vary with the verbal content of the text--from substantial in the case of narrative text to minor when the comprehension of document text is at issue.

3. **Differences across racial/ethnic groups** as a function of related social, cultural, economic, and language differences among the groups.

4. **Differences by level of parental education.** In the absence of a measure of parental occupation, these parental status attainments tap both the educational and social status resources of families. We might expect differences in effect as well between father's and mother's education, particularly among the fourth graders, given the dominant role of mothers in child rearing.

Figure 14-1. Conceptual structure of the Reading Literacy Study data

	ATTRIBUTE	ENVIRONMENT	ACHIEVEMENT
STUDENT	age gender race/ethnicity language		narrative comprehension expository comprehension document comprehension
FAMILY	family composition augmented family father's education mother's education family wealth	regularity of meals study facilities literacy resources TV watching homework help parental interest	
TEACHER	gender race/ethnicity teacher training formal education teaching experience	type emphasis reading homework grouping practices assessment emphasis teacher's reading preference transmission/translation emphasis	
CLASS	size reading resources instructional time (all) instructional time (reading)	% remedial students % remedial assisted	
PRINCIPAL	gender race/ethnicity training (administration) training (reading) experience (total) experience (present school) instructional leadership staff development		
SCHOOL	sector instructional time library resources % specialist teachers % minority teachers size		
COMMUNITY	urbanicity resources region cooperation		

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Figure 14-2. Student-level model

ATTRIBUTE

age
gender
race/ethnicity
language
family composition
agumented family
father's education
mother's education
family wealth

ENVIRONMENT

regularity of meals
study facilities
literacy resources
TV watching
homework assistance
parental interest

ACHIEVEMENT

narrative comprehension
expository comprehension
document comprehension

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

5. **Differences by family wealth** as a function of the economic resources families can command and the goods and services they can purchase.

6. **Differences between families in their composition** based on the presence of one, both, or neither parent, and whether parents are biological parents or stepparents. Effects derive in part from the parental resources available to the student, in part through the experience of family dislocation that attends marriage breakup and/or such difficulties as may be experienced within single-parent families.

7. **A difference based on whether the family is "nuclear,"** in the sense of being composed of parent(s) and siblings, or "augmented" by including grandparents, other relatives, and/or other nonrelatives.

8. **Differences by family language.** In nations where bilingualism is not a particular advantage, operating in a language other than the language of the test is likely to be a handicap. In this case, the distinction is based on whether the student's first language and home language are English, or whether one or both are a language other than English.

These eight exogenous variables appear to be the most reasonable and justifiable among those available in the dataset. They are treated as exogenous within this model because they represent relatively immutable attributes of students and their families whose variation we take as a given. Other models, of course, might treat some or all of these same variables as endogenous. There is nothing about the variables themselves that defines them as exogenous. They are so defined within the context of the particular model as variables that might be seen to affect other variables, but are themselves not influenced by anything within the model.

By contrast, endogenous variables are those whose variation is explained by the model itself, in part by the exogenous variables, in part by other endogenous variables, and in part by undefined influences represented as error terms in the structural equations that capture these structural relations. The group of variables included in the category family environment in Figure 14-1 are endogenous, as are the three reading comprehension measures. Within this model, the variation in narrative, expository, and document comprehension is explained, in part, by the exogenous variables together with those endogenous variables in the family environment group. Variation in the family environment variables is explained by the exogenous student and family attribute variables alone.

Seen another way, the model indicates that part of the influence of the exogenous variables on reading comprehension is transmitted through the endogenous family environment variables, and part is direct. In more substantive terms this means, for example, that the influence of family wealth on reading comprehension comes about in part because wealthier families can provide more in the way of literacy resources, private study facilities, and regular meals, all of which influence reading comprehension, and in part because family wealth influences reading comprehension in ways such as the provision of better schools and better neighborhoods. Respectively, these are the indirect and direct effects of family wealth on reading comprehension. Together they make up the total effect of family wealth.

Thus, we are saying that the several aspects of family environment represented differ among families in parallel with those attributes of families shown. This point is most easily argued in connection with the social and economic status variables; better educated and wealthier families are more likely to control TV watching, encourage reading, and provide literacy resources, a place to study, and regular meals. Parallel arguments can be developed for the other exogenous variables.

The form of the analyses is now governed by a structure that dictates and justifies the variables to be controlled statistically when estimating the net effects of a variable of interest. The analyses gain extra meaning through the (theoretical) distinction between exogenous and endogenous variables, and the reflection of this distinction in multiple equation models in which one can represent total, direct, and indirect effects. For example, it is possible to estimate the total effect of family wealth and other aspects of family background, and then consider how much of the difference between the "rich" and the "poor" is brought about by differences in the environments of families as a direct result of their economic circumstances.

14.2.3.2. Classroom-Level Model

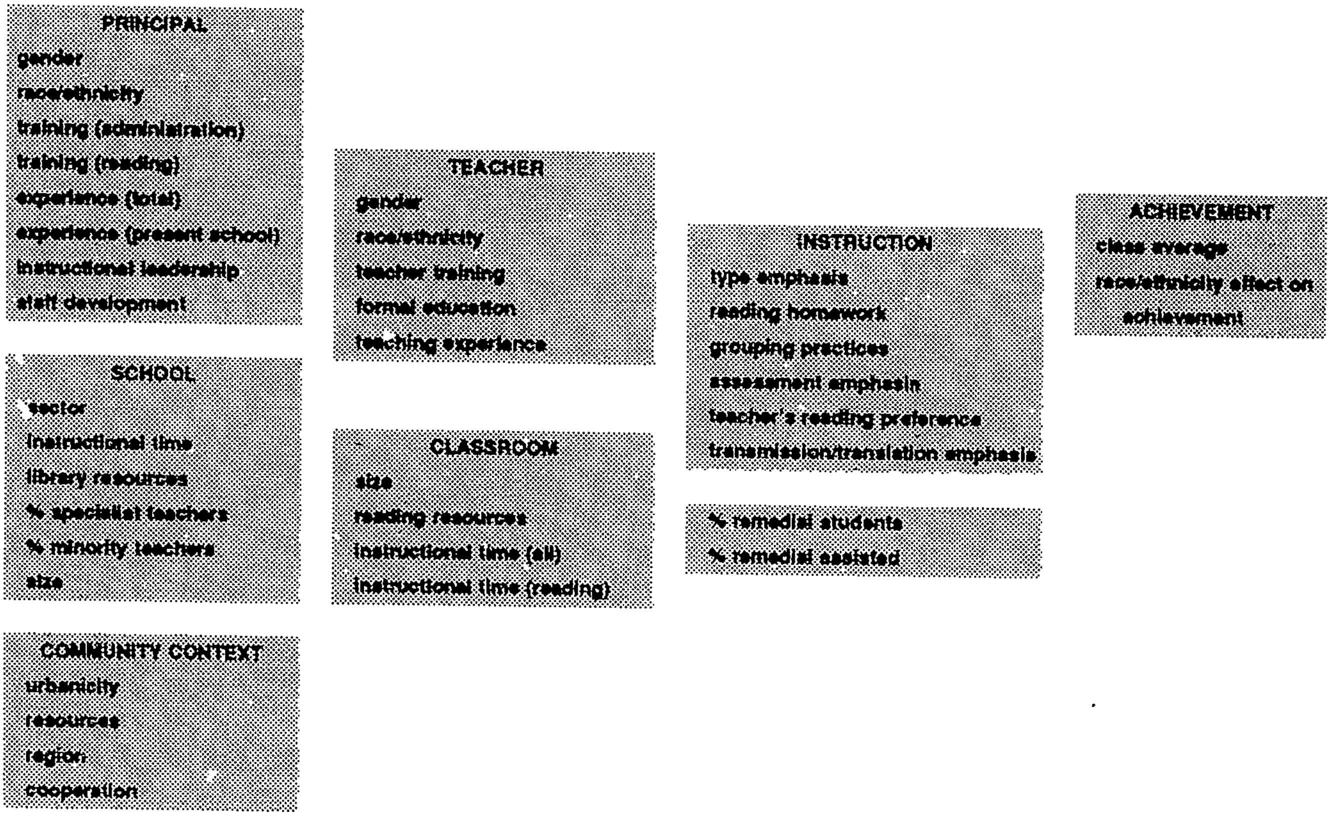
The basis for a classroom-level model, one which examines the influence of communities, schools, and teachers on students' reading comprehension, lies below the line in Figure 14-1. The immediate focus of this model is not the achievement differences between students, but rather

a. Achievement differences between classrooms, as these are expressed in the average achievement levels of their students; and

b. Differences between classrooms in the effects of student-level attributes on achievement.

The explanatory content of this model is defined by the attributes of communities, schools, principals, classrooms, and teachers, plus the mix of teacher behaviors and classroom characteristics that compose classroom environments. However, while this categorization was a useful first step, it makes sense to depart from the structure shown to develop a more differentiated classroom-level model. In this model (Figure 14-3), the attributes of schools, their leaders, and the surrounding communities are treated as exogenous variables. Characteristics of the teachers and the classrooms in which they work are considered endogenous, an outcome of the exogenous school-level variables and a source of variation in both instructional practices and reading achievement.

Figure 14-3. Classroom-level model



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

As noted above, although the student-level model and the classroom-level model are not estimated separately in the statistical analysis, it makes sense to consider their substantive meaning separately in the first instance. Where the student-level model considers the pattern of student and family influences on the achievement levels of individual students, the classroom-level model focuses on differences in the reading achievement levels of classrooms and seeks to understand these in terms of parallel differences in the attributes of the communities, schools, principals, classrooms, and teachers.

Further, in addition to examining the sources of between-classroom variation in average reading achievement, the classroom-level model looks at between-classroom variation in some of the effects from the student-level model. What this means is that we need not treat the student-level effect of, for example, family wealth on reading comprehension as though it were independent of what goes on within classrooms. It is possible that this effect varies across classrooms because communities, schools, principals, classes, and/or teachers differ in what they are or do. In some classrooms the effect of family wealth on achievement might be wiped out by compensatory programs. In others, a laissez faire approach to disadvantage may well maintain the strength of this relationship. For these reasons, Figure 14-3 shows both the average reading achievement of the class as a whole, together with the effect of family wealth on achievement within this class, as the ultimate outcome variables in the classroom-level model. Family wealth is chosen only to illustrate that we can model classroom slopes as well as intercepts.

Figure 14-3 also shows the distinction made between exogenous and endogenous groups of variables. Community context, school attributes, and principal attributes are treated as exogenous, while classroom and teacher attributes and instructional practices are considered as endogenous. We also make some distinctions between the endogenous groups in terms of where they stand relative to each other. As postulated, teacher and classroom variables are affected only by the exogenous variables, but the instruction group is influenced by all of the foregoing.

Without going into detail, the propositions reflected in this model are quite straightforward. The environments within classrooms and the attributes of teachers are a function, in part, of the community context in which schools operate, the attributes of the school itself, and the characteristics of the principal. It is reasonable to suppose further that all of these influence the instructional practices adopted by teachers, and all in turn have the potential to affect the average level of reading comprehension of the class, as well as the within-class influence of student-level variables on reading achievement.

14.3. The Statistics of Hierarchical Models

Consistent with the general argument advanced earlier, once the conceptual model is in place, the next step is to develop a statistical model to reflect the theoretical/substantive arguments represented in the conceptual model and the structure of the data.

As noted above, the student-level and classroom-level models are really two levels of a single model, not two separate models. The basic issue left unresolved in the past was how to link the two in a way that allowed an examination of the effects of group characteristics on individual behaviors. The matter tended to be addressed either by aggregating the student-level data to the classroom level, in which case classroom means become the dependent variable and the individual-level data were ignored, or by disaggregating the classroom-level variables to the student level such that each student in a class has an identical value on classroom-level variables. Both approaches have their problems, as we discussed below.

14.3.1. Multilevel Models

A statistical model that incorporates multiple levels simultaneously to relate the effects of group characteristics to individual behaviors is of more than passing interest. Educational research focuses much attention on the effects of schools, classrooms, and teachers on the behaviors of individual students. In sociology the behaviors of individuals set in social structures/aggregations is a central issue for the discipline. Economists face this question when linking micro- and macroeconomic models--human capital models, for example. Demographers face similar problems in multinational studies of fertility; one of the first applications of hierarchical models was developed in the context of the World Fertility Survey. Psychological studies of learning growth can formulate the measurement of individual change in these terms: Burstein (1980), Hannan (1971), Mason, Wong, and Entwisle (1983), and Bryk and Raudenbush (1987) offer perspectives on this matter.

In common with most other work, IEA analyses to date have tended to consider this as a unit-of-analysis problem and have worked at either or both of the classroom and individual levels, though not simultaneously. Classroom-level analyses have two main difficulties. First, they ignore the variation that exists between students within classrooms, generally the greater part of all the variation. Second, interpretations often run into the problem of inferring individual-level effects from aggregate-level relationships, a matter termed the "ecological fallacy" by Robinson (1950).

Analyses limited to the individual level have their own problems. Either they have to ignore the group-level variables, whose effects are the whole point of the study, or they disaggregate the group measures across the individuals in the group. Unfortunately, the latter procedure violates the assumption of the independence of observations needed for the statistical models used. Though not widely adopted in IEA studies, this procedure was a standard approach in the "contextual effects" literature that grew in sociology during the 1970s (for example, Meyer 1970; Alexander and Eckland 1975) but faded in the face of counterintuitive findings and methodological criticisms (Hauser 1970; Hauser 1974; Hannan 1971; Hannan and Burstein 1974).

The conceptualization of statistical models appropriate for the analysis of hierarchical data, and the development of the technology needed to estimate these models, came together in the 1980s in the work of Mason, Anderson, and Hayat (1988), Bryk et al. (1988), Longford (1988), and Rabash, Prosser, and Goldstein (1989). The analyses presented below follow Raudenbush and Bryk (1986), Bryk et al. (1988), and Bryk and Raudenbush (1992), and use the computer program HLM developed by these authors. We use their term to identify these models as "hierarchical linear models."

14.3.1.1. Specifying Multilevel Models

Conceptually, hierarchical linear models address the problem of multilevel analyses by abandoning the conventional individual-level model estimated on the sample of individuals as a whole. Instead, one can think of the process as one in which the individual-level model is estimated separately within each of the second-level groups--classrooms in this instance. Consider, for example, a model in which minority status is used as a single explanatory variable to predict narrative comprehension. Following Bryk and Raudenbush (1992, ch. 2), within each classroom we might estimate a regression equation of the following form:

$$Y_i = \beta_0 + \beta_1 X_i + r_i \qquad (14.1)$$

If X_i is centered around the classroom mean for minority status, then Y_i is the narrative comprehension score of student "i," β_0 is the classroom mean for narrative comprehension, β_1 is an increment that comes from the student's own minority status, and r_i is the error term, the random component associated with student "i."

With one of these equations estimated for each classroom, the fourth grade sample produces estimates of some 300 classroom means (β_{0j}) that are likely to vary across the "j" classrooms--some classrooms will show high levels of achievement, others will show low levels, and most will be in between. Figure 14-4 illustrates the extent of this variation with respect to the means for narrative comprehension in fourth grade classrooms.

Similarly, there will be some 300 slope coefficients (β_{1j}) measuring the effect of minority status on achievement in each of the "j" classrooms. These too could be expected to vary across classrooms, perhaps as a function of what schools and teachers do, perhaps simply as a function of the selection of students into schools and classrooms. Figure 14-5 provides an illustration based on the effect of minority status estimated for each of the 300 or so fourth grade classrooms. Negative values indicate classrooms in which the effect of being a member of a minority group is to reduce the student's achievement score to a value below the classroom mean. Positive values indicate classrooms in which minority students have scores above those of nonminority students. Slopes are estimated within classrooms by Ordinary Least Squares, but are weighted to take into account their varying reliability as estimates of the true classroom slope (see Bryk and Thum 1989, 9).

Figure 14-4. Variation in class mean reading proficiency scores, narrative comprehension: Grade 4

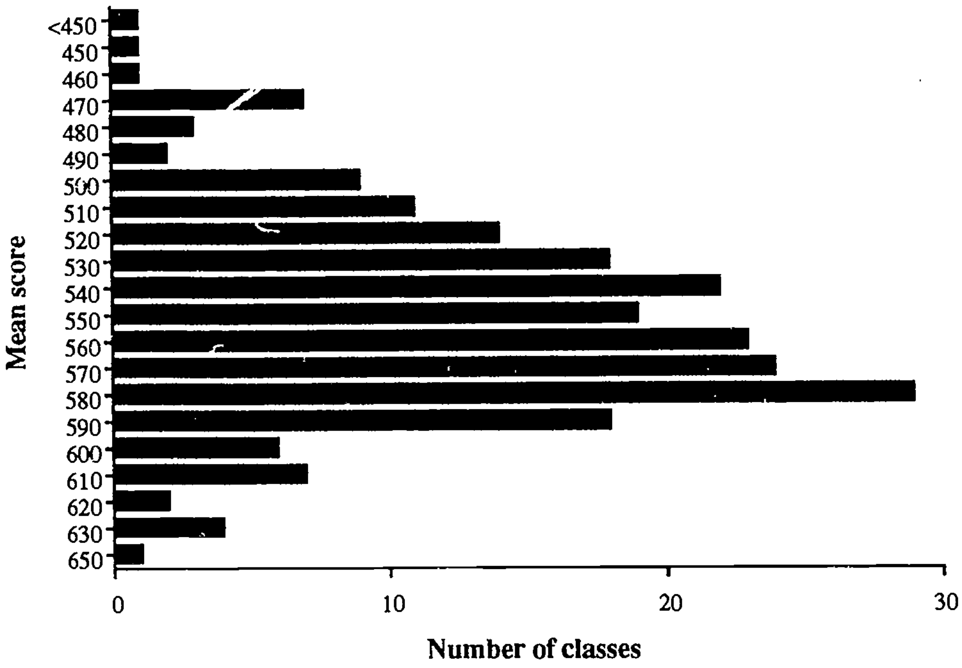
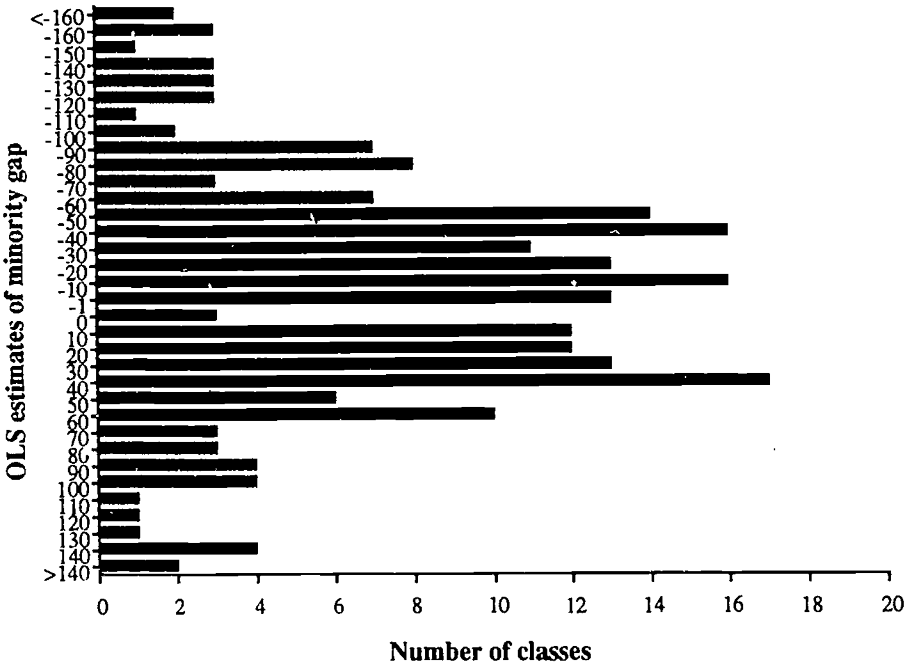


Figure 14-5. Variation in class minority gap, narrative reading comprehension: Grade 4



OLS = Ordinary Least Squares

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

It follows that we could think of these intercepts and slopes as characteristics of the classrooms themselves. And their variation across classrooms might be modeled by group-level variables associated with the classrooms--the community, school, class, and teacher attributes we have noted. In this way, it becomes possible to look at the extent to which, say, differences in instructional practices across classrooms are linked to differences in the average level of achievement of students. By way of example, consider "homework" as the instructional practice of interest. If we were to represent that in an equation, with "W" the measure of homework, it would have the following general form:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \tag{14.2}$$

With W_j centered around the grand mean, β_{0j} is the mean achievement level in classroom "j," γ_{00} is the weighted average of classroom means, γ_{01} is the increment to this grand mean due to the level of W (amount of homework) in classroom "j," and u_{0j} is the component of variation in β_{0j} that is unique to classroom "j" and whose magnitude varies randomly across classrooms. That is, the mean achievement score of a particular classroom is given by the average of means across all classrooms plus an increment due to the level of homework provided in the particular classroom, plus a increment that stems from (unknown) factors unique to that classroom.

It is possible to model variation in the slope coefficients across classrooms according to the same logic. That is, we might argue that the effect of minority status on achievement is less in classrooms where there are high levels of homework. The equation would look like:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \tag{14.3}$$

β_{1j} is the minority status-achievement slope for classroom "j," γ_{10} is the average of the minority status-achievement slopes across classrooms, γ_{11} is the increment to the minority status-achievement slope due to a one-unit difference in homework provision, and u_{1j} is the unique classroom contribution to the relationship between minority status and achievement. Assuming that γ_{11} was negative, for each unit increase in homework provision there is a γ_{11} decrease in the relationship between minority status and achievement--the more homework, the less the effect of disadvantage associated with minority status.

Further, since minority status is a dichotomous variable, β_{1j} is the achievement gap between minority and nonminority students in the classroom, γ_{10} is the gap overall, and γ_{11} is the contribution to closing this gap derived from the level of homework provided in the classroom.

By casting the model in these terms, it becomes possible to say that other things equal, the students in some classrooms do better where.......; other things equal, in some classrooms the achievement deficit of minority students is minimized when......; and so on. This has the potential advantage that analyses may suggest ways in which the achievement of all students could be promoted and ways in which disadvantage may be ameliorated.

Substituting equations (14.2) and (14.3) into equation (14.1) provides a view of the combined model for the achievement score of student "i" in school "j." Recall that X_{ij} is centered around the classroom mean and W_j around the grand mean.

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + u_{0j} + \gamma_{10}X_{ij} + \gamma_{11}W_jX_{ij} + u_{1j}X_{ij} + r_{ij} \tag{14.4}$$

This equation shows that an individual's achievement score Y_{ij} can be seen as having the following components:

1. A base-level component in common with all persons, the grand mean, γ_{00}, which is the average achievement level for the grade;
2. An increment due to the average level of achievement in the classroom in which student "i" is located, $\gamma_{01}W_j$, an increment resulting from the level of homework provided;
3. An increment due to effects on the classroom average that are unique to that classroom, u_{0j};
4. An increment due to the average effect of minority status in all classrooms, $\gamma_{10}X_{ij}$;
5. An increment due to the effect of minority status in the classroom in which student "i" is located, $\gamma_{11}W_jX_{ij}$, an effect that varies between classrooms according to the amount of homework provided;
6. An increment due to unknown classroom-specific factors that affect the size of the minority status-achievement slope within the classroom, $u_{1j}X_{ij}$; and
7. An increment due to unknown person-specific factors, r_{ij}.

This model may be generalized to multiple X_{ij} and W_{ij}.

14.3.1.2. Developing Multilevel Models

The nature of the Reading Literacy Study research design is such that model development involves a good deal of inference, exploration of the data, and judgment. Given this, and keeping in mind that the overall model will place constraints across levels, it seems sensible to develop the multilevel model in two stages.

In the first stage, we define a level 1 model that links the attributes of students and their families to the reading achievement of individual students within each classroom. The effects of some of these attributes will be treated as constant across classrooms, while others will be seen to vary. As a consequence of modeling within classrooms, this level 1 model focuses only on within-classroom variance unconfounded by between-classroom variance. (We actually define a limited level 2 model at the same time but this is less of an explanatory model than a means to provide better estimates of the level 1 effects.)

In the second stage, a level 2 model is proposed, in this case at the classroom level. This model is designed to explain the variation in average achievement across classrooms and the effects of student attributes on achievement as these vary across classrooms. As potential sources of this variation, we include attributes and/or behaviors of the community in which the school is located, the school, the principal, the classroom, the teacher, and the instructional practices of teachers thought to foster reading comprehension. One advantage of this multilevel approach is that community, school, classroom, and/or teacher attributes are used to explain only that part of the total variance one could reasonably expect them to influence--the between-classroom component. This avoids a problem with earlier approaches to modeling school/classroom effects that attempted to explain the total variance with variables that could logically only explain the between-school component. Since the between-school component is usually small relative to the within-school component, school effects, so defined, have been hard to find.

14.3.2. Variance Within and Between Classrooms

As a first step, we establish the magnitude of what it is that is being explained with the multilevel model -- variance in reading comprehension between-persons within classrooms, and variance in the average levels of achievement between classrooms. The estimates of within- and between-classroom variance are developed from the fully unconditional two-level model; that is, a model without X_{ij}'s or W_j's. In this instance, the level 1, level 2, and combined models can be represented, respectively, as follows:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad (14.5)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (14.6)$$

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad (14.7)$$

The variance of r_{ij} in equation (14.7) is the within-classroom variance, while the variance of u_{0j} is the between-classroom variance.

Table 14-1 displays estimates of the variance within and among classrooms for each of the three reading comprehensions measures separately for fourth and ninth grade students. These data make it clear that there are statistically significant amounts of variation among persons within classrooms, and among classroom means, for all three measures at both age levels. The intraclass correlation coefficient (rho) indicates that about 20 percent of the variation in fourth grade reading comprehension scores is between-classroom variance. The comparable figure for ninth grade students is closer to 40 percent.

14.4. Developing the Student-Level Model

In developing further the general structure for the model advanced in Figure 14-2, three matters need further consideration:

a. The specification of structural equations that will capture the distinction between exogenous and endogenous variables and, in so doing, provide for a calculus of the direct and indirect effects;

b. The question of variation in classroom means--whether the level 1 model suggests that we should attempt to model classroom means in the level 2 model; and

c. The matter of whether there is statistically significant between-classroom variation in slopes that can be modeled at the classroom level, and, on the assumption that not all slopes will show significant amounts of variation, the identification of those slopes that do.

Table 14-1. Variance within and between classrooms: One-way ANOVA model, grades 4 and 9

Fixed effect	Grade 4		Grade 9	
	Coefficient	Standard error	Coefficient	Standard error
1. Narrative average class mean	556.5	2.7	537.9	5.0
2. Expository average class mean	540.8	2.3	538.9	5.5
3. Document average class mean	552.3	2.3	527.4	4.0

Random effect	Grade 4					Grade 9				
	Variance component	d.f.	χ^2	p-value	ρ	Variance component	d.f.	χ^2	p-value	ρ
1. Narrative										
class mean	1,768	297	1,680	.000	.19	3,823	164	2,051	.000	.39
level 1 effect	7,615					5,991				
2. Expository										
class mean	1,325	297	1,853	.000	.21	4,500	164	2,007	.000	.39
level 1 effect	5,085					7,168				
3. Document										
class mean	1,312	297	1,748	.000	.20	2.367	164	1,692	.000	.34
level 1 effect	5,322					4,591				

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

14.4.1. Structural Equations for the Student-Level Model

The student-level (level 1) model shown in Figure 14-2 has a basic structure in which the attributes of students, their parents, and their families as treated as exogenous variables, while aspects of family environments are considered endogenous. Implicit in this ordering of the variables is the notion that the kinds of environments provided within families are likely to vary with the attributes of parents, particularly their status attainments, and the structure and composition of the family itself. Given this, two structural equations describe the model: one involving only the exogenous variables (the exogenous equation); and a second involving both exogenous and endogenous variables (the all-variable equation).

This configuration implies that we estimate the model in two stages and, in so doing, provide for a partial calculus of the direct and indirect effects. By estimating the effects of the exogenous variables alone, we gain a notion of the total net effect of each exogenous variable. Estimating the effects of both exogenous and endogenous variables together gives the direct effect of each variable. Comparisons between these estimates for each of the exogenous variables allows a notion of the extent to which, for example, differences in reading achievement between poor and wealthy families come about because of differences in the family environments of the two groups. Logic dictates that the effects of the endogenous family environment variables are total effects that are also the direct effects--there are no intervening variables.

14.4.2. Variation in Classroom Means

Table 14-1 establishes the case for variation in classroom means, the extent of which is illustrated in Figure 14-4. The between-classroom variation is statistically significant for each of the three outcome measures. As a result, we will consider the intercept to be random in the level 1 model and will attempt to model its variation with classroom-level variables in the level 2 model.

14.4.3. Fixed and Random Slopes

This matter is complicated by the likelihood that not all slopes in the level 1 model (the student-level model) will necessarily show significant amounts of variation across classrooms. Some, perhaps most, will be more or less the same in each classroom, since they are relatively unaffected by selection into schools and/or are not influenced much by what schools do. Others will take on different values in different classrooms, reflecting variation in the policies of the schools and the behaviors of their teachers, or simply the fact of socioeconomic and related segregation by school/classroom.

Logically, only slopes that vary across classrooms can be modeled. Thus, the question is one of how to identify those slopes to be treated as random, and those to be considered fixed. Since the modeling of variation in slope coefficients is a relatively recent phenomenon, there is not a lot of guidance in the research literature or assistance from the theoretical/substantive literature, which, on the whole, tends not to be presented in these terms. However, evidence on this matter is not completely absent: Raudenbush and Bryk (1986) show that the SES-achievement slope varies with school sector; Bryk and Raudenbush (1987) show changes in growth curves; Bryk and Thum (1989) identify SES and attitudinal effects on dropping out of school and show how these vary with school climate; Lee and Bryk (1989) show effects of sector, school composition, and climate on the effects of minority status, socioeconomic status, and academic background on mathematics achievement; Raudenbush, Rowan, and Cheong (1992) report that the effect of tracking on teachers' levels of efficacy varied by subject area; and Gamoran (1992) shows how the structural characteristics of tracking influence the achievement gap between tracks.

Common sense suggests that this between-classroom variation in level 1 slopes probably comes about, in part, because some schools and/or teachers treat children unequally. This treatment could be deliberate or incidental, well intentioned or not. Compensatory education programs designed to ameliorate disadvantage are deliberate, well intentioned, and designed to change the relationship between an attribute of students and their achievement. Forms of discrimination are also deliberate, affect some students but not others, probably change the relationship between student attributes and achievement, and are not well intentioned. In addition, the deliberate actions of schools and teachers likely have unanticipated consequences on relationships apart from the one that is the focus of the actions--grouping students for teaching purposes probably affects the relationship between self-concept and achievement, for example. Likewise, the nondeliberate actions of schools and teachers that stem from who and what they are may well alter the relationship between certain student attributes and achievement-- self-fulfilling prophecies are an example. Additionally, the relationship between student attributes and achievement may well vary between schools independently of what schools themselves may do, simply as a function of differential selection of student populations.

It is reasonable to suppose as well that the relationships most likely to vary across schools as a result of what schools and teachers do are those involving readily identifiable attributes of students: gender or minority status, for example, rather than less visible attributes like family composition and parental education. Student gender and race/ethnicity are seen to be related to achievement and, at times, are the subject of policy action designed to change the relationship in question.

Given the prominence of racial/ethnic differences in achievement as a fact of American education, and the programs of compensatory education designed to eliminate this relationship, it is reasonable to suppose that the relationship between minority status and reading achievement will vary across classrooms. With the exception of student socioeconomic background, arguments to support such variation in other level 1 relationships are not well established. And it is not possible to consider all slopes random as the number of variance-covariance components to be estimated increases rapidly with increase in the number of random slopes (Bryk and Raudenbush 1992, 202). Given that we have only 25 students per classroom, the limit is probably two or three random slopes.

14.4.4. Searching for Random Slopes

In the light of this uncertainty, we resort to exploratory analyses to establish which slopes should be treated as constant across schools and which should be designated random. Following Bryk and Frank (1991) the 15 level 1 variables are assigned to five substantively meaningful groups in order to explore the variation in slopes by focusing, one set at a time, on the limited number of variables within a group. These groups are identified in Table 14-2, along with information on the form of the variables used in the analyses reported below.

As a second step, separate models are estimated, one for each group. In each of the five models the three level 1 variables defining the group in question are treated as random, and the remaining variables are fixed. And in each instance three corresponding classroom means are introduced into the level 2 model for the intercept in order to reduce the impact of between-classroom variance on the level 1 estimates.

Further, in specifying these models a distinction is made according to whether the three variables in question are exogenous or endogenous variables. In the case of models 1, 2, and 3, which involve only exogenous variables, each model involves three random slopes and six fixed slopes. In the case of models 4 and 5, which involve endogenous variables as well, the general approach is to consider the three endogenous variables in question as random, while the remaining 12 variables are fixed. However, to the extent that we identify random slopes among the exogenous variables in models 1, 2, and 3, these are carried forward into models 4 and 5 with the consequence that the number of random slopes is increased and the number of fixed slopes decreased.

By way of illustration, in model 1, age, gender, and minority status are treated as random, while father's education, mother's education, family wealth, family composition, augmented family, and language are fixed. Additionally, the classroom means for age, gender and minority status are introduced into the level 2 model for the intercept. In model 2, father's education, mother's education, and family wealth are treated as random, the remaining six level 1 exogenous variables are fixed, and the classroom means for father's education, mother's education, and family wealth are introduced into the level 2 model for the intercept.

In the case of model 4, this same logic dictates that meals, literacy possessions, and study place are treated as random, while age through language plus TV watching, homework help, and parental interest are fixed. (As it turns out, minority status is identified as a random slope in the analyses estimating models 1, 2 and 3, so it is carried forward as a random slope into model 4 resulting in 4 random variables and 11 fixed variables.)

To identify random slopes, the basic logic adopted was to ask whether specifying a slope as random provides a better fit of the model to the data than simply assuming the slope to be fixed across classrooms. Thus, for each of the five models, two equations were estimated. In the first of these--a kind of null model--all slopes are treated as invariant across classrooms. In the second equation the slopes of

Table 14-2. Variable names, descriptions, and scales for the level 1 model

Description	Continuous/Dichotomous	Scale
Measures of Reading Comprehension		
1. narrative comprehension	Continuous	IRT score
2. expository comprehension	Continuous	IRT score
3. document comprehension	Continuous	IRT score
Model 1: Student Attributes		
4. age	Continuous	months
5. gender (sex)	Dichotomous	0=(male), 1=(female)
6. minority status (minor).......	Dichotomous	0=(white+Asian), 1=(black+Hispanic+American Indian)
Model 2: Parental Status Attainments		
7. father's education	Dichotomous	0=(HS grad or less), 1=(HS grad+further edn)
8. mother's education	Dichotomous	0=(HS grad or less), 1=(HS grad+further edn)
9. family wealth	Continuous	factor score
Model 3: Family Attributes		
10. family composition	Dichotomous	0=(<2 biological parents present), 1=(2 biological parents present)
11. augmented family	Dichotomous	0=(nuclear family), 1=(nuclear family plus others)
12. language	Dichotomous	0=(English is first & home language), 1=(other)
Model 4: Family Provision		
13. regular meals	Dichotomous	0=(<3 regular meals/day), 1=(3 regular meals/day)
14. literacy possessions	Continuous	factor score
15. place to study	Dichotomous	0=(no), 1=(yes)
Model 5: Family Interaction		
16. TV watching	Continuous	hours/day
17. help with homework	Continuous	factor score
18. parental interest	Continuous	factor score

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

the three variables that define the "random group" are treated as random, while the remaining slopes remain fixed. In both equations the classroom means for the variables in question are introduced into the level 2 model for the intercept. These two equations are identified respectively as "fixed" and "random" in the following discussion.

Consider the following example. In the case of model 1, which explores the question of random slopes among the three student attribute variables (age, sex, minority status), these two equations are specified as follows. In the fixed equation the slopes for all nine predictors (age, gender, minority status, father's education, mother's education, family wealth, family composition, augmented family, and language) are treated as invariant across classrooms, and classroom means for age, sex, and minority status are included in the level 2 model as predictors of the intercept. The random equation for this model differs only in that the slopes for age, sex, and minority status are allowed to vary across classrooms.

The basic approach then is to compare the fit of the two models using the variance-covariance components test provided within HLM (Bryk et al. 1988, 69). Unless the fit of the random model is significantly better than that of the fixed model, and/or unless the parameter variance statistics show significant variance in one or more of the slopes, we adopt the most parsimonious model and consider that the slopes in question can be seen as invariant across classrooms. If the random model shows a significant improvement in fit over the fixed model, then one or more of the slopes in question can be seen as a candidate for inclusion in the final level 1 model as a random slope.

Even if the random model does not show a statistically significant improvement in fit over the fixed model, since this is a global test it may well obscure significant variation in the slope of one variable. To address this issue, the parameter variance statistics for the random model are examined to determine if significant variation in the slope of any of the three parameters exists. If so, then this parameter too becomes a candidate for inclusion as a random slope in the final model. We emphasize the word "candidate." Since these are not all-inclusive models at this point, and because we are developing these analyses in an exploratory manner, the final decision about which slopes are to be treated as random is essentially a judgmental one to be informed by these findings and by other more theoretical/substantive considerations.

14.4.5. Identifying Random Slopes

This model development process was applied independently to each of the three reading comprehension measures in each of the two grade samples. The results of these exploratory analyses are summarized in Table 14-3.

Thus, in the case of model 1 for fourth grade students the overall global test of difference in fit between the fixed and the random model was statistically significant for narrative comprehension, but not for expository and document. That is, in the case of narrative comprehension, having age, sex, and minority status as random slopes resulted in a significant improvement in fit over a model in which all slopes were fixed. In the case of expository and document, the random slope model offered no improvement in fit over the fixed slope model. And, in only one other instance--model 2 for narrative--was the random slope model an improvement over the fixed slope model. For the ninth grade sample a random slope model offered a better fit than the fixed slope model only in two cases, both concerning document comprehension--model 1 and model 4. Where the univariate tests of the significance of parameter variance are concerned, only 3 of the 18 tests of fit suggest that a model involving random slopes offers much more than one in which all slopes are fixed across classrooms.

Statistics on parameter variance were used to temper the conclusions derived from the overall variance-covariance components test. These statistics point to where significant amounts of between-classroom variation in individual slopes exist. Here, too, evidence to support a random slope model is marginal. Given the number of instances in which there are statistically significant amounts of parameter variance, and the consistency of this occurrence across scales, one could reasonably argue that all slopes could be treated as fixed and that the level 2 model should focus only on between-classroom variation in intercepts.

Table 14-3. Random slope analyses: Summary, grades 4 and 9

Variable	Grade 4						Grade 9					
	Test of fit of model			Parameter variance			Test of fit of model			Parameter variance		
	Narrative comprehension	Expository comprehension	Document comprehension	Narrative comprehension	Expository comprehension	Document comprehension	Narrative comprehension	Expository comprehension	Document comprehension	Narrative comprehension	Expository comprehension	Document comprehension
Model 1: Student Attributes												
4. age				+	+	-				-	-	-
5. gender	+	-	-	-	-	-	-	-	+	-	-	-
6. minority status				+	+	+				+	-	-
Model 2: Parental Status Attainments												
7. father's education				-	-	-				-	-	+
8. mother's education	+	-	-	-	-	-	-	-	-	-	-	-
9. family wealth				+	-	-				-	-	-
Model 3: Family Attributes												
10. family composition					-	-				-	-	-
11. augmented family	-	+	+	-	-	-	-	-	+	-	-	-
12. language				-	-	-				-	-	-
Model 4: Family Provision												
minority status				+	-	-				+	-	+
13. regular meals	-	-	+	-	-	-	-	-	+	-	-	+
14. literacy possessions					-	-				-	-	+
15. study place				-	-	-				+	-	+
Model 5: Family Interaction												
minority status				+	-	-				-	-	-
16. TV watching	-	-	-	-	-	+	-	-	-	-	-	-
17. help with homework				-	-	-				-	-	+
18. parental interest				-	-	-				-	-	-

KEY: + = statistically significant (x = .05)
- = not statistically significant

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

However, the extent to which there is consistency in these statistics points to the "randomness" of the minority status effect, a fact that fits with theoretical expectation. It was, therefore, judged sensible -- and defensible -- to settle on a level 1 model for the exogenous variables in which the slope of minority status was considered to vary across classrooms. By contrast, the effects of age, gender, father's education, mother's education, family wealth, family composition, augmented family, language, and the endogenous family environment measures are considered invariant across classrooms. Further, because these are exploratory analyses of data, and in the absence of any contrary theoretical/substantive arguments, we have opted for this one model to apply to all three reading comprehension scales in each of the two populations.

While exploratory analyses of this kind are not entirely desirable and raise the likelihood of capitalizing on chance variation in the data, they complement theoretical/substantive speculations in what is a somewhat unstructured research design involving large numbers of variables. The final decision about random slopes is a judgmental one, but one informed by scientific common sense and the data.

14.5. Estimating the Student-Level Model

As noted earlier, the level 1 model was characterized by two equations. The first uses just the exogenous variables to establish the total net effect of each on reading comprehension. The second introduces the endogenous variables into the equation to provide estimates of the direct effect of each variable. In the case of the exogenous variables, the difference between the two estimates represents that part of the total effect that is transmitted via the endogenous variables.

These two equations are estimated within each classroom under a set of constraints that assumes the effects of all level 1 variables other than student minority status to be identical in all classrooms; allows for separate estimates of the effect of minority status on reading comprehension in each classroom and, hence, allows for variation across classrooms; allows the average achievement of classrooms to vary; and, includes "proportion of minority students" as a level 2 classroom variable predicting the classroom intercept for reading achievement in order to better estimate the minority-achievement slope within classrooms (by controlling for between-classroom variance in student minority status). Thus, in all we estimate 12 equations, 2 for each of the three measures of reading comprehension at each of the two grade levels.

Estimates for the 12 equations are presented in Tables 14-4 and 14-5. Each table is divided into two panels: the upper panel shows estimates for the exogenous equation; the lower, the estimates for the all-variable equation. Each panel contains three kinds of coefficients distinguished, on the one hand, by whether they refer to the level 2 (classroom) model or the level 1 (student) model and, on the other, by whether they vary across classrooms or are fixed. In the level 1 model the coefficient for minority status is random, while those for the remaining variables are fixed.

We display metric coefficients, standard errors, and standardized coefficients for each variable in each of the three equations. Parameters significantly different from zero (at a nominal a=.05) are shown in boldface type under the column for metric coefficients; those not significantly different from zero are shown in italic type. In this same column we also show the proportion of within-class variance explained by these student-level variables. In the case of standardized coefficients, only values for continuous variables are shown. Since it makes little interpretative sense to standardize dichotomous variables, we have not done so and these coefficients are omitted from the standardized coefficient column.

Each equation offers five basic types of information.

a. An estimate of the grand mean for each reading comprehension scale--the intercept--a weighted average of the classroom means.

b. An estimate of the effect of the proportion of minority students in the class on the average achievement level of the class--an apparent compositional effect that we will explore a little further on in the discussion.

Table 14-4. Student-level model: Grade 4

Variable	Exogenous equation								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Intercept	**556.2**	2.0	556.2	**540.5**	1.8	540.5	**551.9**	1.6	551.9
Level 2 Model									
propn. minor. S's in class	**-65.9**	6.7	-19.7	**-59.6**	5.9	-17.8	**-61.0**	5.4	-18.3
Level 1 Model									
Random Effects									
minority status	**-16.0**	3.9		**-12.6**	3.1		**-18.0**	3.2	
Fixed Effects									
age	**-2.0**	.2	-12.9	**-1.5**	.1	-10.0	**-1.1**	.1	-7.1
gender	**14.7**	2.3		**6.4**	1.9		*-2.8*	1.9	
father's education	**9.1**	2.8		**10.8**	2.3		**6.3**	2.3	
mother's education	*4.6*	2.8		*3.4*	2.3		**7.6**	2.3	
family wealth	**6.4**	1.3	6.0	**6.3**	1.0	5.9	**6.8**	1.1	6.4
family composition	**15.4**	2.4		**11.1**	2.0		**18.8**	2.0	
augmented family	**-20.3**	2.5		**-16.6**	2.0		**-15.4**	2.1	
language	**-9.5**	2.8		-7.0	2.3		**-12.2**	2.3	
Variance Explained (%)	**8**			**6**			**7**		
	All-variable equation								
Intercept	**556.0**	1.9	556.0	**540.4**	1.7	540.4	**551.8**	1.6	551.8
Level 2 Model									
propn. minor. S's in class	**-58.7**	6.6	-17.6	**-53.7**	5.8	-16.1	**-55.8**	5.3	-16.7
Level 1 Model									
Random Effects									
minority status	**-13.8**	3.9		**-10.6**	3.0		**-16.4**	3.1	
Fixed Effects									
age	**-1.9**	.2	-12.4	**-1.4**	.1	-9.5	**-1.0**	.1	-6.7
gender	**15.6**	2.3		**7.2**	1.9		*-1.6*	1.9	
father's education	**8.2**	2.8		**9.9**	2.3		**5.6**	2.3	
mother's education	*3.7*	2.8		*2.8*	2.3		**6.7**	2.3	
family wealth	*-.4*	1.7	-.4	*.6*	1.4	.6	*-.3*	1.4	-.3
family composition	**14.8**	2.4		**10.5**	2.0		**18.4**	2.0	
augmented family	-19.0	2.4		**-15.4**	2.0		**-14.4**	2.0	
language	**-9.0**	2.8		**-6.5**	2.3		**-12.1**	2.3	
regular meals	**10.0**	2.4		**11.8**	2.0		**5.5**	2.0	
literacy possessions	**12.2**	1.8	10.6	**10.2**	1.5	8.9	**12.3**	1.5	10.7
study place	**-10.0**	2.7		**-12.8**	2.2		**-6.1**	2.2	
TV watching	**-1.8**	.6	-11.9	**-1.1**	.5	-7.6	*-.7*	.5	-4.5
homework help	**-4.3**	1.5	-3.5	**-2.8**	1.2	-2.2	**-4.6**	1.2	-3.8
parental interaction	**-5.1**	1.6	-3.8	-3.0	1.3	-2.3	-5.0	1.3	-3.8
Variance Explained (%)	**9**			**8**			**9**		

NOTE: In columns presenting metric coefficients, numbers in boldface indicate parameters significantly different from zero; numbers in italics are not significantly different from zero.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 14-5. Student-level model: Grade 9

Variable	Exogenous equation								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Intercept	**538.5**	3.9	538.5	**539.7**	4.3	539.7	**528.0**	3.1	528
Level 2 Model									
propn. minor. S's in class	**-102.3**	13.3	-30.6	**-112.4**	14.8	-33.6	**-85.5**	10.5	-25.6
Level 1 Model									
Random Effects									
minority status	**-22.4**	5.3		**-27.1**	5.4		**-27.4**	4.2	
Fixed Effects									
age	**-1.5**	.2	-11.1	**-1.5**	.2	-11.8	**-1.4**	.2	-10.4
gender	**16.9**	2.9		*-2.6*	3.2		**-6.1**	2.5	
father's education	**8.7**	3.3		**8.1**	3.6		**8.5**	2.9	
mother's education	*5.0*	3.3		**11.4**	3.6		*1.9*	2.9	
family wealth	**3.0**	1.5	2.8	*-1.2*	1.7	-1.2	*-1.2*	1.3	-1.1
family composition	*1.0*	3.1		*-.9*	3.4		*-1.9*	2.7	
augmented family	*-3.3*	4.0		*-7.3*	4.4		*-6.2*	3.5	
language	**-9.5**	3.7		*-4.3*	4.1		**-7.5**	3.2	
Variance Explained (%)	**5**			**3**			**4**		
	All-variable equation								
Intercept	538.7	3.8	538.7	**539.9**	4.2	539.9	**528.2**	2.9	528.2
Level 2 Model									
propn. minor. S's in class	-95.6	13.0	-28.6	**-104.0**	14.5	-31.1	**-77.7**	10.1	-23.2
Level 1 Model									
Random Effects									
minority status	**-19.5**	5.2		**-23.8**	5.3		**-24.3**	4.2	
Fixed Effects									
age	**-1.4**	.2	-11.0	**-1.5**	.2	-11.4	**-1.3**	.2	-10.2
gender	**16.4**	3.0		*-1.2*	3.3		**-5.7**	2.6	
father's education	**8.5**	3.3		**7.3**	3.6		**7.8**	2.9	
mother's education	*4.6*	3.3		**9.9**	3.6		*.9*	2.9	
family wealth	**-5.7**	2.4	-5.5	**-14.2**	2.6	-13.6	**-11.6**	2.1	-11.2
family composition	*3.0*	3.1		*.3*	3.4		*-.4*	2.7	
augmented family	*-3.6*	4.0		*-6.5*	4.4		*-5.8*	3.5	
language	**-10.5**	3.7		*5.2*	4.1		**8.2**	3.2	
regular meals	*-2.7*	3.1		*6.2*	3.4		*1.5*	2.7	
literacy possessions	**12.1**	2.5	11.1	**18.4**	2.8	16.9	**14.8**	2.2	13.6
study place	*-3.3*	2.9		**-9.1**	3.2		**-7.5**	2.6	
TV watching	**-2.7**	.8	-4.9	**-2.4**	.9	-4.3	**-2.8**	.7	-4.9
homework help	**-4.8**	1.9	-3.8	*-2.5*	2.1	-2.0	*-2.6*	1.6	-2.1
parental interaction	**-4.5**	-1.6	-4.0	**-4.8**	1.8	-4.3	**-5.3**	1.4	-4.7
Variance Explained (%)	**6**			**5**			**6**		

NOTE: In columns presenting metric coefficients, numbers in boldface indicate parameters significantly different from zero; numbers in italics are not significantly different from zero.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

c. The minority-achievement slope, a student-level relationship that varies across classrooms--the coefficient shown is a weighted average of some 300 coefficients in the fourth grade sample and 160 in the ninth grade sample.

d. Fixed coefficients representing the effects of student-level attributes on achievement, effects assumed not to vary across classrooms.

e. The proportion of variance in reading comprehension explained by each equation.

14.6. Interpreting the Level 1 Model

In analyses such as these, two basic kinds of interpretations are feasible. First, it is possible to talk about how well the model explains the between-student variation in reading comprehension within classrooms. Assuming that a good part of this between-student variation is potentially explainable and not due simply to random events, then the amount of variance explained tells us something about the adequacy of the model. Second, we can talk about the effects of student attributes on reading comprehension, effects that have been adjusted for the confounding influence of other variables in the model and that allow an "other things equal" interpretation.

14.6.1. Variance Explained

On the matter of the explanatory power of these models, basically they do not explain a great deal of the variance in any of the three measures of reading comprehension at either grade level. As one might, expect somewhat more of the variance in reading comprehension is explained among younger students using the family-related variables noted, but in each case it is less than 10 percent for the exogenous variables. This is about what one would expect using the variables at hand. However, unless we believe that reading comprehension is inherently inexplainable in the main, then one cannot dismiss the possibility that the model is mispecified by the omission of important variables.

Further, the addition of the endogenous variables to the equation adds little in the way of explanatory power--1 to 2 percentage points at the most. On the surface, this finding is somewhat surprising. One might expect the family process variables to add considerably to variance explained by status variables like those in the exogenous equation. Given that the family process variables in question do not, then we are led to the conclusion that the process variables included in the research design were inappropriate for the development of reading comprehension, inappropriate for reading comprehension at these two ages, and/or poorly represented by the indicators used.

In fact, the literature might lead us to the second position, at least with respect to the parent-child interaction variables. While there is a good deal of research on the salutory effects of parent-child interaction on the reading achievement of children, most of the work relates to children in the preschool and elementary grades (Phillips and Bolt 1992). There seems to be little evidence on the aspects of family environments that affect the reading comprehension of students in the middle grades. They may be the same, but one would guess that the emphasis on reading aloud to parents, which figured prominently in the Student Questionnaire, may have been misplaced with students of this age.

14.6.2. Interpreting Coefficients as Measures of Effect

Structural equation models of this general form are reasonably common in the literature, and are usually estimated with Ordinary Least Squares (OLS) multiple regression to provide partial regression coefficients interpretable as measures of effect. In the present case, the same kinds of interpretations of coefficients apply.

At this point, the interpretations provided are illustrative and quite literal, based on the exogenous variables equation for narrative comprehension in the fourth grade. There seems no point in going further at this stage as the level 2 model is not yet developed. We go into more detail about coefficients and their meaning at a later point in connection with the combined (and refined) model that results from these exploratory analyses.

The coefficients in question are the left-most set in the upper panel of Table 14-4. We follow these with interpretations of the coefficients for the endogenous variables, the coefficients in the left-most set in the lower panel of Table 14-4.

14.6.2.1. Mean Achievement

The estimate of 556.2 for the intercept of the exogenous variable equation for narrative comprehension in grade 4 can be seen as a weighted average of classroom averages. The interpretation of this coefficient depends on the centering of the variables in the equation. The fixed student-level variables (age through language) and the level 2 variable (proportion of minority students in class) were grand-mean centered. That is, each student's score on these variables was converted into the deviation from the mean of all students, and the proportion of minority students in each classroom was transformed into a deviation from the mean of the proportions of minority students in each of the several hundred classrooms. By contrast, the student attribute whose effect was considered to vary across classrooms--minority status--was group-mean centered; that is, each student's score was expressed as a deviation from the mean score of the students in the classroom where the student was located. It follows then that, since minority students are coded 1 and nonminority students 0, the intercept can be seen as the mean across classrooms of the average score of nonminority students in each classroom adjusted for differences between students overall in the student attributes age through language, and differences between classrooms in the proportion of minority students in the class.

14.6.2.2. Proportion of Minority Students in the Class

This is a level 2 variable, a characteristic of classrooms rather than individual students. It is included as a level 2 predictor of the intercept when estimating the level 1 equations in order to control for between-classroom variation in minority status when estimating the within-classroom minority-achievement slope. In other words, we are controlling for potential between-classroom compositional effects of minority status that stem from the proportion of minority students in the classroom. Though the study of compositional effects is interesting in its own right, in the present case we introduce this variable into the equation for the intercept in order to get a better estimate of the effect of student minority status on achievement. However, the difference between the between-classroom coefficient (level 2) and the average within-classroom coefficient (level 1) indicates the size of the compositional effect (Bryk and Raudenbush 1992, 121). In this case, it is not an effect to be ignored. The difference is quite substantial, amounting to a 50-point gap (-65.9 - (-16.0)). This is more than half a standard deviation for classrooms at the two extremes -- all majority versus all minority. We take up this matter again later in the discussion.

14.6.2.3. Minority Status-Achievement Slope

Minority status is a dichotomous variable (nonminority=0, minority=1), and in the model the effect of this attribute of students is assumed to vary across classrooms. Since the measure is dichotomous, we can think of the effect as the average achievement gap between minority and nonminority students adjusted for the differences between students in the student/family characteristics noted, and for the proportion of minority students in the class. This achievement gap varies across classrooms and averages out as -16; other things equal, in the average classroom, minority students score some 16 points below nonminority students on this test of narrative comprehension.

14.6.2.4. Age

Age is the first of eight fixed-effect variables to be considered, and one to which it is difficult to assign an unequivocal meaning. The variable captures, in some unknown mix, developmental effects, opportunity to learn, and the effects of grade retardation, themselves a mix of the effects of ability and disadvantage. Since the unit of measurement in this case is months, the -2 value shown for the coefficient indicates that, other things equal, with each month increase in age the average score of students in fourth grade classrooms declines by two points. The effect is small though statistically significant. The negative sign is almost certainly indicative of grade retardation in the sense that the older students in the class are less capable readers who have not been promoted on a yearly basis.

14.6.2.5. Gender

The measure of gender is scored as 0/1 with females shown as 1. Thus, the effect shown indicates that, other things equal, on the narrative comprehension measure females score some 15 points on average higher than males, a finding consistent with what is known about gender differences in reading and in verbal ability at this age.

14.6.2.6. Parental Education

Both father's education and mother's education are treated as dichotomies, scored 0 for persons with no more education than high school graduation, and 1 if education goes beyond high school, including college. Thus, other things equal, fourth grade students whose fathers have more than a high school diploma have average narrative comprehension scores some 9 points higher than those whose fathers have lower levels of formal education. In the case of mother's education the results are mixed overall, but for narrative comprehension they indicate that, other things equal, the level of mother's education has an effect about half that of father's education, and is not statistically significant.

14.6.2.7. Family Wealth

Family wealth is an indirect measure based on family possessions and was created as a factor score. Since this score is standardized to a metric of standard deviation units, the unit of measurement is a standard deviation. This means that, other things equal, fourth grade students from families one standard deviation apart in wealth are themselves some 6.4 points apart on average in their narrative comprehension scores. Since the standard deviation of the narrative achievement scale scores is close to 90, it is clear that for fourth grade students effects due to the economic circumstances of families are not large.

14.6.2.8. Family Composition

Family composition is a dichotomy within which students reporting the presence of a mother and father were coded 1, and all others were coded 0. The coefficient of 15.4 is sizable, relatively speaking, and indicates the net advantage of having both parents present in the family. This achievement gap is of the same order as the minority status gap.

14.6.2.9. Augmented Family

This variable measures whether or not persons other than members of the nuclear family are present in the household--grandparents, other relatives, and/or other persons. Families in which others are present are coded 1, and nuclear families, 0. Thus, the effect of -20 points indicates the cost, other things equal, of living in an augmented rather than nuclear family. Keep in mind that this effect of augmented family is not confounded by parental education, family wealth, minority status, and language, since these are controlled statistically in arriving at this estimate. As such, it represents an effect of family structure and composition over and above that due to the low levels of status attainments traditionally associated with such household arrangements.

14.6.2.10. Language

The language variable makes a distinction between those students whose first and home languages are English, and all others. The latter group is coded 1, the former 0. As a consequence, the effect of -10 indicates the reading comprehension disadvantage that accrues from having a first and/or home language other than the language of the test.

14.6.2.11. Regularity of Meals

At this point attention is directed to the coefficients in the all-variable equation, the left-most set in the lower panel of Table 14-4. The coefficients for the meals variable and those that follow refer to endogenous variables and represent a measure of the direct effect (which is also the total effect) of the endogenous variable on reading comprehension. It is the effect net of all the other variables in the equation.

Regular meals is a dichotomous variable. The coefficient of 10 points offers the interpretation that, other things equal, those students who receive three regular meals per day outperform, by some 10 points on average, those who have less regular meals. In the interpretation of this coefficient we may be looking at a nutritional effect, but more likely the variable taps parental concern and interest in the student's well-being.

14.6.2.12. Three Apparently Counterintuitive Effects

For convenience we consider the meaning of the coefficients for study place, homework, and parental interaction at the same time. All three effects are negative, and all three are counterintuitive in the sense that we would normally expect that more of any of these would be associated with higher levels of reading comprehension. As it is, "more" is associated with lower levels of comprehension.

Other things equal, students with a specific place to study show lower levels of narrative comprehension on average, some 10 points lower than those who report that they have no particular place in which to do their homework. Similarly, students whose parents read aloud to them at home and help them with their reading and students who receive more parental help with their homework in reading do not read as well as those who do not have this parental intervention.

We cannot be certain, of course, but these counterintuitive results are probably a further consequence of a family environment model specified for younger children and not really applicable to children in the middle grades. Presumably the explanation is that, by age 10, if parents are still providing for these kinds of interactions with their children it is because the students are not doing well in reading. That is, they provide behaviors appropriate to younger children because the students in question are reading at the level of competence of younger children. In short, parents read aloud to them at home, provide them with a special place to study, and help with reading homework, all with the view to improving the child's reading.

14.6.2.13. Hours of TV Watching

The effect of TV watching offers a more conventional interpretation. This variable retains its original metric of hours; thus, the interpretation of the coefficient is that, other things equal, each additional hour of TV watching reduces the average score of students by two points--a statistically significant but rather minor effect.

14.6.2.14. Literacy Possessions

Literacy possessions is a composite derived from reports of literacy-related possessions in the home or actually owned personally by the student. Encyclopedias, dictionaries, computers, and the like are the items in question. These items appear with some regularity in the IEA studies and in NAEP, and show a consistent relationship with school achievement generally, and with reading proficiency in particular. Like parental support and homework, this variable is a factor score. Since factor scores are standardized, they represent the effect of a one standard deviation difference between students. In the case of the effect of literacy possessions, this means that students one standard deviation apart in literacy possessions are, other things equal, some 12 points apart in their narrative comprehension scores. Having more of these possessions appears to enhance reading, but in a relatively minor way.

14.6.3. Family and Student Effects on Reading Comprehension

These same kinds of considerations apply across the models estimated for each of the three reading comprehension scores in each of the two grade populations. As the data in Tables 14-4 and 14-5 indicate, the similarity between the six models is considerable in terms of the overall pattern of effects, although there are some differences across the scales within and between the two grades. We comment first on the common aspects of the level 1 model, and then highlight pattern differences we see as having potential meaning.

Note, however, that it is not legitimate to make direct comparisons of the size of effects across scales within samples or within scale types between samples. Each of the six scale scores were derived separately and scaled arbitrarily to a mean of 500 and standard deviation of 100. Thus, the six achievement scores are not on the same scale, and the effects of family and student attributes on reading comprehension cannot be compared directly across the several equations.

Given the high correlations among the three scales, it is not too surprising to find that the patterns of effects are much the same from scale to scale. The interpretations of effects advanced above in connection with the narrative scale for fourth grade students apply equally well to expository and document comprehension at this grade level. Most of the family effects expected on the basis of past research into the academic achievements of elementary school students are to be found here as well. Students from higher status and wealthier families do better, children from two-parent families have higher achievements, and children not burdened by working in a new language and a new culture are similarly advantaged. Gender differences also work in the expected direction, favoring females when the comprehension material is heavily verbal, as in the narrative scale, but disappearing as the verbal content decreases to the level found in the document scale. The one student-level effect we might not have anticipated is that due to augmented family, though possibly this has similar origins to the negative effect of family size often reported in the literature. The more children have to share the attention of their parents with others (other adults in this case), the lower their achievements. However, multifamily and/or multigeneration households are more characteristic of some subpopulation groups than others. We may be seeing as well reflections of social, economic, ethnic, and perhaps rural-urban differences not captured in the parental education, family wealth, and ethnicity variables included explicitly in the equation.

Minority status remains a disadvantage for the development and display of reading comprehension skills among fourth graders, even after adjustment for the status attainments of parents and other attributes of students' families. However, it is not markedly different from a number of the other effects shown. Other things equal, being a minority student is about as much of a disadvantage as being without both parents, or growing up with a language other than English--or being male when the tests are loaded heavily with verbal items.

Among ninth grade students the effects of family attributes seem to be of less importance, as one might expect. Table 14-5 provides estimates for all six equations. These variables explain less of the variance in reading comprehension and, overall, the effects appear to be much smaller. This is accounted for, in part, by the fact that 14-year-olds are subject to a wider variety of influences than those arising within their families. It may be a function as well of the fact that the somewhat suspect family process model for the younger students was adopted without change for the older group, where it would be even more suspect. The items in each of the two questionnaires are virtually identical. Overall, these data did not allow the development of a particularly persuasive model of the influence of families on the development of reading comprehension skills.

14.7. Compositional Effects

As the name suggests, compositional effects (often called contextual effects or structural effects) arise from the composition of the classroom group in which these students find themselves. The basic argument rests on the notion that, for example, a student's educational ambitions are affected by his/her achievement and, if a compositional effect is present, by the aggregate achievement level of the classroom as well. Other things equal, students in high achieving classrooms are likely to be more ambitious because high achieving classrooms have high aggregate levels of ambition, and the class acts as both a normative and comparative reference group for the student (see Kelley 1952; Kemper 1968). Another version of this argument suggests that the effect of student achievement on ambition is conditioned by the aggregate level of achievement in the classroom. This is the so-called relative deprivation (Stouffer 1949) or "frogpond" argument (Davis 1966). High achieving students in high achieving classrooms are likely to have lower levels of ambition than students with comparable levels of achievement in low-achieving classrooms--small frog, big pond versus big frog, small pond, respectively.

We noted earlier the matter of the inclusion of the proportion of minority students in the class as a level 2 variable in these models. We noted as well that the effect of such a variable could be interpreted as a compositional effect--in this case, an effect on the average achievement of the class and/or an effect on the minority status- achievement slope, and hence, on the level of student achievement. Since these compositional effects are consistent across all the equations estimated to date and those that will be presented later, and because they appear to be sizable, they cannot be ignored. However, we need to be clear about what they mean and whether we can legitimately ascribe the effects shown to the proportion of minority students in the class.

In the equations estimated the individual-level effect of minority status is group-mean centered, while the group-level measure is grand-mean centered. Simple subtraction of the student-level effect from the classroom-level effect gives us the compositional effect (Bryk and Raudenbush 1992, 121). With reference to the equations displayed in Tables 14-4 and 14-5, the compositional effect amounts to somewhere between 40 and 50 points--close to half a standard deviation. Note, however, that we have measured the minority composition of the class as the proportion of minority students in the class. Thus, the range of variation in the variable is from 0 (no minority students) to 1 (all minority students). It follows that a compositional effect of 50 points represents the difference, other things equal, between the average achievement of a class in which there are no minority students and the average achievement of a class in which all students are minority group members. Another way of looking at this effect is to recalculate it as the effect of each 1 percent increase in minority students. Under these conditions, the compositional effect can be interpreted as other things equal, each 1 percent increase in minority students reduces the class average by about 0.5 of a point.

One might want to worry about the size of this effect from another perspective, in the sense that minority status tends to be associated with lower levels of status attainments. The observed effect may represent, at least in part, a mix of social, educational, and economic status compositional effects. We looked into this proposition by estimating a parallel set of models to those reported in Tables 14-4 and 14-5. In these analyses we included, along with the measure of class minority composition as a predictor of the intercept, aggregate measures of parental education and family wealth--respectively, the proportion of students whose fathers had completed college, the proportion of students whose mothers had completed college, and the mean level of family wealth for the students in the class.

The results of these analyses suggested that the compositional effect due to the proportion of minority students in the class was reduced by the inclusion of the status attainment variables as classroom aggregates, but it was not eliminated. Under these conditions the size of the minority status compositional effect is about 30 points--a third of a standard deviation difference between classes in which there are no minority students and those in which all students are members of minority groups, other things equal. Note, however, that the other things held equal in these models are only the level 1 variables. It is almost certain that a number of community, school, classroom, and teacher attributes vary with the minority composition of classrooms and that, in fact, the effect shown is overestimated. We take up this matter later in connection with the development of level 2 models.

14.8. Formulating Classroom-Level Models

Student-level models are not the main interest of IEA studies. Rather, once the comparisons of national achievement scores have been made, interest turns to the way in which schools and their teachers influence student achievement. The effects of policy variables at the level of classroom, school, and community remain a dominant interest. In particular, there is interest in the effects of resource variables-- physical resources (e.g., libraries, computers) as well as human resources in the form of teacher education and training -- and their application in the form of instructional practices. This interest is driven by the notion that, since these resources are alterable by educational policy, if a resource effect can be identified, then educational policy can be used to change educational practice and, eventually, the educational achievements of students in the aggregate.

In short, the ideal is that IEA studies would be able to demonstrate such observations as, "other things equal, an infusion of library books into classrooms will raise the level of achievement of the students in those classrooms"; or "better trained teachers produce higher levels of achievement"; or "students excel in classes where teachers give 10 hours of homework a week"--a sort of international "what works." This very worthwhile cause is usually predicated on analyses that seek to account for between-classroom/school differences in average achievement levels in terms of parallel physical and human resource differences, or differences in teaching behaviors and strategies.

The same questions focus our development of the level 2 models that attempt to explain between-classroom variation in average reading comprehension, together with the effect of minority status on achievement, in terms of between-classroom differences in context, resources, and pedagogy. The basic dimensions of these contexts, resources, and instructional behaviors, as implied by the data, were outlined in Figure 14-3. Six categories of context/resource/instruction variables were defined, and logic suggests that these categories can be arrayed as indicated to distinguish groups of variables we see as exogenous from those that are endogenous. Community context, school, and principal are treated as groups of exogenous variables that influence the attributes of classrooms and the attributes of the teachers that make up the schools' work forces. All of these, in turn, are seen as influences on the kinds of instructional practices teachers use in classrooms to promote reading comprehension.

In the first instance this model is more an organizing principle than a model to be estimated. While the classification of these variables (Figure 14-1) introduced a measure of theoretical/substantive order, and we have conjectured a little further about the broad structure of the underlying models in Figures 14-2 and 14-3, it would not make a lot of sense to estimate the model as it stands. It contains 43 predictors, for one thing. Since the first cut at a conceptual organization of the data was only a classification of variables into relatively homogeneous categories, it is likely that there is redundancy within these categories. If so, and if we went ahead to include all the variables within the one equation, collinearity among the variables would likely obscure effects of interest. Besides there are simply too many variables in the model for it to make much sense.

Table 14-6 provides a summary statement on the level 2 variables under consideration.

Table 14-6. Classroom-level variables and measurement: Grades 4 and 9

Variable	Continuous/Dichotomous	Scale
Compositional		
1. proportion of minority students in class	Continuous	proportion
Community Context		
2. urbanicity	Continuous	community size
3. region	Dichotomous	dummy variables (NE, SE, Central)
4. community resources	Dichotomous	(0=low) (1=high)
5. parental cooperation	Continuous	rating by principal
School Attributes		
6. public/private	Dichotomous	(0=public) (1=private)
7. school size	Continuous	pupil population size
8. hours instruction/week	Continuous	hours
9. proportion of minority students	Continuous	minority teachers/all teachers
10. proportion of specialist teachers	Continuous	specialists/all teachers
11. library books/student	Continuous	# books/# students
Principal Attributes		
12. gender	Dichotomous	(0=male) (1=female)
13. minority status	Dichotomous	(0=majority) (1=minority)
14. educational administration training	Continuous(4)	factor score
	Dichotomous(9)	(0=low) (1=high)
15. reading training	Dichotomous(4)	(0=low) (1=high)
	Continuous(9)	factor score
16. experience as principal	Continuous	years
17. experience this school	Continuous	years
18. instructional leadership	Continuous	factor score
19. staff development orientation	Continuous(4)	factor score
	Continuous(9)	(0=low) (1=high)
Classroom Attributes		
20. class size	Continuous	# of students
21. % remedial students in class	Continuous	remedial students/all students
22. % remedial students helped	Continuous	students helped/remedial students
23. library resources	Continuous	reading material/student
24. instructional time (all)	Continuous	hours/week
25. instructional time (reading)	Continuous	hours/week
Teacher Attributes		
26. gender	Dichotomous	(0=male) (1=female)
27. minority status	Dichotomous	(0=majority) (1=minority)
28. formal education	Dichotomous	(0=no higher degree) (1=higher degree)
29. teacher training	Continuous	
30. teaching experience	Continuous	years
Instructional Practices: Grade 4		
31. instructional orientation	Continuous	
32. teacher's reading preferences	Continuous	factor score
33. assessment practices	Continuous	
34. hierarchical instruction	Dichotomous	(0=low) (1=high)
35. use of grouping	Dichotomous	(0=none) (1=same)
36. instructional time (reading)	Continuous	hours
37. homework emphasis	Continuous	factor score
Instructional Practices: Grade 9		
38. construction of meaning	Continuous	factor score
39. extend meaning	Continuous	factor score
40. student directed learning	Continuous	factor score
41. teacher directed learning	Continuous	factor score
42. teacher's professional reading	Continuous	factor score
43. teacher's general reading	Continuous	factor score

NOTE: The numbers 4 and 9 in parentheses refer to grade 4 and grade 9.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

14.8.1. Separate Models for Each Category

In view of this situation the approach adopted was to develop, as a first step, separate models for each of the six level 2 categories of variables. The first of these models seeks to explain the between-classroom variation in average achievement, and in the minority-achievement slope, with the set of community context variables alone. A second model looks in the same way at the effects of the variables characterizing schools; a third model focuses on the characteristics of principals as influences on classroom achievement; and so on. This approach has been adopted by other investigators dealing with relatively large numbers of variables (for example, Bryk and Thum 1989; Arnold, Kaufman, and Sedlacek 1992). While probably unavoidable in the present circumstances, this form of model development has obvious limitations; the effects of school characteristics, for example, are considered apart from those contained within other categories of the model--communities, principals, teachers, and so on. Since we would not want to argue that these categories are orthogonal to each other, each of the six models is probably specified incorrectly. Nevertheless, in these submodels each variable gets its best chance to show an effect on reading comprehension, an effect that may not persist in any model combining elements of each of the six submodels.

14.8.2. Explaining Variation in Classroom Means and Classroom Slopes

Within each of these six models the variables explaining between-classroom variation in reading comprehension levels are specified a little differently from those that explain the between-classroom variation in slopes. In the case of the classroom means the selection of variables is fairly straightforward, as each variable included in the equation contributes only one additive term to be estimated. Given this, all variables were considered as potential influences on the variation in mean achievement across classrooms.

By contrast, each variable considered as an effect on the minority status-achievement slope adds a cross-level interaction term to the equation and makes the equation to be estimated more complex. The difficulty is compounded somewhat by the fact that we have little guidance from the literature or theory about what it is that affects minority status-achievement slopes. In view of this, and consistent with the general tenor of these analyses, we chose to identify the variables predicting the variation in slopes by an exploratory procedure. The approach taken was to use the approximate "t-to-enter" statistics produced by HLM through the regression of the empirical Bayes residuals on the variables specified (Bryk et al. 1988, 54). Any level 2 predictor exhibiting a sizable relationship with the minority-achievement slope was considered a candidate for inclusion as a predictor of this slope.

In operation, this means that the first level 2 equation estimated in each group specified predictors for the intercept but none for the random minority status-achievement slope. If an examination of the (zero-order) relationships between the residual for the slope and the variables included as predictors of the intercept suggested a sizable and meaningful effect on the minority status-achievement slope, these variables were included as slope predictors and the equation reestimated. The number of predictors identified in this way was small and inconsistent across equations, as will become clear from the following analyses. Basically, we were not able to say much about the sources of between-classroom variation in slopes.

14.8.3. Six Separate Classroom-Level Models

Although the tables that follow do not show the coefficients for the level 1 model estimated in each case, a level 1 model was specified and had the form of the all-variables equation in Tables 14-4 and 14-5. All level 1 variables with fixed effects were grand-mean centered, the single random slope for the effect of minority status was group-mean centered, and all level-2 variables were grand-mean centered.

Additionally, the compositional variable, proportion of minority students in the class, was included in all six models. It was treated as a predictor of the intercept, and as a potential predictor of the minority-achievement slope. The inclusion of this variable, and the fact that it has a sizable effect on reading comprehension, led to the development of two estimates of the variance explained by each equation. We estimated the total variance explained by all level 2 variables, together with the proportion of variance explained by level 2 variables other than the compositional variable. This allowed an examination of the explanatory power of the variables unique to the particular model-- that is, over and above the variance explained by the compositional variable.

Since these models represent developmental stages in the progress toward a combined model, as with the student-level model discussed above, we do not devote much space to the interpretation of particular coefficients at this point. This is undertaken in connection with the composite model to be discussed later.

14.8.3.1. Reading the Tables

Since Tables 14-7 through 14-12 and 14-15 and 14-16 have the same general format, it may be helpful if we provide an extended treatment of Table 14-7 as an example of how one reads the tables for the meaning of the findings. We refer to the models for narrative comprehension for the fourth grade in developing specific illustrations. Table 14-7 is divided into two panels, the upper one referring to fourth grade students and their schools, the lower to ninth grade students and their schools. In these panels, we report the results of estimating the classroom-level community context submodel, as shown, simultaneously with a student-level model as defined by the all-variable equation detailed in Tables 14-4 and 14-5, but not shown here.

Each panel contains the results of estimating six models, two for each of the three reading comprehension measures. One of these is the model for the variation in classroom means. The other is the model for the variation in minority status-achievement slopes across classrooms. The model for the classroom means contains a number of explanatory variables grouped under the heading "fixed effects" and is shown in the upper section of each panel. The model for the slopes contains only one predictor at the most (urbanicity, in this case) and is reported in the lower section of each panel, below the model for the classroom means.

In each model, we report estimates of the coefficients for the (in this case) community context variables specified. In the left-most column these are shown as metric coefficients; that is, the coefficients retain their original units of measurement. The coefficient of -50.0 for the proportion of minority students in the class is in "proportion units" and, as such, has a range from 0 to 1. Each of these coefficients is accompanied by an estimate of the standard error of the estimate -- 7.1 for this same coefficient. Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface type. Other coefficients are shown in italic typeface. Standardized coefficients are shown for each variable other than dichotomies. The metric for these coefficients is standard deviation units, and the estimate for proportion of minority students in the classroom is -15.0.

Table 14-7. Classroom-level model: Effects of community context, grades 4 and 9

Community context	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.1**	1.8	556.1	**540.5**	1.6	540.5	**551.9**	1.5	551.9
Fixed Effects									
propn. minor S's in class	**-50.3**	7.1	-15.0	**-44.2**	6.4	-13.2	**-50.5**	5.9	-15.1
urbanicity	*.3*	1.4	.4	*-.6*	1.3	-.9	*.2*	1.2	.3
region: Northeast	*1.5*	5.3		*1.6*	4.8		*2.7*	4.4	
region: Southeast	**-19.4**	5.2		**-12.1**	4.7		**-13.3**	4.4	
region: Central	**-11.7**	4.8		*-.9*	4.4		*-3.9*	4.0	
community resources	**12.7**	6.1		**11.2**	5.5		**10.4**	5.1	
parental cooperation	**10.0**	2.1	9.2	**7.5**	1.9	6.9	**4.6**	1.8	4.3
Variance Explained (%)									
total	**43**			**36**			**43**		
context	**20**			**11**			**12**		
Model for Minority Slope									
minority gap	**-14.8**	3.9	-14.8	**-10.7**	3.0	-10.7	**-16.4**	3.1	-16.4
urbanicity (unique)	**6.4**	2.9	8.9						
Variance Explained (%)	**2**			NA			NA		
	Grade 9								
Model for Class Means									
Intercept	**538.7**	3.8	538.7	**539.9**	4.2	539.9	**528.2**	2.9	528.2
Fixed Effects									
propn. minor. S's in class	**-93.6**	15.8	-28.0	**-107.2**	17.7	-32.1	**-81.9**	12.3	-24.5
urbanicity	*2.0*	3.2	2.9	*4.7*	3.6	6.8	**4.2**	2.5	6.0
region: Northeast	*12.2*	11.0		*8.4*	12.2		*4.6*	8.5	
region: Southeast	*.7*	10.3		*1.0*	11.5		*-5.4*	8.0	
region: Central	*8.3*	10.8		*12.8*	12.0		*5.6*	8.3	
community resources	*3.1*	8.7		*-6.5*	9.7		*-3.6*	6.7	
parental cooperation	*1.3*	5.2	1.1	*-1.0*	5.8	-.8	*-1.4*	4.0	-1.2
Variance Explained (%)									
total	**26**			**23**			**30**		
context (unique)	*			*			**1**		
Model for Minority Slope									
minority gap	**-19.3**	5.1	-19.3	**-23.8**	5.3	-23.8	**-24.3**	4.2	-24.3
Variance Explained (%)	NA			NA			NA		

NA = slope not modeled; * = less than 1 percent.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The percentage of variance explained by each of the models is shown as well. In the case of the model for classroom means two estimates are provided, one showing the total variance explained (43 percent), the other showing the percentage of between-classroom variance explained by the (in this case) community context variables (20 percent). Similarly, we see that urbanicity explains some 2 percent of the between-classroom variation in the minority status-achievement slopes for narrative comprehension. The lack of entries for the expository and document models indicates that we were unable to find predictors of the variation in these slopes.

14.8.3.2. Interpreting Table 14-7

In the first instance, consider the three equations predicting the intercept for narrative, expository, and document comprehension for fourth grade students. The patterns of effects are similar across the three reading comprehension measures. As far as the community context variables are concerned, region, community resources, and parental cooperation all make a difference to the average level of achievement within classrooms. In the case of narrative comprehension the effects of region are negative for the Southeast and Central regions, relative to the West. For expository and document comprehension, only the Southeast shows a significant level of disadvantage. Community resources and parental cooperation both exert positive effects on reading comprehension. With these variables alone we are able to explain between 10 and 20 percent of the variance in the three reading comprehension measures. With the addition of the compositional variable, the between-classroom variance explained increases to around 40 percent.

By contrast, the equations for the ninth grade students show no statistically significant effects at all for the community context variables. We need to keep in mind that the sample size in this group is half that of the fourth grade and, hence, the standard errors are larger. However, the effects overall are small, and the amount of variance explained by the community context variables alone is negligible. In part this is probably due to the fact that high schools are not as closely tied to their communities as elementary schools, and are located in larger communities where the matter of community resources (as measured--libraries, bookstores, other secondary school, museums, etc.) is not an issue. The overall noneffect of region is more difficult to account for.

On the matter of predictors of the minority status-achievement slope, only urbanicity has an effect, and then only in the case of narrative comprehension among fourth graders. In this instance the effect estimates are suggesting that the more urban the community the smaller the minority gap. The base minority gap is 15 points for narrative comprehension in the fourth grade. However, each unit of urbanicity reduces this gap by a little more than 6 points. Seen another way, the minority status-achievement slope becomes flatter as one moves from rural to urban classrooms; the more urban the classroom, the more equitable it is in this respect.

Note, however, that we are explaining only 6 percent of the between-classroom variation in this slope, and that we find this effect in only one of the six equations. Overall, then, the data suggest that one should be suitably modest about these community context models. The study design did not set out explicitly to model community context effects of the kind Coleman and Hoffer (1987) talk about, so perhaps one should not expect too much of these analyses. Nevertheless, we have demonstrated community context effects and we explain a respectable proportion of the variation in the intercept among fourth grade classrooms (but none at all among ninth grade classrooms). And, there is some evidence of effects of urbanicity on the minority status-achievement slope.

14.8.3.3. The Effects of Community Context

The basic notion underlying this model is that characteristics of the community in which the school is set will influence the school itself, in terms of the resources made available to it, and the demands made upon it. Through these, the community influences the achievements of the students (Coleman and Hoffer 1987). Additionally, one might expect that the characteristics of the community would influence student achievement directly in that different communities may offer varying levels of support for education generally, and different sets of opportunities for students to engage in achievement-promoting behaviors outside of school.

These kinds of arguments are examined in a limited way. We have a measure of the urbanicity of the school and hence of the community in which it is located. That community can be set among other communities within the same region to look at influences that arise in this broader context. In addition, two more community-specific variables are included in the model. One is a measure of the kind of educational provision available generally within reach of the school. The second is a measure of the integration of the school within the community, in the sense of the degree of parental cooperation experienced. Thus, the community context model includes the following predictors of between-classroom differences in mean achievement levels of narrative, expository, and document comprehension:

a. The urbanicity of the school;

b. Three dummy variables representing the four regions;

c. Community resources; and

d. Parental cooperation.

These community context variables were obtained from either the School Questionnaire or the sampling frame. Descriptions of these variables and their derivation from the questionnaire data can be found in Chapter 13, and a summary statement is provided as Table 14-6. All except region are measured on ordinal scales, which we choose to treat as though they were interval scales. Region is, of course, an unordered categorical variable and is represented by three dummy variables whose effects are to be interpreted relative to the omitted category, the West. In the first equation estimated, all the variables noted above were treated as predictors of the intercept--the average achievement level of classrooms. No predictors of the slope were specified, though each variable was considered as a candidate to explain variation in the minority status-achievement slope, and the appropriate exploratory statistics were produced to aid selection of slope predictors. This exploratory aspect of the analyses suggested that urbanicity, while not a statistically significant effect on the intercept, may well affect the slope for narrative comprehension, but only among fourth grade students.

The results of reestimating this one equation with the full set of predictors of the intercept together with urbanicity as a predictor of the minority status-achievement slope are shown in the left-most columns of the upper panel in Table 14-7. The coefficients for the other five equations estimated without predictors of the slope are shown in analogous fashion, the two remaining equations for the fourth grade population in the upper panel, and the three for the ninth grade in the lower panel.

14.8.3.4. The Effects of School Attributes

This submodel is concerned with the effects of school attributes on the average level of reading comprehension, and on the effect of minority status on achievement, in the nation's fourth and ninth grade classrooms. There is sizable literature on the way in which schools as organizations affect the learning of students (for example, Bidwell 1965; Lee, Bryk, and Smith 1993), including that generally identified with school effectiveness (Mortimore et al. 1988). The data available to us have aspects of this literature represented, but there is nothing like a sustained attempt to focus on any particular part of it. In fact, the data on the schools themselves consists of fairly conventional administrative record variables. So, at the outset we should be fairly modest about our capability to develop a model of organizational effects on achievement.

Working at a common sense level, it seems that the data available allow one to think about four categories of school-level variables likely to influence the aggregate achievement of students in classrooms. The first refers to rather gross organizational characteristics of schools: whether they are public or private; their size, in terms of student numbers; and the number of hours that they are open for business each week. Second, we think of the human side of this organization and include the racial/ethnic makeup of the teaching staff as an attribute of the organization itself, an indicator of more about a school than simply the literal meaning of the variable. Third, we think of resource differences between schools and include two measures--library resources per student and specialist teacher support for students. (If students needing special attention are segregated by school to some degree, then the latter indicator has a meaning over and above its resource implications.)

Thus, the school attribute predictors in the equations for the intercept and the slope are as follows:

a. Public/private school;

b. School size;

c. Instructional time per week;

d. Racial/ethnic makeup of teaching staff;

e. Library resources; and

f. Specialist teacher support.

In estimating this model, the approach used in the community context submodel was used again. In the first instance the equation estimated contained the six school variables plus the minority composition variable as predictors of the intercept. The minority status-achievement slope was treated as random but was included without predictors. The results of this estimation indicated that, for both grades, there were no variables that could be seen as significant influences on the minority status-achievement slope. As a result the final form of the model is as described immediately above. Details of the parameter estimates obtained are provided in Table 14-8 in the conventional format.

Table 14-8. Classroom-level model: Effects of school attributes, grades 4 and 9

School attributes	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.3**	1.9	56.2	**540.6**	1.7	540.6	**552.0**	1.5	552.0
Fixed Effects									
propn. minor. S's in class	**-35.2**	12.0	-10.5	**-32.6**	10.6	-9.8	**-47.2**	9.8	-14.1
public/private sector	*.1*	6.5		*2.4*	5.7		*6.1*	5.3	
school size	*0*	0	.2	*0*	0	.1	*0*	0	-2.2
hours instruction per week	**2.1**	.6	6.4	**2.0**	.5	6.2	**1.7**	.5	5.3
propn. minor. teachers . .	**-.3**	.2	-6.8	**-.3**	.2	-6.2	*-.1*	.1	-2.1
propn. specialist teachers	*.1*	.2	.9	*.2*	.1	2.3	*.1*	.1	1.4
library books/students . .	**.5**	.2	5.8	**.4**	.2	4.6	*0*	.1	.1
Variance Explained (%)									
total	**35**			**34**			**39**		
school (unique)	**9**			**8**			**5**		
Model for Minority Slope									
minority gap	**-13.7**	3.9	-13.7	**-10.6**	3.0	-10.6	**-16.5**	3.1	-16.5
Variance Explained (%) . . .	NA			NA			NA		
	Grade 9								
Model for Class Means									
Intercept	**538.4**	3.7	538.4	**539.5**	4.0	539.5	**527.9**	2.8	527.9
Fixed Effects									
propn. minor. S's in class	**-114.9**	16.9	-34.4	**-132.9**	18.4	-39.8	**-88.8**	13.2	-26.6
public/private	**24.9**	11.3		**27.3**	12.3		**15.0**	8.7	
school size	*0*	0	5.6	*0*	0	6.7	*0*	0	5.8
hours instruction per week	-.1	.9	-.6	.5	1.0	2.1	.3	.7	1.2
propn. minor. teachers . .	*.2*	.3	3.9	*.4*	.3	8.1	*.1*	.2	1.0
propn. specialist teachers	*-.6*	.3	-8.6	**-.8**	.3	-12.7	*-.4*	.2	-6.1
library books/student . . .	*-.4*	.2	-8.3	**-.5**	.2	-9.6	*-.2*	.2	-5.0
Variance Explained (%)									
total	**33**			**34**			**36**		
school (unique)	**9**			**12**			**9**		
Model for Minority Slope									
minority gap	**-19.7**	5.2	-19.7	**23.9**	5.3	-23.9	**-24.3**	4.2	-24.3
Variance Explained (%) . . .	NA			NA			NA		

NA = slope not modeled.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Overall the school model explains close to one-third of the variance in the intercept in both populations. However, in the fourth grade population most of this is due to the influence of the minority composition variable; the school variables contribute a maximum of only 4 percentage points. Most of this appears to be attributable to differences between schools in hours of instruction and library resources.

In the ninth grade population the school variables add a greater component to the explained variation; however, a large part of this contribution comes from the public/private distinction and favors the students in private schools. In this connection it is worthwhile keeping in mind that we have gone some way toward adjusting for differences in student attributes between the two types of schools in the student-level model, thereby reducing the effects of selective recruiting of students on this outcome. Thus, it may well be that the private schools are offering something deliberately, in the form of program, and/or unwittingly, in the way of compositional effects stemming from selection processes (Murnane 1986), that promotes achievement among ninth graders. The issue is a contentious one, and the arguments are well enough known as to not need revisiting here within the context of a rather limited model.

14.8.3.5. The Effects of the Principal

Most of the variables identified as attributes of the principal were derived from the School Questionnaire and are self-reports by the principals concerned. In addition, we have included two items taken from the Teacher Questionnaire. These concerned aspects of the principal's leadership style and represent teacher perceptions of the extent to which the principal emphasizes instructional leadership or staff development when dealing with staff.

The submodel developed is based on four groups of variables: two principal attributes (gender and minority status), training, experience, and leadership. The arguments embodied in the model suggest that there may be differences according to gender and minority status, but we do not predict the sign of these effects. Better trained and more experienced principals are more likely to promote higher levels of achievement in their schools; and there will be differences across schools according to whether the principal exercises more or less leadership in instructional and/or staff development domains, but the direction of these differences is unclear.

The variables used in the model are as follows:

a. Gender;

b. Minority status;

c. Training in educational administration;

d. Training in reading;

e. Total experience as principal;

f. Experience in this school;

g. Emphasis on instructional leadership; and

h. Emphasis on staff development.

In attempting to estimate the independent effects of these characteristics on classroom achievement, and on the relationship between minority status and achievement within classrooms, the same approach used for the previous two models was used again. Since the exploratory analyses indicated no statistically significant relationships of these variables with the minority status-achievement slopes, the final equations include only predictors of the intercept. The results of estimating the three equations in each of the two grades are shown in Table 14-9.

The statistics presented in these tables indicate that, over and above the class compositional variable, the measures of principal attributes that we have do not explain much of the variance between classrooms in average reading comprehension in either of the two grades -- somewhere between 0 and 4 percent. And perhaps one should not expect them to, as the kind of influence that principals have is applied at some distance from the classroom and probably not directly to the teaching or promotion of reading. However, we would expect whatever principal effects there are to show up most clearly in elementary schools, where the principal is likely to be more directly involved in the teaching work of the school.

Overall, the analyses suggest this is the case. What significant effects exist are in the fourth grade models--principals whose style is to promote staff development have higher levels of reading achievement in their schools, and minority principals are linked to lower average levels of achievement. We do not suppose that the latter is an effect of the minority status of the principal as such, but more likely an artifact of the assignment of minority principals to schools that show lower levels of achievement for reasons other than those included in this model.

14.8.3.6. The Effects of Classroom Attributes

The Reading Literacy Study data appear to tap a set of common sense notions about how classrooms work. The larger the class size, the harder it is for teachers to do what teachers do. The more remedial students there are, the more difficult the teacher's job, and nonremedial students (along with the classroom average) suffer. The more textbooks there are, the better will students learn. The more time teachers spend teaching, the more students learn and the higher the average achievement level of the classroom. This view of classroom influence is captured in the following set of variables, which are used to explain between-classroom variation in the intercept for each of the three forms of reading comprehension. They are used as well in exploratory analyses designed to identify predictors of the minority status-achievement slope. The variables in question are identified below:

a. Class size;

b. Proportion of students needing remedial help;

c. Proportion of remedial students receiving help;

d. Availability of textbooks;

e. Instructional time (all); and

f. Instructional time (reading).

Table 14-9. Classroom-level model: Effects of principal attributes, grades 4 and 9

Principal attribute	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.1**	1.9	556.1	**540.4**	1.7	540.4	**551.8**	1.6	551.8
Fixed Effects									
propn. minor. S's in class	**-40.5**	9.0	-12.1	**-38.6**	8.0	-11.6	**-50.0**	7.4	-15.0
gender	*-2.3*	4.3		*-0*	3.8		*-0*	3.5	
minority status	**-25.1**	8.3	-25.1	**-15.3**	7.4		*-7.3*	6.8	
ed. admin. training	*-8.3*	5.4	*1.1*	1.1	4.8	*-.1*	-.1	4.4	
reading training	**9.1**	4.2		*-1.9*	3.7		*-.7*	3.5	
experience overall	*-.4*	.4	-3.6	*-.4*	.3	3.7	*-.4*	.3	3.6
experience this school	*.4*	.5	2.8	*.8*	.4	4.8	*.5*	.4	2.8
instructional leadership	*-4.0*	2.1	2.1	**-4.3**	1.9	-4.2	*1.9*	1.8	-1.9
staff development orientation	**5.1**	2.0	5.1	**4.0**	1.8	4.1	*2.2*	1.6	2.2
Variance Explained (%)									
total	**32**			**29**			**35**		
principal (unique)	**4**			**2**			*****		
Model for Minority Slope									
minority gap	**-12.9**	3.7	12.9	**-10.6**	3.0	-10.6	**-16.4**	3.1	-16.4
Variance Explained (%)	NA			NA			NA		
	Grade 9								
Model for Class Means									
Intercept	**538.5**	3.8	538.5	**539.7**	4.2	539.7	**528.0**	2.9	528.0
Fixed Effects									
propn. minor. S's in class	**-111.0**	17.3	-33.2	**-123.2**	19.0	-36.9	**-89.0**	13.2	-26.7
gender	*-2.9*	11.4		*-1.5*	12.6		*.3*	8.7	
minority status	*8.9*	15.6		*12.1*	17.2		*4.4*	11.9	
ed. admin. training	*-13.7*	9.6		*-10.3*	10.5		*-8.1*	7.3	
reading training	*-9.7*	8.0	-1.0	*-1.3*	9.4	-1.3	*1.0*	6.5	1.0
experience overall	*-.2*	.8	-1.5	*-.6*	.9	-4.4	*-.4*	.6	-3.0
experience this school	*-1.1*	.9	-7.0	*-1.2*	1.0	-7.5	*-.8*	.7	-5.0
instructional leadership	*-1.5*	3.9	-1.5	*-3.1*	4.3	-3.1	-3.0	3	-3.0
staff development orientation	*-9.7*	8.0		*-13.3*	8.8	-13.3	**-12.5**	6.1	-12.5
Variance Explained (%)									
total	**26**			**25**			**32**		
principal (unique)	*****			**1**			**3**		
Model for Minority Slope									
minority gap	**-19.6**	5.2	-19.6	**-23.8**	5.3	-23.8	**-24.2**	4.2	-24.2
Variance Explained (%)	NA			NA			NA		

NA = slope not modeled; * = less than 1 percent.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The results of estimating equations based on these predictors of classroom achievement levels are shown in Table 14-10 for both the fourth and ninth grade populations, respectively. As before, the first equation estimated in each case specified these variables as predictors of the intercept and provided for an examination of their potential as predictors of the minority status-achievement slope. In all except one of the six equations, there was little to support the notion that classroom variables of this kind affect the size of the minority status-achievement slope in different classrooms. The exception was for expository comprehension in the fourth grade population where the proportion of remedial students helped has a small effect. (Note that the equations estimated differ between the two populations in that the fourth grade equations contain an additional resource variable to do with reading texts per student that was not available in the ninth grade data.)

The results of estimating these equations suggest that as far as classrooms are concerned, the number of students in the class makes a difference in elementary school--other things equal, for each additional student in the class, the class average goes down by a little more than a point. Seen another way, increasing the class size from 20 to 30 students is likely to reduce the class average by 10 to 15 points, some 16 percent of a standard deviation. In high school the effect is not statistically significant, or even particularly meaningful, since these data refer only to one class out of many -- the language-arts class -- and reading as such is not taught in high school. The only other effects of note are due to the proportion of remedial readers in the class, a substantively trivial effect that works in the expected direction: the higher the proportion of remedial readers in the class, the lower the class average; and the greater the proportion of these remedial readers who receive help, the higher the class average.

Oddly, we seem to be explaining a substantial amount of variance in classroom means with this set of variables--some 20 to 30 percent. Although this seems strange given the small size of the effect estimates, the metric of these estimates is cast in small units: per student, for class size; single percentage points for the remedial variables; single books for the library resources variable; and hours for the two instructional time variables.

14.8.3.7. The Effects of Teacher Attributes

The model for the effects of teacher attributes on average reading achievement levels takes something like a human capital approach. Its basic proposition is that the more we invest in the education, training, and experience of teachers, the better teachers they will be, and this will be reflected in the achievement levels of their classrooms. To such human capital measures--formal education, teacher training, and years of experience--we add gender and minority status as ascribed attributes of teachers with some potential to affect the achievements of students.

The variables in question were specified as follows:

a. Gender;

b. Minority status;

c. Formal education;

d. Teacher training; and

e. Teaching experience.

Table 14-10. Classroom-level model: Effects of classroom attributes, grades 4 and 9

Classroom attributes	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.4**	1.8	556.4	**540.8**	1.5	540.8	**552.2**	1.5	552.2
Fixed Effects									
propn. minor. S's in class	**-38.3**	6.5	-11.5	**-35.2**	5.7	-10.5	**-41.9**	5.4	-12.6
class size	**-1.1**	.3	-5.7	**-1.4**	.3	-7.4	**-1.0**	.3	-5.3
% remedial students	**-.8**	.1	-14.7	**-.8**	.1	-13.0	**-.6**	.1	-10.2
% remedial students helped	*-0*	.1	-.1	*-0*	0	-0	*-0*	0	-.6
library resources	*-0*	0	2.2	*0*	0	1.7	*0*	0	-.12
instructional time (all)	*.1*	.4	.6	*-.3*	.3	-1.4	*-.4*	.3	-1.7
instructional time (reading)	*-.3*	.6	-.9	*0*	.5	.1	*-.3*	.5	-1.0
Variance Explained (%)									
total	**49**			**47**			**54**		
class (unique)	**28**			**26**			**29**		
Model for Minority Slope									
minority gap	**-14.0**	3.9	-14.0	**-10.0**	3.0	-10.0	**-16.7**	3.1	-16.7
% remedial students helped				**.2**	.1	5.9			
Variance Explained (%)	NA			**12**			NA		
	Grade 9								
Model for Class Means									
Intercept	**538.3**	3.3	538.3	**539.5**	3.7	539.5	**527.8**	2.5	527.8
Fixed Effects									
propn. minor. S's in class	**-72.0**	11.8	-21.5	**-81.6**	13.5	-24.4	**-59.5**	9.3	-17.8
class size	*.8*	.6	4.8	*1.0*	.6	6.0	*.4*	.4	2.5
% remedial students	**-.8**	.2	-17.1	**-.8**	.2	-16.8	**-.7**	.2	-14.6
% remedial students helped	**.2**	.1	11.0	**.2**	.1	12.1	**.1**	.1	6.6
instructional time (all)	*-.3*	.3	-2.8	*-.3*	.4	-2.9	*-.3*	.3	-3.3
instructional time (reading)	*-3.1*	3	-3.6	*.3*	3.4	.3	*-1.6*	2.3	-1.9
Variance Explained (%)									
total	**50**			**44**			**51**		
class (unique)	**31**			**26**			**31**		
Model for Minority Slope									
minority gap	**-19.6**	5.2	-19.6	**-23.4**	5.4	-23.4	**-24.5**	4.2	-24.5
Variance Explained (%)	NA			NA			NA		

NA = slope not modeled.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Only two of the teacher attributes considered seem to exert an effect on the achievement levels of classrooms, and then only for fourth grade classrooms. If one takes a fairly relaxed stance with regard to statistical significance, then teacher gender and perhaps teacher education could be considered as influences on classroom achievement. Female teachers and teachers with higher levels of formal education seem to promote higher levels of achievement in reading comprehension in elementary schools. Other attributes of teachers appear to make little difference to the achievement levels of classrooms. Further, equations estimated in the usual way failed to reveal any attributes with consistent statistically significant effects on the average level of reading comprehension or on the minority status-achievement slope. In total, this set of predictors explains, with the exception of fourth grade expository comprehension, less than 5 percent of the variation in classroom means. Table 14-11 displays the results of these analyses.

The apparent conclusion is that teachers, characterized in this way, do not make much of a difference. However, it is difficult to accept the notion that better educated, better trained, and more experienced teachers are not better teachers whose work results in higher levels of achievement in their classrooms. Most likely we are seeing at least four sets of influences at work in this respect. First, the measures of reading comprehension used evaluate basic reading skills, the teaching of which, conceivably, does not require especially high levels of education, training, or experience. Second, the range of the education and training of U.S. teachers is truncated at the lower end. That is, most teachers in the U.S. are educated and trained beyond the point at which education and training make a difference in the teaching of reading. Third, quality of training is more important in this respect than quantity of training, the measure we have. Fourth, and probably most important, it is not entirely reasonable to think that the effects of better qualified and more experienced teachers would radically alter the achievement level of a classroom during the course of a single year when the achievement of the students concerned is the cumulation of either 4 or 8 years of schooling with other teachers. These attributes of teachers might be expected to effect growth in achievement during a school year, but we are unable to examine this proposition with the cross-sectional data at hand.

14.8.3.8. The Effects of Instructional Practices

The data on instructional practice are fairly extensive, but problematic as well in two main ways. First, much of the data is organized in multi-item blocks that take the form of single questions in the Teacher Questionnaire, each with the same stem and the same response scale but with up to 30 subitems requiring a response. Presumably these are meant to measure one or more--most likely more, given the number of items--latent instructional variables from a domain thought to represent important influences on the development of reading comprehension. While we do not quite know what these items were designed to measure, exploratory analyses have provided interpretable latent variables.

The second problem stems from the fact that reading as a formal subject is not taught in the ninth grade, though it may well be "taught" as part of other subjects. So, it is not entirely clear what an instructional model would look like in this instance. It is difficult, therefore, to know what to make of the ninth grade instruction data; consequently, while we have estimated equations with these data, not much can be said about the results.

As noted in Chapter 13, exploratory factor analyses of the multi-item block questions in the fourth grade Teacher Questionnaire revealed a latent structure that could be anchored in the literature on reading. Many of the factor scores produced to represent these latent variables were highly correlated and subject to similar interpretations even though they were derived from different sets of items in the questionnaire. Under these circumstances, it was necessary to make some judgments about variables to

Table 14-11. Classroom-level model: Effects of teacher attributes, grades 4 and 9

Teacher attributes	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.0**	1.9	556.0	**540.4**	1.6	540.4	**551.7**	1.5	551.7
Fixed Effects									
propn. minor. S's in class	**-51.2**	7.4	-15.3	**-46.3**	6.4	-13.9	**-50.8**	6.0	-15.2
gender	*10.4*	5.6		**14.8**	4.8		**11.1**	4.5	
minority status	*-11.7*	7.6		*-11.0*	6.6		*-7.3*	6.1	
formal education	*10.0*	5.4		**13.3**	4.7		*4.3*	4.3	
training	*-2.7*	2.6	-2.3	*-3.9*	2.2	-3.4	*-.5*	2.1	-.4
experience	*-.4*	.2	3.1	*.4*	.2	3.3	*.4*	.2	3.7
Variance Explained (%)									
total	**32**			**35**			**39**		
teacher (unique)	**4**			**9**			**5**		
Model for Minority Slope									
minority gap	**-13.7**	3.9	-13.7	**-10.5**	3.0	-10.5	**-16.4**	3.1	-16.4
Variance Explained (%)	NA			NA			NA		
	Grade 9								
Model for Class Means									
Intercept	**538.6**	3.7	538.6	**539.8**	4.2	539.8	**528.1**	2.9	528.1
Fixed Effects									
propn. minor. S's in class	**-91.7**	14.5	-27.4	**-103.4**	16.2	-30.9	**-76.7**	11.3	-22.9
gender	*6.1*	8.8		*4.6*	9.8		*6.9*	6.8	
minority status	*-17.7*	17.9		*-9.7*	19.9		*-8.3*	13.9	
formal education	*6.2*	8.5		*4.4*	9.4		*6.3*	6.6	
training	*-.8*	4.2	-.8	*0.8*	4.7	-.8	*-.1*	3.3	-.1
experience	*.8*	.5	6.8	*1.0*	.6	8.9	*.5*	.4	3.9
Variance Explained (%)									
total	**29**			**26**			**32**		
teacher (unique)	**3**			**3**			**3**		
Model for Minority Slope									
minority gap	**-20**	5.2	-24.2	**-23.9**	5.3	-23.9	**-24.2**	4.2	-19.6
Variance Explained (%)	NA			NA			NA		

NA = slope not modeled.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

include on substantive grounds and with a view to avoiding redundant measures and excessive collinearity among measures.

The variables included in the fourth grade instructional model are listed below.

a. Emphasis on student interaction with text;

b. Teacher's own reading preferences;

c. Emphasis on assessment;

d. Emphasis on transmission/translation approach;

e. Use of grouping;

f. Instructional hours on reading; and

g. Emphasis on homework.

Bearing in mind the issues raised in connection with the ninth grade instructional model, we identified the following variables as the most likely and justifiable candidates for inclusion in that model. Note that, as a result, the fourth and ninth grade models are not the same.

a. Activities to construct meaning;

b. Activities to extend meaning;

c. Student-directed instruction;

d. Teacher-directed instruction;

e. Teacher reading, professional; and

f. Teacher reading, general.

Equations predicting class means for narrative, expository, and document comprehension were estimated in the same fashion as previously. In one case the exploratory analyses designed to identify predictors of the minority status-achievement slope did so. Emphasis on homework appeared as a predictor for document comprehension in the fourth grade data. In this case the equation was reestimated. The results of estimating the three equations for each population of students are shown in Table 14-12.

The prediction of between-class differences in achievement with the teacher instructional variables at hand shows little of interest and explains next to nothing of the between-classroom variance. The one exception occurs in connection with document comprehension in the fourth grade. The more homework a teacher gives, the lower the minority achievement gap.

Table 14-12. Classroom-level model: Effects of instructional practices, grades 4 and 9

Instructional practice	Grade 4								
	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Model for Class Means									
Intercept	**556.0**	2.0	556.0	**540.4**	1.7	540.4	**551.8**	1.6	551.8
Fixed Effects									
propn. minor. S's in class	**-54.7**	6.8	-16.6	**-52.2**	6.0	-15.6	**-54.7**	5.5	-16.4
instructional orientation	*3.0*	2.6	2.9	*.9*	2.3	.9	*-.3*	2.1	-.3
T's reading preferences	*-4.2*	2.6	-3.8	*-1.2*	2.3	-1.1	*.4*	2.1	.4
assessment practices	*-4.1*	2.6	-3.7	*-2.0*	2.3	-1.8	*-1.7*	2.1	-1.5
hierarchical instruction	*-.1*	2.4		*-.3*	2.1		*.4*	2.0	
use of grouping	*1.0*	4.3		*-.1*	3.8		*3.0*	3.5	
instructional time (reading)	*-.2*	.2	-3.1	*-.1*	.1	-1.1	*-.1*	.1	-1.4
homework emphasis	*-1.1*	1.7	-1.3	*-2.2*	1.5	-2.6	*-.6*	1.4	-.7
Variance Explained (%)									
total	**29**			**27**			**35**		
instruct (unique)	*			*			*		
Model for Minority Slope									
minority gap	**-13.7**	3.9	-13.7	**-10.6**	3.0	-10.6	**-15.8**	3.1	-15.8
homework							**5.8**	2.6	6.8
Variance Explained (%)	NA			NA			17		
	Grade 9								
Model for Class Means									
Intercept	**538.6**	3.8	538.6	**539.8**	4.2	539.8	**528.0**	3.0	528.0
Fixed Effects									
propn. minor. S's in class	**-95.5**	13.4	-28.6	**-101.9**	14.7	-30.5	**-78.1**	10.4	-23.4
construct meaning	*8.6*	6.3	7.7	**15.3**	7.0	13.6	**11.0**	4.9	9.8
extend meaning	*-8.5*	6.1	-7.0	*-11.0*	6.7	-9.0	*-7.9*	4.7	-6.5
student directed	*.6*	4.8	.6	*1.1*	5.3	1.0	*-1.5*	3.7	-1.3
teacher directed	*-6.7*	4.6	-5.9	**-11.7**	5.1	-10.5	*-4.0*	3.6	-3.6
T. professional reading	*-1.6*	5.1	-1.5	*-.8*	5.6	-.7	*-4.2*	3.9	-3.9
T. general reading	*-.5*	4.6	-.4	*-8.3*	5.1	-7.6	*-3.4*	3.6	-3.1
Variance Explained (%)									
total	**25**			**24**			**28**		
instruct (unique)	*			*			*		
Model for Minority Slope									
minority gap	**-19.4**	5.2	-19.4	**-23.8**	5.3	-23.8	**-24.3**	4.2	-24.3
Variance Explained (%)	NA			NA			NA		

NA = slope not modeled; * = less than 1 percent.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Our interpretations of these findings are not meant to imply that teachers' instructional practices and orientations do not make a difference in elementary school. The arguments raised in connection with the effects of teacher attributes apply here as well. Conceivably, important instructional variables were not included in the questionnaires and, hence, are not part of our model; quite possibly the important variables are included but not well measured by these questionnaires; or, it is essentially impossible to demonstrate the effects of instructional practice on achievement during the course of a single school year.

This is not meant to imply that the teaching practices item group was without meaning. The analyses undertaken identified some latent variables consistent with the theory and substance of reading research. However, as a whole, the constructs identified seemed to capture only fragments of prevailing theories about the effects of reading instruction. This point is made with some force in Chapter 13, where the constructs identified are mapped onto a typology of reading theories.

But, even if given the best variables well measured, there is a generic problem with research designs of this kind. Without a pretest measure of reading comprehension taken at the beginning of the year, the chances are slim that one would find effects that can be attributed to what teachers do. The reading capabilities of students in fourth and ninth grades are well established at the beginning of the year when they enter these grades, and they reflect the cumulation of past teaching within school and without. It would be unrealistic to think that, over the period of a school year, between-classroom differences in teachers' methods of instruction and personal orientations would alter substantially the between-classroom distribution of achievement. To get at instructional effects one needs a pretest/posttest design. In this instance it would be possible to look at growth across the year in the average achievement of students within classrooms, a measure more likely to be influenced by differences in instructional practice over the course of a school year. Unfortunately, the cost and practical difficulties of such research designs usually preclude them coming into being.

Further, if this explanation is close to the truth, then the effects we should find in the present circumstances are those cumulative effects from students' experiences in the longer term, those which have been shaping their lives since birth, and/or for the total of their school years. As it turns out, the bulk of the effects are of this kind, especially for the younger group of students--effects due to the attributes of the community context in which students live, the number of hours of instruction offered by the school, the library resources of the school, and the size of the class the student is in (assuming these school and class attributes are matters of school policy and, hence, of an enduring nature).

14.9. A General Model of Reading Comprehension

The intent of the model development process described above was to formulate a single meaningful and reasonably defensible model to explain between-student and between-classroom variation in reading comprehension. This development had many of the characteristics of a secondary analysis of data from a large omnibus survey. However, the general approach differs in this case in that the convention in IEA studies is to treat the data as though they were generated by a single comprehensive model and should be exploited to their fullest extent in estimating such a model. This expectation shapes an analysis plan under which attempts are made to use the full scope of the data available in a single analysis. In contrast, in the secondary analysis of omnibus surveys one tends to focus on selected aspects of the data with the view to producing a number of smaller, more contained analyses based on limited models that address specific issues.

The development of the general model relied in good part on conceptual inference and statistical exploration. However, these inferences and explorations were undertaken within theoretical/substantive structures seen to be consistent with what is known about the processes by which students come to be differentiated in terms of school achievement and, more specifically, with what the literature on reading has to say in this respect. While no great claims will be made for the working models developed in this way, they are necessary to lend a degree of theoretical coherence to the analyses.

14.9.1. Defining a Student-Level Model

In developing the structure and content of the general model, we chose to stay with the form of the student-level model as defined earlier. Although the model did not explain a great deal of the between-student variance in reading comprehension, it did encompass the important student-level variables available, and most of these had statistically significant effects on the three reading comprehension outcome measures. Further, the structure and content of the model itself, though somewhat limited, was consistent with the substantial status attainment literature on models of school achievement.

14.9.2. Defining a Classroom-Level Model

The development of the classroom-level model from the analyses reported immediately above is less straightforward. The classification of these classroom-level variables into six broad categories provided an initial handle on the data. The development and estimation of models within each of these categories, models designed to explain between-classroom variation in classroom means and slopes, was the next step toward clarification. In this overtly exploratory way, an attempt was made to refine the variables to a manageable set to be included in a composite model of the sources of variation in reading comprehension.

The model development procedure adopted was designed to build on the findings of the six submodels developed above. Variables shown to have statistically significant effects on the intercept and/or slopes within any of the six submodels were considered candidates for inclusion in a single larger model. Table 14-13 summarizes the HLM coefficients presented in Tables 14-7 through 14-12 in terms of their statistical significance. The entries in the tables can be interpreted as follows: a "+" indicates that the coefficient in question is equal to or greater than 1.96 times its standard error (5 percent level of confidence); a "?" points to a coefficient equal to or greater than 1.9 times its standard error; a "-" indicates an effect coefficient less than 1.9 times its standard error (6 percent level of confidence); and an "NA" identifies variables "not applicable" in the model, as a function of not being available for the grade in question.

Our model development strategy used this kind of information, together with a degree of judgment about what is meaningful in the context of the model as a whole, to select a subset of variables for a single model whose structure parallels that of the overall model shown in Figure 14-3. In the interest of simplicity, the model adopted has the same structure for both grade 4 and 9 samples, even though variables with significant effects on reading comprehension at one grade level often did not demonstrate these effects at the other grade. The basic rule of thumb was that a variable would be considered a candidate for inclusion in the final model if at least three of the six coefficients were statistically significant at the chosen 5 percent level, or close to it. As always, judgments based on substantive grounds would allow for the bending of this rule. The results of this selection process are shown in Table 14-14.

Table 14-13. Summary of effects in classroom-level models: Grades 4 and 9

Model variable	Grade 4			Grade 9		
	Narrative comprehension	Expository comprehension	Document comprehension	Narrative comprehension	Expository comprehension	Document comprehension
Community Context						
urbanicity of the school	-	-	-	-	-	+
region; Northeast				-	-	-
region; Southeast	+	+	+	-	-	-
region; Central	+	-	-	-	-	-
community resources	+	+	-	-	-	-
parental cooperation	+	+	+	-	-	-
School						
public/private sector	-	-	-	+	+	+
school size	-	-	-			
instructional time per week	+	+	+	-	-	-
proportion minority teachers	+	+	-	-	-	-
library books per student	+	+	-	?	+	-
proportion specialist teachers	-	-	-	?	+	?
Principal						
gender	-	-	-	-	-	-
minority status	+	+	-			
training in educational administration	-	-	-	-	-	-
training in reading	+	-	-	-	-	-
total experience as principal	-	-	-	-	-	-
experience this school	-	-	-	-	-	-
emphasis on instructional leadership	-	+	-	-	-	-
emphasis on staff development	+	+	-	-	-	+
Classroom						
class size	+	+	+	-	-	-
proportion students needing remedial help	+	+	+	+	+	+
proportion remedial students receiving help	-	-	-	+	+	+
textbooks per student	-	-	-	NA	NA	NA
hours per week instructional time	-	-	-	-	-	-
hours per week reading instruction	-	-	-	-	-	-
Teacher						
gender	?	+	+	-	-	-
minority status	-	-	-	-	-	-
formal education	-	+	-	-	-	-
teacher training	-	-	-	-	-	-
years of teaching experience	-	-	-	-	-	-
Instructional Practice (Grade 4)						
emphasis on student's interaction with text	-	-		NA	NA	NA
extent of teacher's own reading	-	-	-			
emphasis on assessment	-	-	-			
emphasis on hierarchical skills		-	-			
use of grouping	-	-	-			
instructional hours on reading	-	-	-			
amount of reading homework	-	-	-			
Instructional Practice (Grade 9)						
activities to construct meaning	NA	NA	NA	-	+	+
activities to extend meaning				-	-	-
student-directed instruction				-	-	-
teacher-directed instruction				-	+	
teacher reading; professional				-	-	-
teacher reading; general				-	-	-

KEY:
+ = statistically significant (α = .05)
- = not statistically significant
? = statistically significant (α = .06)
NA = not applicable

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 14-14. Makeup of composite model: Grades 4 and 9

Model variable	Grade 4			Grade 9		
	Narrative comprehension	Expository comprehension	Document comprehension	Narrative comprehension	Expository comprehension	Document comprehension
1. Exogenous Variables						
Community Context						
region; Northeast	-	-	-	-	-	-
region; Southeast	+	+	+	-	-	-
region; Central	+	-	-	-	-	-
community resources	+	+	-	-	-	-
parental cooperation	+	+	+	-	-	-
School						
public/private sector	-	-	-	+	+	+
instructional time per week	+	+	+	-	-	-
library books per student	+	+	-	?	+	-
proportion specialist teachers	-	-	-	?	+	?
Principal						
emphasis on staff development	+	+	-	-	-	+
2. Endogenous Variables						
Classroom						
class size	+	+	+	-	-	-
proportion students needing remedial help	+	+	+	+	+	+
Teacher						
gender	?	+	+	-	-	-

KEY: + = statistically significant (α = .05)
- = not statistically significant
? = statistically significant ((α = .06)

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

It is clear that we are violating convention in using tests of significance as a filtering device. We use the standard errors of coefficients as a guide to the identification of variables expected to have effects on reading comprehension. These standard errors are used to indicate the likelihood that a variable in a submodel will account for between-student or between-classroom variation in reading comprehension in a composite model if such a model should be estimated with data from another sample from the same grade. While clearly a somewhat atheoretical approach to model building, it does takes place within a predefined structural model.

The variables identified in this way cover most of the categories noted in the general model shown in Figure 14-3, though the coverage of categories is uneven. The general model contains three attributes of the context within which the school is set (region, community resources, and parental cooperation); four characteristics of the school itself (school sector, instructional time, library resources, specialist teachers); a single measure to do with the principal's leadership style; three classroom attributes, basically characteristics of the student body (class size, proportion of minority students, proportion of remedial students); and one characteristic of teachers (gender). As noted earlier, we were unable to establish a case for the inclusion of any of the instructional practice variables.

We are treating the compositional variable "proportion of minority students in class" a little differently in this composite model. During the model development phase, we included this variable in the model essentially as a statistical control used to obtain better estimates of other parameters. As such it was in all equations and was treated, essentially, as a classroom-level exogenous variable. In the composite model described below, the compositional variable takes its logical place within the model as an attribute of classrooms, which makes it an endogenous variable in this model.

14.9.3. Estimating the Reading Literacy Model

The results of estimating the model so defined are presented in Tables 14-15 and 14-16. The structure of the models for both grades 4 and 9 is identical at both the student level and classroom level. And, since we were unable to identify any predictors of the minority status-achievement slope, the analyses do not report a modeling of the variation in classroom slopes. However, consistent with the structure of the model proposed originally, the specification of an exogenous and an all-variable equation at each level was retained. This provided for a minor degree of complexity in the estimation. The complexity is based on the notion that the equation of interest at any level ought to be estimated in a model that includes statistical controls on all the variables at the other level. In practice this meant that when estimating both exogenous and all-variable equations for the student-level model, the all-variable equation was used at the classroom level; when estimating the exogenous and all-variable equations for the classroom-level model, the all-variable equation was used at the student level.

Panel 1 in each table displays the coefficients for the exogenous equation in the between-student level of the model. Panel 2 provides information for the all-variable equation at the same between-student level. Panels 3 and 4 display coefficients estimated for the two analogous between-classroom equations.

14.9.4. Between-Student Differences in Reading Comprehension

The development of the classroom-level model proved to have little effect on the expression of effects within the student-level model. The pattern of effects in panels 1 and 2 of Tables 14-15 and 14-16 are almost identical to those discussed earlier in connection with Tables 14-7 and 14-8 where the classroom-level model was limited to a single variable -- proportion of minority students in class. Here, as there, the exogenous variables appear to be doing all the work in the sense of explaining between-student variance in reading comprehension. In total they explain about 8 percent of the variability between fourth grade students in these respects, and rather less among ninth grade students. The addition of the endogenous family environment variables to the equation adds only 1 or 2 percentage points to these values.

Further, the specification of exogenous and endogenous variables and the estimation of two equations with the view to examining total, direct, and indirect effects turned out to be a needless complication, for the most part. The family environment variables specified did not transmit the influence of the family status variables and, in fact, appeared to be somewhat unrelated to both the family status variables and the measures of reading comprehension. With the exception of family wealth, the effects of the exogenous variables remained largely unchanged between the exogenous and all-variable equations.

The meaning we can assign to this fact is that the intervening family process variables specified in the research design were largely irrelevant to the explanation of reading comprehension per se. They were also largely irrelevant to the explanation of why it is that the ascribed and achieved statuses, and the structural attributes, of students and/or their families make a difference in this respect. The exception was family wealth, whose effects were transmitted substantially by literacy possessions as an intervening variable. However, in this instance we would want to be a little careful in assigning much

Table 14-15. Composite model: Grade 4

	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Panel 1: Exogenous equation, between student									
Random Effects									
minority status	-16.2	4		-12.8	3.1		-18.2	3.2	
Fixed Effects									
age	-1.9	0.2	-13	-1.5	0.1	-9.9	-1.1	0.1	-7
gender	14.9	2.3		6.6	1.9		-2.7	1.9	
father's education	7.8	2.8		9.8	2.3		5.6	2.3	
mother's education	3.4	2.8		2.6	2.3		7	2.3	
family wealth	6.1	1.2	5.7	6.1	1	5.7	6.6	1	6.2
family composition	14.3	2.4		10.3	2		17.9	2	
augmented family	19.5	2.4		-15.9	2		-15	2.1	
language	10.4	2.8		-7.5	2.3		-12.9	2.3	
Variance Explained (%)	8			7			8		
Panel 2: All-variable equation, between student									
Random Effects									
minority status	-14.1	3.9		-10.9	3		-16.7	3.1	
Fixed Effects									
age	-1.8	0.2	-12	-1.4	0.1	-9.4	-1	0.1	-6.6
gender	15.8	2.3		7.3	1.9		-1.6	1.9	
father's education	6.9	2.8		8.9	2.3		4.8	2.3	
mother's education	2.6	2.7		2	2.3		6.1	2.3	
family wealth	-0.5	1.6	-1	0.6	1.4	0.6	-0.3	1.4	-0.3
family composition	13.7	2.4		9.6	2		17.4	2	
augmented family	-18.3	2.4		-14.8	2		-4	2	
language	-9.9	2.8		-7	2.3		-12.8	2.3	
regular meals	9.9	2.8		11.7	1.9		5.6	2	
literacy possessions	11.7	1.8	10	9.8	1.5	8.5	11.9	1.5	10.4
study place	-9.1	2.7		-12.2	2.2		-5.5	2.2	
TV watching	-1.7	0.6	-11	-1.1	0.5	-7.2	0.6	0.5	-4.1
homework help	-3.6	1.4	-3	-2.2	1.2	-1.8	-4.1	1.2	-3.4
parental interaction	-5	1.5	-4	-2.9	1.3	-2.2	-4.8	1.3	-3.6
Variance Explained (%)	9			8			9		

Table 14-15. Composite model: Grade 4 (continued)

	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Panel 3: Exogenous equation, between classroom									
Model for Class Mean									
Intercept	**-556.4**	1.9		**540.7**	1.7		**552.1**	1.6	
Fixed Effects									
region; Northeast.	9.4	6.0		8.4	5.4		9.5	5.2	
region; Southeast.	**-24.7**	5.4		**-16.8**	4.9		**-18.5**	4.7	
region; Central	-7.1	5.2		3.3	4.7		2.3	4.5	
community resources . . .	0.2	4.1		-3.0	3.7		-2.4	3.6	
parental cooperation	**13.7**	2.3	12.7	**11.1**	2.1	10.3	**9.0**	2.0	8.3
school sector	-7.1	6.7		-3.7	6.1		2.4	5.8	
hours instruction/week . .	**3.0**	0.7	9.2	**2.8**	0.6	8.1	**2.5**	0.6	7.5
library book/S	**0.5**	0.2	6.6	**0.4**	0.2	5.2	0.1	0.2	1.9
% specialist teachers . . .	0.1	0.2	1.4	0.2	0.2	2.4	0.1	0.1	1.6
P. staff development . . .	3.3	1.9	3.4	2.7	1.8	2.7	2.3	1.7	2.3
Variance Explained (%) . . .	**35**			**26**			**2.5**		
Model for Minority Slope									
minority gap	**-12.8**	3.9		**-10.0**	3.0		**-15.6**	3.1	
Variance Explained (%) . . .	NA			NA			NA		
Panel 4: All-variable equation, between classroom									
Model for Class Mean									
Intercept	**556.6**	1.6		**541**	1.4		**552.3**	1.4	
Fixed Effects									
region; Northeast.	**11.3**	5		**9.4**	4.5		**10.2**	4.4	
region; Southeast.	**-13.8**	4.6		-6.9	4.1		**-9.2**	4	
region; Central	**-9.7**	4.3		0.7	3.9		-1.4	3.8	
community resources . . .	**6.1**	5.3		4.9	4.8		3.8	4.7	
parental cooperation	**8.5**	1.9	7.9	**6.2**	1.7	5.8	**3.9**	1.7	3.6
school sector	-10.6	5.5		-8.5	5		-1.4	4.9	
hours instruction/week . .	**2.5**	0.6	7.5	**2**	0.5	6.1	**1.8**	0.5	5.6
library book/S	**0.3**	0.2	3.8	0.2	0.1	2.1	-0.1	0.1	-0.9
% specialist teachers . . .	**0.3**	0.1	3.7	**0.4**	0.1	4.8	**0.3**	0.1	3.3
P. staff development . . .	1.6	1.6	1.6	0.9	1.5	0.9	0.9	1.4	
class size	-0.6	0.3	-3.2	**-1**	0.5	-5.4	**-0.7**	0.3	-3.9
% remedial S's	**-0.8**	0.1	-13.3	**-0.7**	0.1	-12.7	**-0.6**	0.1	-10
propn. minority S's	-35.2	6.1	-10.5	**-22.3**	5.5	-9.7	**-40.9**	5.4	-12.3
T gender	**9.5**	4.6		**13.3**	4.1		**9.3**	4	
Variance Explained (%) . . .	**67**			**61**			**59**		
Model for Minority Slope									
minority gap	**-14.1**	3.9		**-10.9**	3		**-16.7**	3.1	
Variance Explained (%) . . .	NA			NA			NA		

NA = slope not modeled.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

Table 14-16. Composite model: Grade 9

	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Panel 1: Exogenous equation, between student									
Random Effects									
minority status	**-22.3**	5.3		**-27**	5.4		**-27.1**	4.3	
Fixed Effects									
age	**-1.3**	0.2	-10	**-1.4**	0.2	-10.7	**-1.2**	0.2	-9.5
gender	**17.1**	2.9		-2.4	3.2		**-5.8**	2.5	
father's education	**8.1**	3.3		**7.6**	3.6		**8.1**	2.9	
mother's education	5.2	3.3		**11.8**	3.6		2.3	2.9	
family wealth	3.3	1.5	3.2	-0.8	1.7	-0.8	-0.6	1.3	-0.6
family composition	1.3	3.1		-0.5	3.4		-1.5	2.7	
augmented family	-2.8	4		-6.6	4.4		-5.5	3.5	
language	**-10**	3.7		-4.8	4.1		**-8**	3.2	
Variance Explained (%) ...	5			3			4		
Panel 2: All-variable equation, between student									
Random Effects									
minority status	**19.6**	5.2		**-24**	5.3		**-24.1**	4.3	
Fixed Effects									
age	**-1.3**	0.2	-10	**-1.3**	0.2	-10.3	**-1.2**	0.2	-9.3
gender	**16.5**	3		-1	3.3		**-5.5**	2.6	
father's education	**8**	3.3		6.8	3.6		**7.5**	2.9	
mother's education	4.8	3.3		**10.3**	3.6		1.3	2.9	
family wealth	**-5.1**	2.4	-4.9	**-13.5**	2.6	-13	**-10.9**	2.1	-10.4
family composition	**3.4**	3.1		0.8	3.4		0.1	2.7	
augmented family	-3.1	4		-5.9	4.4		-5.3	3.5	
language	**-11.1**	3.7		-5.6	4.1		**-8.8**	3.2	
regular meals	-2.8	3.1		6	3.4		1.3	2.7	
literacy possessions	**11.8**	2.5	10.8	**18**	2.8	16.5	**14.4**	2.2	13.3
study place	-3.2	2.9		**-8.9**	3.2		**-7.3**	2.6	
TV watching	-2.6	0.8	-4.7	**-2.3**	0.9	-4	**-2.6**	0.7	-4.7
homework help	**-4.9**	1.9	-4	-2.6	2.1	-2.1	-2.8	1.6	-2.2
parental interaction	**-4.5**	1.6	-4	**-4.8**	1.8	-4.3	**-5.2**	1.4	-4.7
Variance Explained (%) ...	6			5			7		

Table 14-16. Composite model: Grade 9 (continued)

	Narrative			Expository			Document		
	Metric	Standard error	Standard	Metric	Standard error	Standard	Metric	Standard error	Standard
Panel 3: Exogenous equation, between classroom									
Model for Class Mean									
Intercept	**538.6**	4.2		**539.8**	4.6		**528.2**	3.3	
Fixed Effects									
region; Northeast.	17.9	12.5		14.0	13.7		10.7	9.7	
region; Southeast.	-7.1	11.5		-10.1	12.6		-11.9	8.9	
region; Central	17.8	11.8		22.0	12.8		14.1	9.1	
community resources . . .	-4.5	9.3		-13.0	10.1		-5.9	7.2	
parental cooperation	**14.9**	5.2	12.6	**14.3**	5.6	12.1	**10.1**	4.0	8.5
school sector	3.3	13.2		6.7	14.4		-2.0	10.3	
hours instruction/week . .	-0.0	1.1	-0.0	0.2	1.2	0.9	0.3	0.4	1.3
library book/S	-0.4	0.2	-7.4	-0.5	0.2	-9.5	-0.2	0.2	-4.7
% specialist teachers . . .	-0.5	0.3	-8.2	**-0.8**	0.3	-13.0	-0.4	0.2	-6.6
P. staff development . . .	-1.2	8.8	-1.2	-4.0	9.6	-4.8	-5.3	6.8	-5.3
Variance Explained (%) . . .	**8**			**9**			**10**		
Model for Minority Slope									
minority gap	**-19.5**	5.2		**-23.6**	5.4		**-23.4**	4.2	
Variance Explained %	NA			NA			NA		
Panel 4: All-variable equation, between classroom									
Model for Class Mean									
Intercept	**538.2**	3.2		**539.3**	3.6		**527.8**	2.4	
Fixed Effects									
region; Northeast.	1.8	10		-4.4	11.1		-3.2	7.5	
region; Southeast.	-4.5	8.9		-8.1	10		-11.7	6.7	
region; Central	5.5	9.4		7.2	10.6		1.7	7.1	
community resources . . .	-4	7.2		-12.1	8		-4.4	5.4	
parental cooperation	-1.4	4.4	-1	-4	4.9	-3.4	-3.2	3.3	-2.7
school sector	20.6	10.9		**27.2**	12.1		14.8	8.1	
hours instruction/week . .	0.2	0.8	0.7	0.4	0.9	1.5	0.5	0.6	1.9
library book/S	**0.5**	0.2	-10	**-0.6**	0.2	-12.9	**-0.4**	0.1	-8
% specialist teachers . . .	**-0.5**	0.2	-7.9	**-0.8**	0.2	-12.6	-0.4	0.2	-5.8
P. staff development . . .	-2.3	7		-6	7.8		-8.8	5.2	
class size	0.1	0.6	0.8	0.1	0.7	0.4	-0.1	0.5	-0.4
% remedial S's	**-1**	0.2	-23	**-1.1**	0.2	-23.8	**-0.8**	0.1	-17.6
propn. minority S's	**-83.6**	13.8	-25	**-95.9**	15.4	-28.7	**-72.8**	10.4	-21.8
T gender	14.4	7.8		16.5	8.7		**14.7**	5.8	
Variance Explained (%) . . .	**51**			**48**			**57**		
(Compositional effect) . .	**33**			**32**			**39**		
Model for Minority Slope									
minority gap	-19.6	5.2		**-24**	5.3		**-24.1**	4.3	
Variance Explained (%) . . .	NA			NA			NA		

NA = slope not modeled.

NOTE: Coefficients equal to, or greater than, 1.96 times their standard error are shown in boldface; other coefficients are shown in italics.

SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

meaning to this since both variables were measured in terms of possessions within the family and may simply reflect parallel views of family wealth.

With regard to the meaning to be assigned to the coefficients themselves, with one or two exceptions the picture that emerges from the analyses is a fairly conventional one. And it seems to apply across the six models estimated--one for each of the three reading comprehension scores in each of the two grades. As one might expect, given the high correlations among the reading comprehension scores, the similarity between the six models is considerable in terms of the overall pattern of effects, although there are some differences across the scales within and between the two grades. We comment first on the common aspects of the level 1 model and then highlight some of the between-model differences that seem to have meaning.

Most of the family effects expected on the basis of past research into the academic achievements of elementary school students are to be found in these analyses as well. With regard to the status attainments of students' families, these figure prominently in the explanation of reading comprehension differences between students (Wigfield and Asher 1984). In the model used, the social side of socioeconomic status was represented by parental educational attainments, while the economic side was captured in a measure of family wealth. Father's education was consistent in its demonstrated positive effect on reading comprehension in both student populations. The effect of mother's education was less consistent across the three measures in each of the two populations. However, without measures of parental occupations in the equation to control for the social status of the family, the attribution of meaning to the difference in effect of father's and mother's education probably is not warranted. Most likely we are seeing some mix of effects from both social and educational attainments.

The economic aspect to socioeconomic status played a part over and above that due to the social/educational status attainments of parents. Other things equal, students from poor families fared less well than did those from economically advantaged backgrounds. This is not a trivial observation: the important point is that the differences in question are taking into account the influences that arise from the other aspects of family background. It follows that we are much closer to providing evidence that family economic circumstance affects reading comprehension in its own right, something that is not always immediately obvious from the observed relationship, which is confounded by the social, educational, ethnic, structural, and language attributes of families.

Family structure seems to matter as well. Consistent with the majority of findings on this matter, students from two-parent families have higher achievements than those from single-parent or blended families. However, one other aspect of family structure demonstrated an unanticipated effect on the development of reading comprehension. The attribute we have called "augmented family" has a negative effect on reading achievement. Other things equal, students whose families include adults in addition to their parents show lower levels of reading comprehension. This effect may have similar origins to the negative effect of family size often reported in the literature. The more children have to share the attention of their parents and parent-child interaction with others (siblings and/or other adults in this case), the lower their achievements. However, multifamily and/or multigeneration households are more characteristic of some subpopulation groups than others. Though many of these confounding effects have been controlled for statistically, we may still be seeing reflections of social, economic, ethnic, and perhaps rural-urban differences not captured in the parental education, family wealth, and ethnicity variables included explicitly in the equation.

Gender differences work in the expected direction, favoring females when the comprehension material is heavily verbal, as in the narrative scale, but disappearing as the verbal content decreases to the level found in the document scale. Other things equal, females do better than males in reading comprehension tasks more or less in proportion to the verbal content of these tasks. Where the verbal content of the comprehension task is low, as in the document comprehension measure, there is no significant difference between males and females among fourth grade students, and among ninth grade students males do significantly better than females. Explanations for these persistent gender differences tend to be cast in terms of gender differences in interests as the intervening variable (Guthrie and Greaney 1991).

Minority status remains a disadvantage for the development and display of reading comprehension skills among fourth graders, even after adjustment for the status attainments of parents and other attributes of students' families. However, the net effect of minority status so estimated is not markedly different in size to a number of the other effects shown. Other things equal, being a minority student is about as much of a disadvantage as being without both parents, or growing up with a language other than English--or being male when the tests are loaded heavily with verbal items.

Among ninth grade students the effects of family attributes seem to be of less importance, as one might expect. The variables in question account for less of the variance in reading comprehension and, overall, the effects appear to be much smaller. This is accounted for, in part, by the fact that ninth grade students are subject to a wider variety of influences than those arising within their families. We may also be seeing the results of applying the somewhat suspect family process model developed for the younger students without change to the older group, where it would be even more suspect. Overall, these data did not allow the development of a particularly persuasive model of the influence of families on the development of reading comprehension skills among ninth graders.

14.9.5. Displaying Effect Estimates

While we have been circumspect about the development of these analyses and the meaning that one can assign to the results, many of the findings themselves are of interest. However, as presented to date, their impact is probably not as strong or broad as it might be, for at least two reasons. First, for statistical and computational convenience, in the development and estimation of the models some simplifying assumptions were made with regard to the form of the variables used. Basically, ordinal scales were treated as continuous if there was a reasonable number of categories; nominal variables, with the exception of region, were dichotomized; and ordinal variables with a limited number of categories were also treated as dichotomies. While this treatment of the variables seemed necessary to facilitate the extended series of exploratory analyses, it may obscure potentially interesting detail in the findings that have emerged. Second, all of the continuous variables were without any natural and concrete metric -- factor scores, for instance. As such they provide for interpretations with less substance to them than one might like -- interpretations based on standard deviation units do not have the same impact as when the units are dollars, or years of education, or some other unit with a more worldly reference.

With the view to remedying this situation in the general model, the form of selected variables was expanded to provide for a more detailed look at their effects on reading comprehension. The variables in question were identified on the basis that they had effects, and that the display of these effects would be of interest relative to what is known, suspected, or not known about the sources of variability in reading comprehension.

More specifically, in the case of dichotomies like father's education, for example, the dichotomy high school or less/more than high school was expanded into four categories--less than high school, high school diploma, high school diploma plus further education, and college degree. In the case of continuous variables, such as the factor score for wealth, we categorized the distribution into quartiles with the view to talking about the lower quartile (somewhat loosely) as poor and the fourth quartile equally loosely as rich, with quartiles 2 and 3 having intermediate positions.

For the purpose of estimation these categories were treated as dummy variables, and the HLM equations were reestimated. However, since coefficients for dummy variables have somewhat awkward interpretations, in the interests of injecting more meaning into the discussion of results the dummy variable coefficients were transformed into adjusted category means. By so doing, summary statistics (slope coefficients) that are unfamiliar to most nonacademic readers were turned into statistics that are part of everyday discourse (means). As well as being easier to understand, such presentations have decidedly more impact than tables of coefficients, although, unfortunately, statistical parsimony is sacrificed in the process. If the intent is to disseminate the findings more broadly than an academic style of presentation will allow, the additional effort may be justified.

In practice this form of statistical display means that the single coefficient of 6.9 shown in Table 14-15 as the effect of father's education on narrative comprehension is now represented as four adjusted category means. In the case of father's education this implies that mean reading comprehension scores are reported separately for students whose fathers had not completed high school, had completed high school but gone no further with education, had further education beyond high school, or had completed college. Further, these category means are to be adjusted means in the sense that the confounding influences of other variables within the model will be controlled statistically. They are estimates of what the reading comprehension levels would be, other things equal, in each of the categories.

In the student-level model there are some 15 coefficients by three scales by two populations; accordingly, this kind of presentation is used sparingly in the discussion that follows. However, the statistical presentation in a subsequent report aimed at providing a comprehensive statement of the findings of the study in a form suitable for a broader audience will take a similar approach and provide a greater range of presentations of this kind.

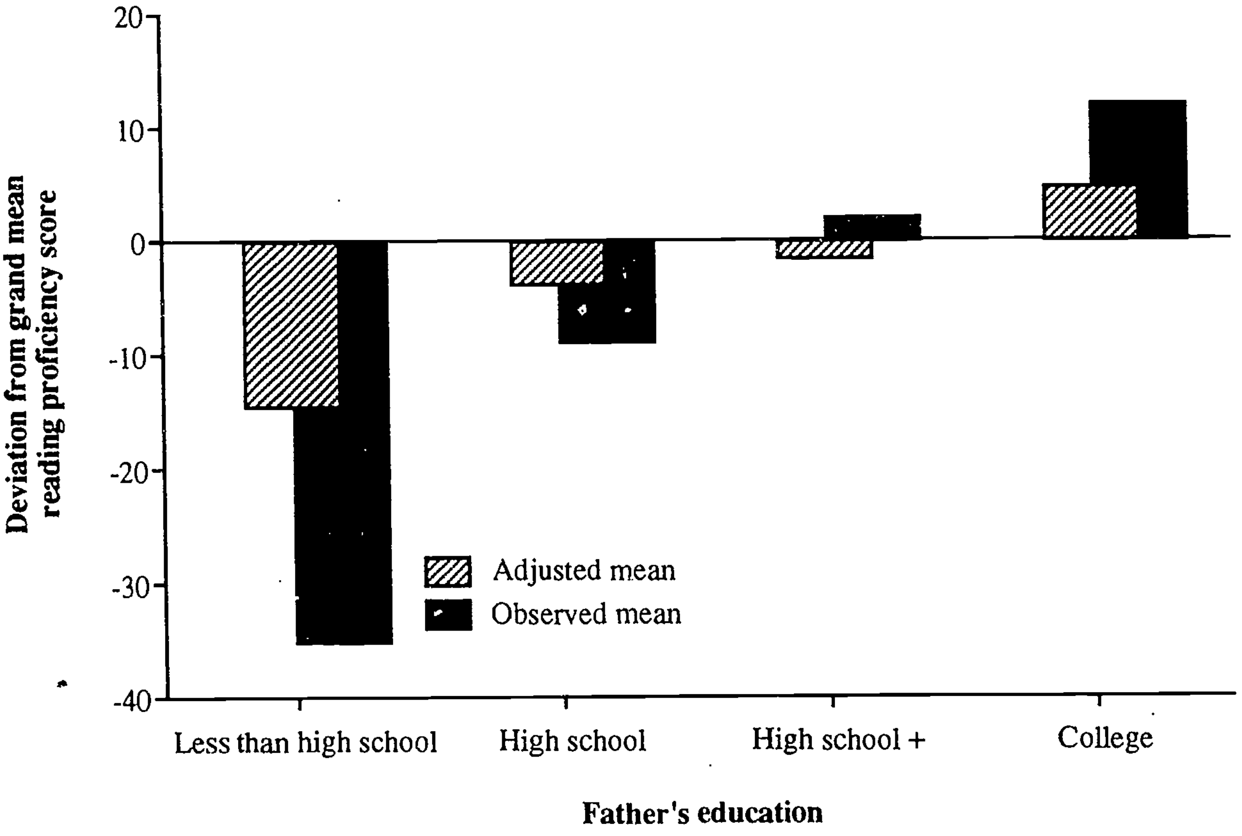
In a further simplification of the statistical presentation we show these category means as deviations from the grand mean, so that category means below the grade average are immediately obvious. Those groups doing better than one might expect on the basis of the grade mean are equally obvious above the line. The standard errors of these estimates are not shown on the display itself since they complicate the visual simplicity of the display. These data are available in the appendix to this chapter, along with an explanation of the way in which these standard errors were estimated -- a process not entirely straightforward. However, the statistical significance of the differences between groups, based on Bonferroni-adjusted pairwise multiple comparisons of these category means, is identified in the figures.

14.9.5.1. Father's Education and Reading Comprehension

The display of estimates for the effect of father's education is shown in Figure 14-6. Both observed and adjusted category means are shown, which allows for an immediate comparison of the effect of adjusting for the other exogenous variables. In this kind of display the effect on -- as distinct from the relationship to -- reading comprehension of father's educational attainments is immediately obvious. The (observed) relationship between father's education and reading comprehension overestimates the actual effect of father's education adjusted for the other confounding influences specified.

The trend in both the observed and adjusted means is obvious and in accord with what one would expect -- the children of better educated parents do better in school. As indicated in Figure 14-6, the only significant difference in these adjusted category means is between the two extremes. Other things equal, the reading comprehension levels of students from the least educated of families are significantly lower than those of students whose parents are college educated. The usual explanations lean toward social learning theory concepts, with better educated parents providing more appropriate stimuli, models, and reinforcement for, in this case, the comprehension of narrative prose. Since father's education may also be tapping aspects of social status, we are probably looking also at the traditional effect of family SES on school-related achievements.

Figure 14-6. Observed and adjusted narrative reading comprehension proficiency mean scores, by father's education: Grade 4

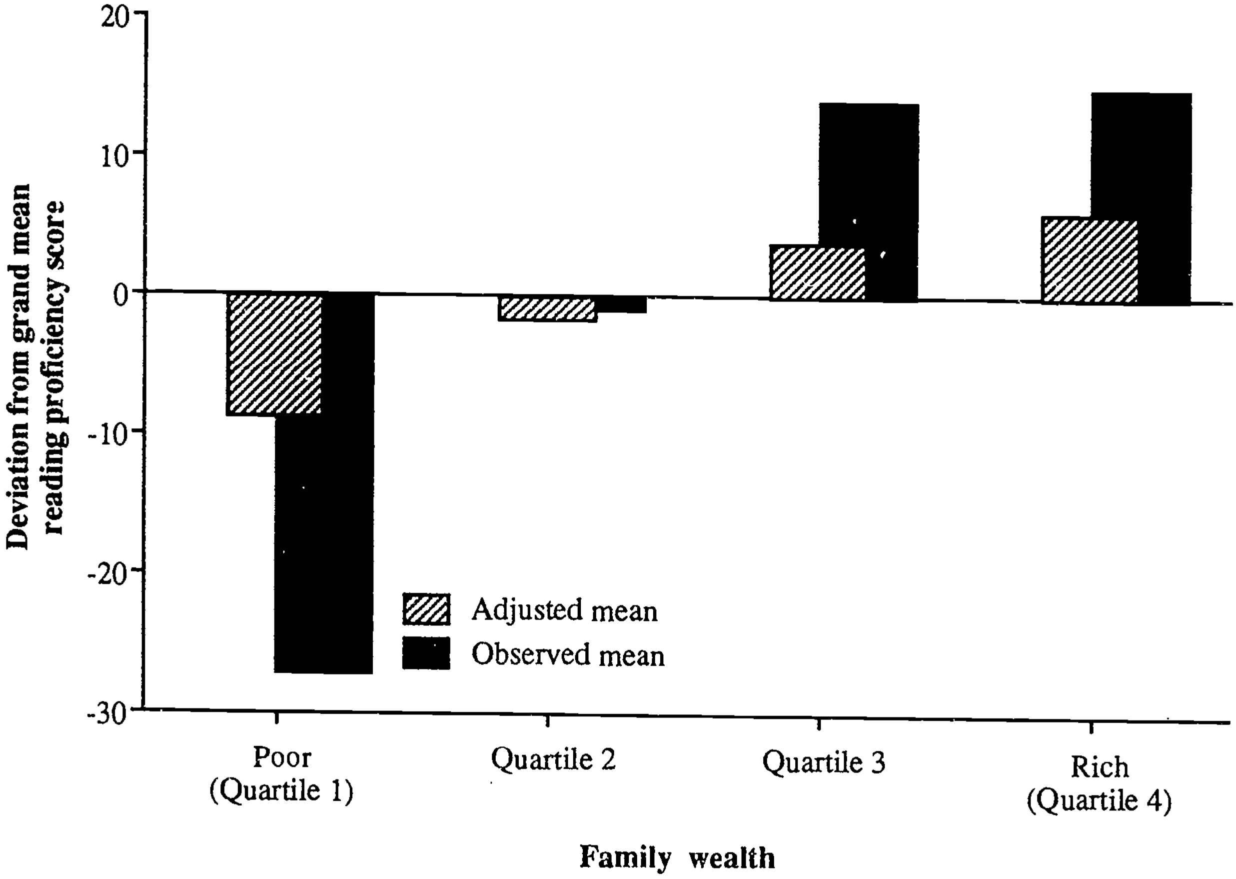


SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

14.9.5.2. Family Wealth and Reading Comprehension

Given the importance of family wealth in most thinking about the sources of educational disadvantage, we display its effects on narrative comprehension in the same way. The adjusted means presented in Figure 14-7 make clear that students from poorer families show the lowest levels of reading comprehension, at least in fourth grade. On the surface this seems less than news since the conventional wisdom is that economic disadvantage is a substantial handicap during schooling.

Figure 14-7. Observed and adjusted narrative reading comprehensive proficiency mean scores, by family wealth: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

However, much of this conventional wisdom is based on the simple observation that rich students outperform poor students on a variety of measures. The information displayed in Figure 14-7 bears directly on this point. The solid bars show the observed category means (as deviations from the grand mean) by level of family wealth, and their trend is indicative of the strength of the observed relationship. The hatched bars take into account the fact that family wealth is related to parental education, minority status, language background and so on, which may themselves actually be the factors that affect the reading comprehension of students.

A comparison of the two sets of bars suggests quite clearly that one would be led astray if, for example, these data were interpreted to mean that poor students score close to 30 points below the grand mean because they are poor. Being poor is part of it, but it is not the whole story. In fact, within the context of this model, other things equal, poor students score less than 10 points below the grand mean. The difference between this value and the observed value is due to other aspects of family background and schooling related to family wealth. In terms of statistical significance, those in the first quartile are reliably different from those in either quartile 3 or 4. Other things equal, reading disadvantage seems to be located among the poorest 25 percent of families. However, the degree of disadvantage itself is not of major proportions; other things equal, poor fourth graders score about 15 points below rich fourth graders, on average, in a test whose standard deviation is 100 points.

14.9.5.3. The Effect of Minority Status

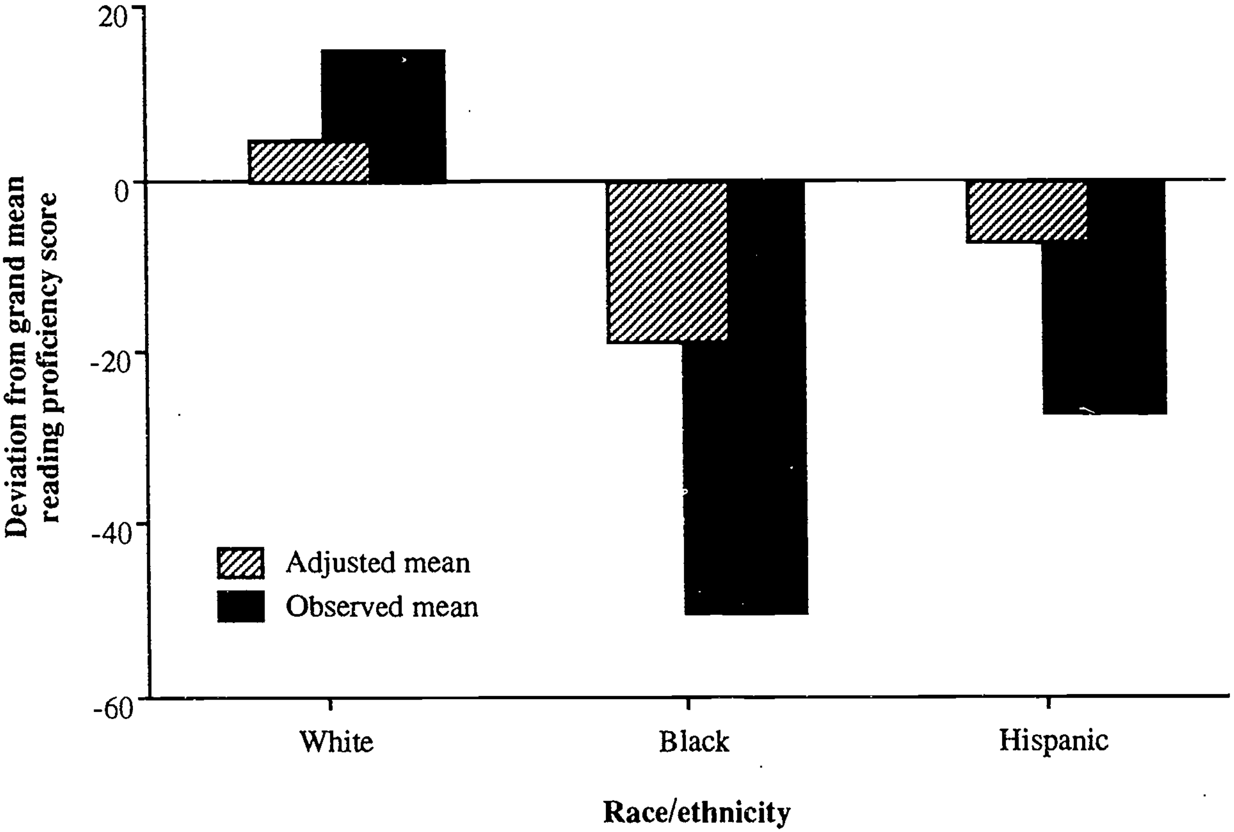
The effect of minority status on reading comprehension, as indicated in Tables 14-15 and 14-16, is consistent across the three comprehension measures and across the two student populations. When other aspects of family background are taken into account, it is an effect similar in magnitude to other attributes of students and their families -- gender, family composition, and language, for example. We have broadened the detail of this variable from a simple dichotomy to one in which there are three categories of race/ethnicity; white, black, and Hispanic. This allows a closer look at the way in which the two largest minority groups differ from the white majority and from each other (Figure 14-8). We have not shown estimates for American Indians and Asians, the other two categories, as their numbers are small and the standard errors large as a consequence.

Recall that minority status was treated as a random variable, one whose effect on reading comprehension was assumed to vary across classrooms. For reasons we need not consider at this time, we are unable to treat minority status in the form of a random set of dummy variables. The meaning of the coefficients changes from the original under these conditions. So, to provide some notion of the observed and adjusted differences between the two minority groups and the white majority, we have treated the dummy variables as fixed and unvarying across classrooms in this instance.

Displaying these estimates of the effect of minority status as adjusted deviations from the grand mean for each of the four categories of race/ethnicity demonstrates again the general point about confounding influences quite clearly, as Figure 14-8 shows. The observed differences between minority and nonminority students are large--the difference between blacks and whites is some two-thirds of a standard deviation. Taking into account that other aspects of family background are associated with minority status gives a better indication of the size of the "true" effect of minority status, which, in this model, is substantially less than the observed effect. An observed black-white gap of 65 points is reduced to one of 24 points after we take into account the other aspects of family background that go along with minority status. This gap between the reading comprehension levels of blacks and whites is a statistically significant one and is the only one of the three differences to reach statistical significance.

However, just what this true effect means is another matter as it is model-specific. At one level it means that it is an effect controlling for the confounding influence of the other aspects of family background specified. However, it is not an effect that is controlling for everything, since we have not measured everything in this model. One might speculate on other differences between minority and nonminority students -- the use of nonstandard English, for example (Heath 1991) -- which, if controlled statistically, would result in the disappearance of the comprehension gap shown.

Figure 14-8. Observed and adjusted narrative reading comprehension proficiency mean scores, by students' race/ethnicity: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

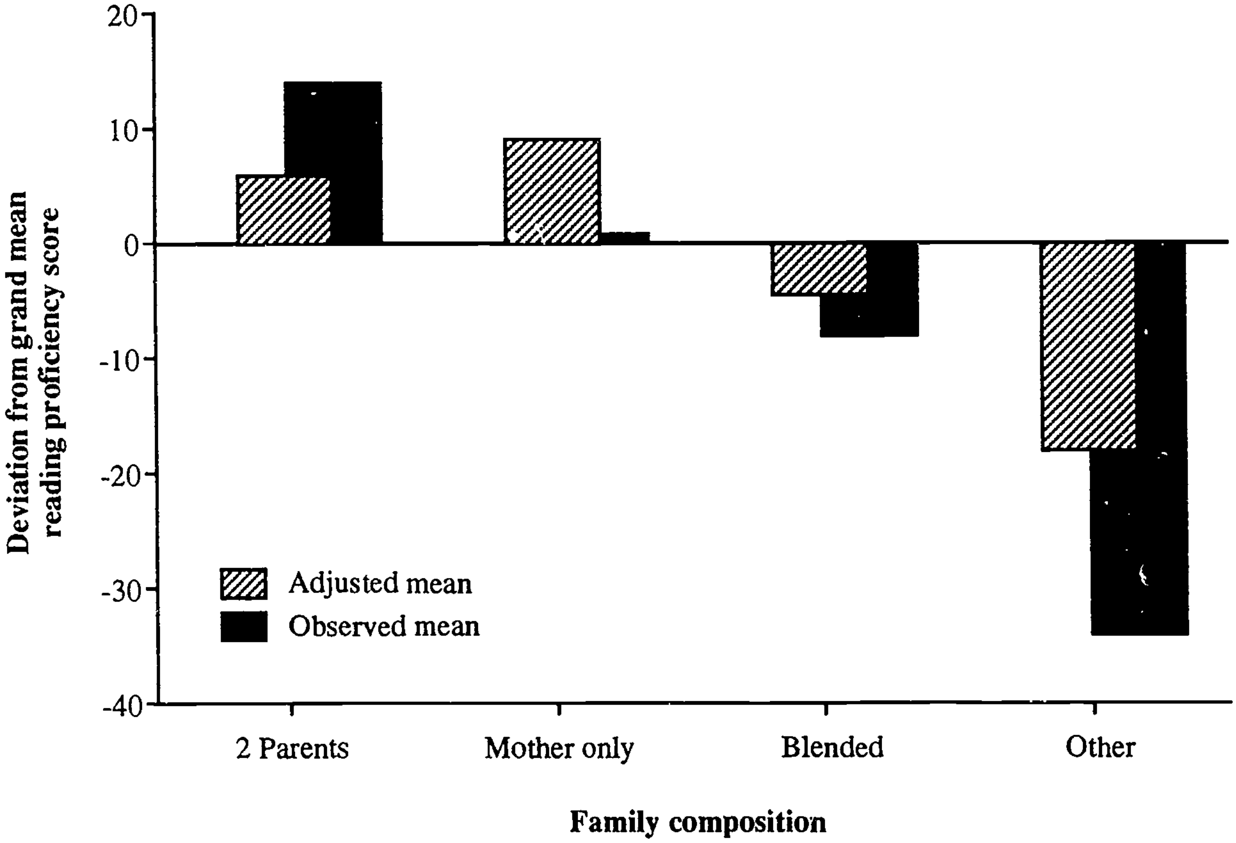
14.9.5.4. Family Structure and Reading Comprehension

The measure of family structure used asked students to report on persons living in the same house. Siblings, grandparents, other relatives and other nonrelatives were included in the list. To define family structure in terms of the nuclear family, a distinction was made between father and mother, on the one hand, and male and female guardians/stepparents/foster parents on the other. Students reporting both father and mother were considered as the standard two-parent unblended nuclear family, although strictly speaking one or both of these parents may not have been the student's biological parents. The important point was that the student saw these adults as parents and not as guardians, stepparents, or foster parents. Single-parent families were distinguished an analogous way, and the definition of blended families follows the same logic. "Other" refers to other reported combinations of parents/guardians, many of which appeared to be improbable to say the least. Observed and adjusted category means for narrative comprehension in the fourth grade population are displayed in Figure 14-9.

Our knowledge about child development generally, and about the development of reading comprehension in children, points to the apparent importance of a stable complement of parents with whom the child can interact during the development of language (Milne 1989). Family disruption through family breakup and reformation is seen as a source of disturbance to this learning process, other things equal (Saltzman 1987). Single-parent families produced by family disruption, or by choice, are thought to be beset by additional problems, many of which stem from the relative economic disadvantage that attends this type of family structure.

The observed category means shown in Figure 14-9 support this view. Students from two-parent unblended families do best, those from one-parent unblended families are next, children of blended families are third, and those in other arrangements show the lowest average levels of performance. However, we know that single-parent families and blended families differ in other ways from the modal two-parent family. Such families tend to have lower levels of social, educational, and economic attainments, and to have minority groups overrepresented among them. Thus, it is conceivable that the between-family differences in reading comprehension observed are due to these factors rather than to the family configuration as such.

Figure 14-9. Observed and adjusted narrative reading comprehension proficiency mean scores, by family composition: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

This point is illustrated nicely in the comparison of the observed and adjusted means for students from single-parent families, all of which are headed by mothers. When we take these other related aspects of family and school into account, students from these families actually do better than one would expect on the basis of simple observation. In terms of fostering reading comprehension in children, after allowance is made for the attendant disadvantages of single-parent family arrangements, they do as well as the model two-parent family.

14.10. Between-Classroom Differences in Reading Comprehension

The notion basic to the structure of the model developed here was that student variation in levels of reading comprehension had both individual and group components. With regard to the individual component the argument was as follows: part of the reason that students vary in their level of reading comprehension is that students and their families vary in terms of the characteristics that contribute to the development of reading comprehension. This variation is uniquely associated with each student. The discussion immediately above focused on some of the findings resulting from the attempt to model this component of student variation.

The group component to this between-student variation in reading comprehension stems from the fact that students are assigned in groups to classrooms. As a result, students within the same classroom are assumed to be exposed to the same learning environment, but different classrooms have different learning environments. Given this assumption, within any one classroom the quality of the learning environment, that is, the characteristics of the classroom being shared by all students within the group, is the same. Consequently, each student's level of reading comprehension is incremented to the same degree. However, since learning environments differ between classrooms, the size of the increment added to a particular student's comprehension level depends on the classroom in which he or she is located.

We attempted to model this group-derived component of the variation in students' reading comprehension by modeling the variation between classrooms in both the average level of reading comprehension and the effect of minority status on reading comprehension. These differences between classrooms were set within a model that postulated their origins in aspects of community context, school characteristics, school leadership, classroom learning environments, and teacher attributes.

The data defined these sources of between-classroom differences, for the most part, in terms of attributes of communities, schools, principals, classrooms, teachers, and learning-related resources available. Additionally, detailed data on teachers' instructional practices were available, but the analyses failed to demonstrate that these instructional behaviors had any effects on levels of comprehension in classrooms. We considered this to be an artifact of the research design rather than a reflection of reality.

The analyses were similarly unsuccessful in accounting for the fact that the handicap attending minority status varies between classrooms. This variation is an important phenomenon that calls for a greater understanding of its origins than is available at present. Such an understanding clears the way for effective compensatory programs designed to foster equity in this respect.

Thus, the actual model estimated to account for the fact that classrooms differ in their capability to foster the development of reading comprehension focused only on variation in the average level of reading comprehension across classrooms, an issue that is not inconsequential. As noted earlier in connection with the interpretation of coefficients in Tables 14-15 and 14-16, it was possible to explain quite respectable proportions of this between-classroom variation. For the fourth grade population, overall, the exogenous community, school, and principal variable groups explain some 25 percent of the between-classroom variation in average levels of document comprehension. This explanation increases to 35 percent when narrative comprehension is the matter at issue. In contrast, the ninth grade models explain about 10 percent of the variance in each case. (Recall that between-classroom variation amounts to about 20 percent of total variation.)

These differences in explanatory power across the grades are consistent with what one might expect. Reading is taught in elementary schools, but not in high schools. It follows that one should expect the attributes of elementary schools to have a greater effect on classroom levels of reading achievement in the fourth grade than in the ninth grade. Besides, differences between ninth grade language arts classrooms in the reading capabilities of students have other sources as well, among them the selection and tracking of students with different levels of reading competence.

In fourth grade classrooms, levels of reading comprehension vary with the region of the nation in which the students are located, being notably lower in the Southeast and higher in the Northeast (relative to the West), other things equal. They vary too with the level of parental cooperation experienced by the school, with school policy-dictated attributes like instructional time and library resources, the latter probably a good indicator of general resource levels of schools and not simply a matter of reading resources as such.

The all-variable equation adds to the list of exogenous variables four endogenous variables--three attributes of classrooms, class size, percentage of remedial students and proportion of minority students, along with teacher gender. In the fourth grade population, the addition of the endogenous variables to the equation increases the proportion of variance explained quite substantially--to 60 percent or more. Class size shows negative effects on all three reading comprehension scores, as does the proportion of remedial students in the class. The effect of the minority composition of the class is important as well; in fact, this variable adds between 15 and 25 percentage points to the total variance explained by the combined exogenous and endogenous variables. In addition, teacher gender shows statistically significant effects favoring female teachers.

In the ninth grade population the addition of the endogenous variables has the effect of increasing the proportion of explained variance by about 40 percentage points in each of the three equations, to the extent that the model explains quite respectable proportions of the variance in classroom means -- between 50 and 60 percent. However, as in the fourth grade, a good part of this increase is due to the compositional variable (a little less than 20 percentage points). Among the remaining endogenous variables, only the remedial student variable percentage of remedial students has consistent statistically significant effects. However, since we are predicting class averages, a negative effect of the proportion of remedial students in the class is not especially informative; other things equal, the more remedial students in the class, the lower the classroom average.

It is worth noting that the introduction of the endogenous variables into the equation has a noticeable effect on the coefficient for the Southeast. Once these endogenous variables are controlled, the size of the coefficient for this dummy variable decreases substantially from its value in the exogenous equation -- to about half. The implication is that part of the total degree of disadvantage that comes from living in the Southeast is a function of regional differences in the endogenous variables, among them the proportion of minority students in the class.

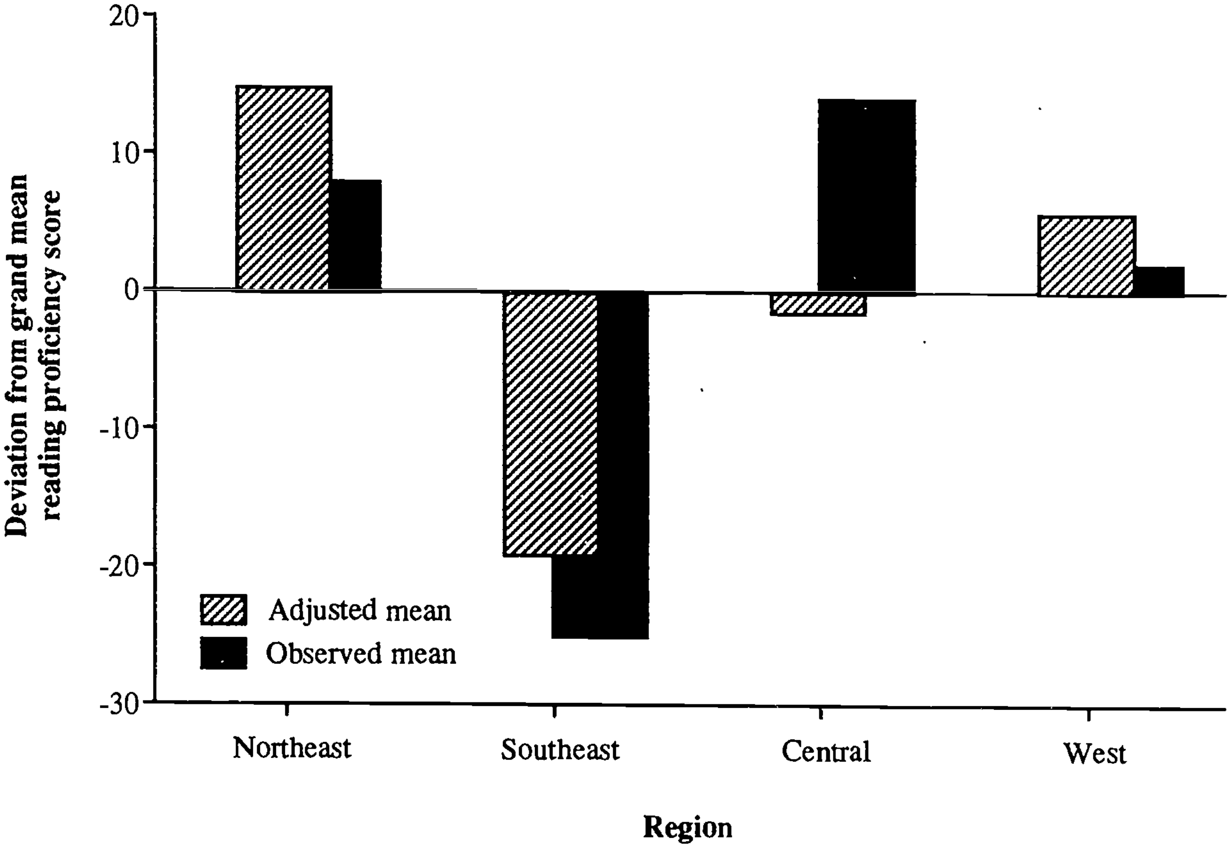
14.10.1. Selected Classroom-Level Effects on Reading Comprehension

As in the case of the student-level model, for some of the classroom-level effects it is instructive to display some of these coefficients in more detail as adjusted category means. Four effects seem of particular interest from the point of view of their implications for educational policy -- region, parental cooperation, instructional time, and class size. As before, these presentations are limited to the results for narrative comprehension for fourth grade students.

14.10.1.1. Regional Differences in Reading Comprehension

The observed differences between the four regions are substantial and of the same order as one usually sees. However, the meaning of these differences is unclear. Regions also differ in terms of the makeup of their populations and their school systems. Figure 14-10 illustrates the observed and adjusted reading comprehension means for classrooms in each region. The adjusted means go some way toward isolating differences that are regional per se from those due to regional differences in population characteristics and schooling in the aggregate. The statistical significance of the differences between these adjusted means shows the Northeast and Southeast as being different from the Central and the West, and from each other.

Figure 14-10. Observed and adjusted narrative reading comprehension proficiency mean scores, by region: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

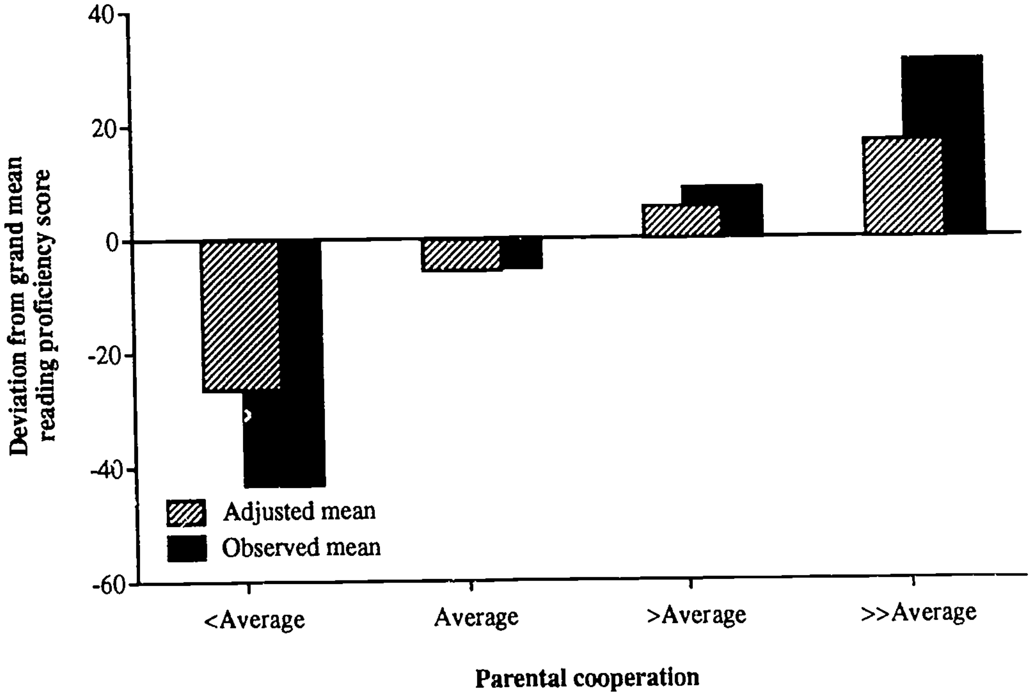
The adjusted data show, for example, that the apparent advantage of living in the Northeast and the disadvantage of living in the Southeast is not as indicated by observed values. In fact these data suggest that if the several attributes of students and classrooms specified by the model were equalized across the nation, classrooms in the Northeast and Southeast would do better than they do now. Since classrooms in the Northeast are apparently advantaged anyway, such an equalization would produce further inequality between regions, though in the process those in the Southeast would be less disadvantaged. Whatever it is that accounts for these regional differences in achievement is, by definition, not included in the model. While we do not know what is responsible for the adjusted regional differences displayed, we do know that it is not regional differences in the attributes of schools, classrooms, and teachers explicitly included in these analyses.

14.10.1.2. Community Support and Reading Comprehension

Local support for education and local control of the provision of education are considered to be part of the American way. While schools are thought to be a reflection of their communities and a marker of their success and affluence, it is not always clear that this view is universal. At a more theoretical level, Coleman and Hoffer (1987) develop the notion of social capital within and beyond the family in this context. Where human capital concerns an individual's skills and capabilities that result from personal investment and can be traded on the open market to produce a return on the investment, social capital is the analogue for social groups: "social capital is less tangible...it exists in the relations between persons...for example, trust is a form of social capital" (Coleman and Hoffer 1987, 221).

In a limited way, we have examined the importance of social capital, in the form of community support for schools, for the development of reading comprehension among students in the selected classes within the school. And the effect is noteworthy, as Figure 14-11 makes clear.

Figure 14-11. Observed and adjusted narrative reading comprehension proficiency mean scores, by degree of parental cooperation: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

The observed relationship between parental support for the school and the mean level of reading comprehension of at least one class of students in the school is quite substantial. Even after adjustment for confounding influences, the relationship remains. Under these circumstance, it seems reasonable to talk about an effect of parental support for a school on the development of reading comprehension within that school. From the point of view of statistical significance, those schools reporting less than average levels of parental support show significantly lower levels of reading comprehension.

In schools where parental support is below average, so is the average level of reading comprehension of the students, even after making allowance for the confounding of this relationship by the variety of other variables noted. Conversely, where parental support for schools is high, the reading comprehension of fourth grade students is enhanced. Further, this effect seems to be confined to fourth grade classrooms, perhaps because elementary schools are more closely identified with their immediate community than are larger high schools, which draw on several communities.

14.10.1.3. Instructional Hours and Reading Comprehension

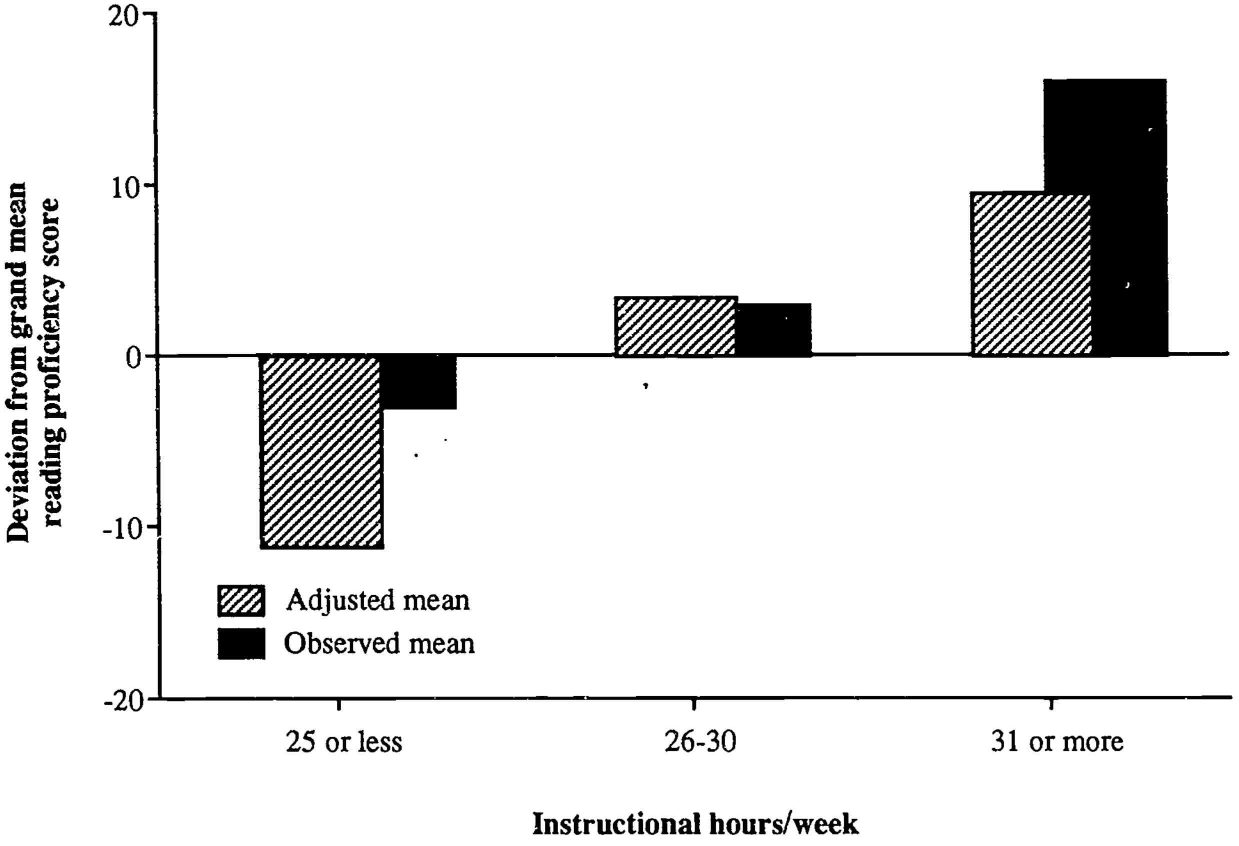
School principals were asked to report on instructional time as a response to the question about the total instructional time made available by the school each week for all subject areas. We saw this variable fitting within the context of the kind of time-on-task concerns typically associated with the well-known Carroll model (Carroll 1963). In this instance, the focus is on Carroll's "time allowed," and arguments on this matter cover the range from length of the school year (Harnischfeger and Wiley 1977) through the length of the school day to the amount of time students actually spend in learning specific subject matter. The general notion, of course, is that the more time available for learning, the more extensive that learning will be, other things equal.

In this particular case we are looking at the question of whether, other things equal, a school that provides for more instructional time in each school week also has higher levels of reading comprehension. And the answer seems to be that it does. Figure 14-12 displays observed and adjusted means for each of the three categories created from this continuous variable. Those comparisons, which are significantly different from each other, are indicated in the usual way.

On the basis of simple observation, we would conclude that schools that offer more instructional time each week also have classrooms with higher reading comprehension levels than do schools offering less time. More specifically, schools that provide more than 30 hours per week of instructional time do much better in this respect than those who provide less. However, following the same general line of argument as before, it is likely that these observed differences are confounded by other attributes of schools and classrooms that also affect reading comprehension levels.

Controlling statistically for these, at least insofar as they are represented in the model, does not change the general conclusion. Other things equal, the more instructional time offered by schools the higher the average level of achievement of their students. In the case of the over-30-hours schools, the effect is not as great as we might infer from simple observation alone. Other attributes of these schools related to instructional time are also playing a part. In the case of schools offering 25 hours a week or less, the effect of controlling for confounding influences is to bring out the fact that these confounding influences are actually compensating for the disadvantaging effects of limited instructional time. If other things were equal, these "low-time" schools would do much worse. Nevertheless, one thing is clear: the more time that schools provide for instruction the more likely this instruction is to take hold and be manifested as higher levels of student achievement.

Figure 14-12. **Observed and adjusted narrative reading comprehension proficiency mean scores, by hours of instruction per week: Grade 4**



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

14.10.1.4. Class Size and Reading Comprehension

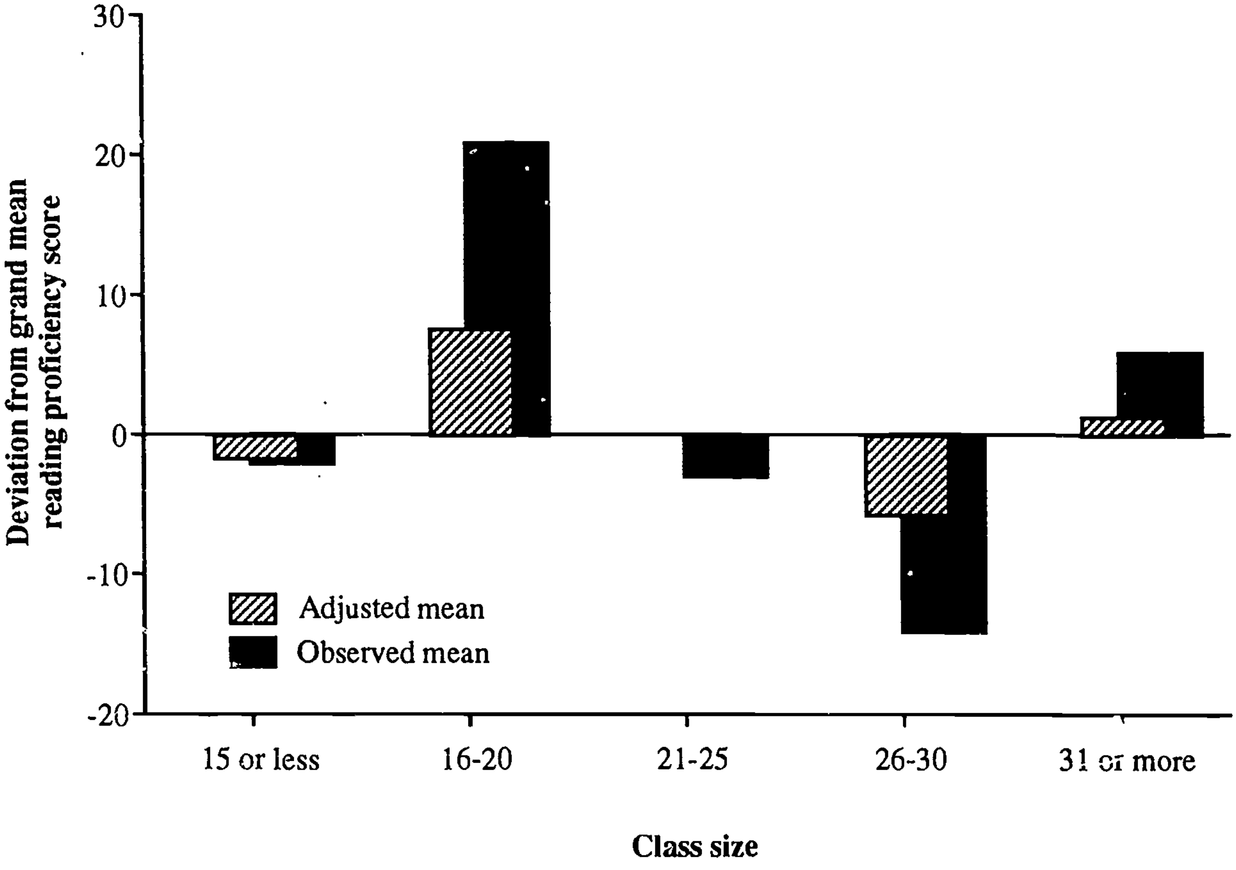
The relationship between class size and achievement has a lengthy and contentious history in educational thought. Large-scale studies demonstrate consistently a negligible or even a positive relationship, though the latter is usually explained in terms of the small size of remedial classes. Teachers and teacher unions are convinced that the relationship is negative, pointing to the difficulties of providing for individual differences in large classes, and asserting that the quality of learning (and teaching) improves as the size of the class gets smaller.

Glass and Smith (1979) reviewed an extensive collection of studies of this phenomenon. Their meta-analysis suggests that when the quality of the various research designs is taken into account, the general conclusion is that learning is better in smaller classes. However, the relationship is not linear. In classes ranging in size from only a few students up to 15 or so, the relationship is strong and in the expected direction. However, there appears to be a threshold in the sense that the effect of increasing class size beyond this point -- from 20 to 30 or 40 -- is relatively minor, though still detrimental to learning.

In fourth grade classes we find the picture about as confusing as most studies report. While the coefficient for the continuous measure of class size was statistically significant -- and negative, in the case of fourth grade classrooms -- suggesting a decrease in achievement with an increase in class size, the implication of the analyses using a categorization of this variable is less clear. Figure 14-13 illustrates this finding for narrative comprehension.

We can say, as before, that the observed relationship between class size and achievement is somewhat misleading as an indicator of the effect of class size on achievement. Statistical adjustment for factors related to both class size and achievement shows that the differences between the categories are reduced. However, with regard to the overall effect of class size, the differences between categories are without a clear trend, although the only statistically significant difference is in the right direction. Other things equal, students in classrooms with about 15 to 20 students do better than those in classrooms with 25 to 30 students. In all, though, we are probably looking at the effects of a variety of unmeasured influences on class size--tracking by ability among them--such that classes of the same size are created for different reasons and with students of different ability levels, a situation guaranteed to confuse the picture, as it appears to have done.

Figure 14-13. Observed and adjusted narrative reading comprehension proficiency mean scores, by class size: Grade 4



SOURCE: IEA Reading Literacy Study, U.S. National Study data, National Center for Education Statistics, 1991.

14.11. Observations on the Modeling of Reading Comprehension

The text has made clear at various points that we did not see this modeling exercise as an unalloyed success by any absolute standard. The overall study seemed to be characterized by an understanding that the data were generated by a single general model with, perhaps, a few ancillary pieces grafted on to account for special interests. Given the amount of data collected, and the fact that the conceptual and measurement structures underlying the research design were largely unspecified in any formal sense, the modeling of reading literacy was something of a challenge. At the operational level, this meant that we engaged in a good deal of inference and exploration, both conceptual and statistical, to arrive at a model.

The end result, however, is probably the most comprehensive statistical model of reading comprehension ever developed. There is nothing of this level of complexity in the literature on reading. What is more, there are findings of interest and with policy relevance. Further, these findings have a statistical foundation that makes their interpretation in these respects less unequivocal than most. We do not base conclusions on simple relationships among variables; rather, we set these relationships within a common sense theoretical model of the sources of variation in reading comprehension. This model dictates the structure of statistical analyses designed to estimate the effects of variables uncomplicated by confounding influences with the result that we are better able to interpret relationships as effects, which, after all, is the point of the exercise.

This model building and associated statistical analysis was undertaken using somewhat technical language to describe the "theoretical" conceptions and statistical estimates. We considered how both of these might be made more accessible to persons not entirely familiar with this language. The approach adopted is only illustrated in this volume, but was developed in more detail in a separate publication designed to provide a wider dissemination of these findings using statistical presentations designed for a broad, general audience.

The development of the analyses is reported in some detail for at least two reasons. First, we engaged in a good deal of statistical exploration during the course of model development, though for the greater part this was exploration within theoretically defined limits. In more than a few instances our use of statistical techniques, particularly tests of significance, would not be sanctioned by statistical purists. It is important that this be made explicit so that the reader can judge the extent to which the conclusions reached are compromised by this process.

The other way to look at this development process is to think of it as quite reasonable under the circumstances. To the extent that this is seen to be the case, this detailed reporting provides a model to guide similar analyses. Others engaged in analyses of this kind, with IEA data in particular, may find the detail instructive as a model of the application of multilevel modeling to data of the kind that has become traditional within IEA studies.

References

Alexander, K.L., and Eckland, B.K. (1975). Contextual effects in the high school attainment process. *American Sociological Review*, 40, 402-416.

Arnold, C.L., Kaufman, P.D., and Sedlacek, D.S. (1992). *School effects on educational achievement in mathematics and science: 1985-86.* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Bidwell, C. (1965). The school as a formal organization. In J.G. March (ed.), *Handbook of Organizations*, 972-1002. Chicago: Rand McNally.

Bielby, W.T., and Hauser, R.M. (1977). Structural equation models. In Inkeles, A. (ed.), *Annual review of sociology*, 137-161. Palo Alto: Annual Reviews, Inc.

Bryk, A.S., and Frank, F. (1991). The specialization of teachers' work: An initial exploration. In S.W. Raudenbush and J.D. Willms (eds.), *Schools, classrooms and pupils: International studies of schooling from a multilevel perspective*, 185-204. Orlando, FL: Academic Press.

Bryk, A.S., and Raudenbush, S.W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage Publications.

Bryk, A.S., and Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147-158.

Bryk, A.S., Raudenbush, S.W., Seltzer, M., and Congdon, R. (1988). *An introduction to HLM: Computer program and user's guide.* Chicago: University of Chicago.

Bryk, A.S., and Thum, Y.M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, 26, 353-383.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, 8, 158-233.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.

Coleman, J.S., and Hoffer, T. (1987). *Public and private high schools: The impact of communities.* New York: Basic Books.

Davis, J.A. (1966). The campus as a frogpond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, 72, 17-31.

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57, 812-828.

Glass, G.V., and Smith, M.L. (1979). Meta-analysis of research on the relationship of class size and achievement. *Evaluation and Policy Analysis*, 1, 2-16.

Guthrie, J.T., and Greaney, V. (1991). Literacy acts. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research.* Vol. 2, 68-96. New York: Longman.

Hannan, M.T. (1971). *Aggregation and disaggregation in sociology.* Lexington, MA: Heath-Lexington.

Hannan, M.T., and Burstein, L. (1974). Estimation from grouped observations. *American Sociological Review*, 39, 374-392.

Harnischfeger, A., and Wiley, D. (1977). Time allocations in 5th grade reading. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Hauser, R.M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology*, 75, 645-654.

Hauser, R.M. (1974). Contextual analysis revisited. *Sociological Methods and Research*, 2, 365-375.

Keeves, J.P., and Morgenstern, C. (1992). Attitudes towards science: Measures and effects. In J.P. Keeves (ed.), *The IEA study of science III: Changes in science education and achievement: 1970 to 1984*, 122-140. Oxford: Pergamon Press.

Keeves, J.P., and Saha, L.J. (1992). Home background factors and educational outcomes. In J.P. Keeves (ed.), *The IEA study of science III: Changes in science education and achievement: 1970 to 1984*, 165-186. Oxford: Pergamon Press.

Kelley, H.H. (1952). Two functions of reference groups. In T.M. Newcomb and E.L. Hartley (eds.), *Readings in social psychology*, 410-414. New York: Holt, Rinehart and Winston.

Kemper, T.D. (1968). Reference groups, socialization and achievement. *American Sociological Review*, 33, 31-45.

Lee, V.E., and Bryk, A.S. (1989). The effects of high school academic organization on the social distribution of achievement. *Sociology of Education*, 62, 172-192.

Lee, V.E., Bryk, A.S., and Smith, J.B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (ed.), *Review of Research in Education*, Vol. 19, 171-267. Washington, DC: American Educational Research Association.

Longford, N.T. (1988). Fisher scoring algorithm for variance component analysis of data with multilevel structure. In R.D. Bock (ed.), *Multilevel analysis of educational data*, 297-310. Orlando: Academic Press.

Mason, W.M., Wong, G.M., and Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (ed.), *Sociological methodology 1983*, 72-103. San Francisco: Jossey-Bass.

Mason, W.M., Anderson, A.F., and Hayat, N. (1988). *Manual for GENMOD.* Ann Arbor: University of Michigan.

Medrich, E.A., and Griffith, J.E. (1992). *International mathematics and science assessments: What have we learned?* (U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.

Meyer, J. (1970). High school effects on college intentions. *American Journal of Sociology*, 76, 59-70.

Milne, A.M. (1989). Family structure and the achievement of children. In W. Weston (ed.), *Education and the American family*. New York: New York University Press.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988). *School matters: The junior years*. Wells: Open Books.

Murnane, R.J. (1986). Comparisons of public and private schools; what can we learn? In D.C. Levy (ed.), *Private education*, 153-169. Oxford: Oxford University Press.

Phillips, L.M., and Bolt, D. (1992). *Family variables which impact on reading literacy. A review.* Paper prepared for Westat.

Postlethwaite, T.N., and Wiley, D.E. (1992). *The IEA study of science II: Science achievement in twenty-three countries*. Oxford: Pergamon Press.

Rabash, J., Prosser, R., and Goldstein, H. (1989). *ML2: Software for two-level analysis: User's guide.* London: University of London.

Raudenbush, S.W., and Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.

Raudenbush, S.W., Rowan, B., and Cheong, Y.F. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education*, 65, 150-167.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.

Rosier, M.J., and Keeves, J.P. (eds.) (1991). *The IEA study of science I: Science education and curricula in twenty-three countries*. Oxford: Pergamon Press.

Rosier, M.J., and Long, M.G. (1991). *The science achievement of year 12 students in Australia.* Hawthorn, Victoria: American Council for Educational Research.

Saltzman, S.A. (1987). Meta-analysis of studies investigating the effects of father absence on children's cognitive performance. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Schmidt, W.H., and Kifer, E. (1989). Exploring relationships across population A systems: A search for patterns. In D.F. Robertaille and R.A. Garden (eds.), *The IEA study of mathematics II: Contexts and outcomes of school mathematics*, 209-231. Oxford: Pergamon Press.

Stouffer, S.A. (1949). *The American soldier*. Princeton: Princeton University Press.

Wigfield, A., and Asher, S.R. (1984). Social and motivational influences on reading. In P.D. Pearson (ed.), *Handbook of reading research*, Vol. 1, 423-452. New York: Longman.

Appendix A14

Introduction

A common and appealing method of presenting the relationship of variables to outcomes is through the presentation of subclass means, presented as deviations from the population grand mean. This method presents results one variable at a time. In an analogous fashion, one can present the results of multiple linear models, including hierarchical linear models (HLMs) by presenting adjusted subclass means, derived from the model parameter estimates, as deviations from the grand mean. We have used this approach in Chapter 14. This appendix provides details as to how the adjusted means and their standard errors were calculated, and tabulates the full set of adjusted means calculated for the "final" model presented in Chapter 14.

Since the set of predictors used to predict a student's reading performance are correlated, it is useful to construct adjusted means for each predictor to represent the effect of that predictor alone "holding constant" all other predictors in the analysis. For example, minority students are more likely to be in a subclass with lower father's and mother's education than are white students. The multivariate modeling process is essentially one of adjusting for these differences. Therefore, adjusted means can be constructed to estimate what the mean would have been if the subclasses had been exactly like the total population in its distribution over all the other predictor classifications.

The calculation of adjusted means, however, is not available in the program, HLM3, used to conduct the HLM analysis in this study. The process of computing for the adjusted means is included in the next section. The adjusted means discussed in this chapter are expressed as deviation from the overall mean. They are constrainted so that when each subclass is weighted by the proportion in each subclass, the adjusted means sum to zero.

THE CALCULATION OF ADJUSTED MEANS AND THEIR VARIANCES

1. Calculating Adjusted Means and Variances in the General Regression Situation

In the general regression situation, the predicted value $\hat{Y}$ given a specific set of $X's$ (independent variables) is:

$$\hat{Y} = b_0 + b_1X_1 + ... + b_pX_p. \tag{1}$$

The variance of $\hat{Y}$ is:

$$\begin{aligned} var(\hat{Y}) &= var(b_0 + b_1X_1 + ... + b_pX_p) \\ &= X'var(b)X, \end{aligned} \tag{2}$$

where $X' = (1\ X_1\ ....\ X_p)$, and $var(b)$ is the variance-covariance matrice of the regression estimates. The $1 - \alpha$ confidence limits on the true mean value of Y at the specified set of X values are given by:

$$\hat{Y} \pm t_{(n-p-1,\ 1-1/2\ \alpha)} * \sqrt{var(\hat{Y})}, \tag{3}$$

(see Draper and Smith, 1981).

1.1 Situation with Two Classes

For qualitative variables with two classes (i.e., gender), a dummy variable, which takes on values of 0 and 1, is used to represent the classes. For example, let X_1 be a dummy variable with 1 = "female" and 0 = "male", the adjusted value of $\hat{Y}$ for females and males, holding the other $X's$ at their mean values, are

$$\hat{Y}_{Female} = b_0 + b_1(1) + b_2\bar{X}_2 + ... + b_p\bar{X}_p,$$

$$\hat{Y}_{Male} = b_0 + 0 + b_2\bar{X}_2 + ... + b_p\bar{X}_p.$$

The corresponding variances of the adjusted means are:

$$var(\hat{Y}_{Female}) = X'var(b)X, \qquad where\ X' = (1\ 1\ \bar{X}_2\ ....\bar{x}_p),$$

$$var(\hat{Y}_{Male}) = X'var(b)X, \qquad where\ X' = (1\ 0\ \bar{X}_2\ ....\bar{x}_p).$$

The variance-covariance matrice *var(b)*, is available from most regression programs. For example, with PROC REG in SAS, the option COVB requests an output of the estimated variance-covariance matrix of the estimates, and PROC IML can be used to perform the matrix multiplication.

1.2 Situation with Multiple Classes

For a qualitative variable with *c* classes, the *c* classes are represented by *c-1* dummy variables. For example, to represent the qualitative variable mother's education, which is classified into four levels (less than high school, high school, some college, and college), three indicator variables are included. For example, let

X_2 = 1 if less high school,
= 0 otherwise,

X_3 = 1 if high school,
= 0 otherwise, and

X_4 = 1 if some college,
= 0 otherwise.

No dummy variable is used to represent the last class so that none of the *X* variables in the regression equation is a linear function of one or more of the other *X* variables.

For each class of a qualitative variable with multiple classes, the adjusted means can be computed using equation (1). The class means for the four groups with different levels of mother's education are:

$$\hat{Y}_{<H.S.} = b_0 + b_1\bar{X}_1 + b_2(1) + b_3(0) + b_4(0) +.... + b_p\bar{X}_p\ ,$$

$$\hat{Y}_{H.S.} = b_0 + b_1\bar{X}_1 + b_2(0) + b_3(1) + b_4(0) +.... + b_p\bar{X}_p\ ,$$

$$\hat{Y}_{Some\ College} = b_0 + b_1\bar{X}_1 + b_2(0) + b_3(0) + b_4(1) +.... + b_p\bar{X}_p\ ,$$

$$\hat{Y}_{College} = b_0 + b_1\bar{X}_1 + b_2(0) + b_3(0) + b_4(0) +.... + b_p\bar{X}_p\ .$$

The variance of these means can be computed in the same way as in section 1.1.

2. Calculating Adjusted Means and Their Variances in Hierarchial Linear Model

With hierarchical linear models (HLM), the equations for the standard regression models can be generalized and be applied to calculate the adjusted means and variances for classes of qualitative variables estimated as fixed effects parameters. In order to run the HLM program (Bryk, Raudenbus, Seltzer, and Congdon, 1989), the X and Y variables are grand mean centered. Therefore the predicted value of the deviation from the grand mean, $\hat{Y} - \bar{Y}$, is:

$$\hat{Y} - \bar{Y} = \gamma_0 + \gamma_1(X_1 - \bar{X}_1) + ... + \gamma_p(X_p - \bar{X}_p),$$

and the variance of the deviation score is:

$$var(\hat{Y} - \bar{Y}) = Var[\gamma_0 + \gamma_1(X_1 - \bar{X}_1) + ... + \gamma_p(X_p - \bar{X}_p)],$$
$$= (X - \bar{X})'Var(\gamma)(X - \bar{X}) \ ,$$

where $(X - \bar{X})' = [1 \ (X_1 - \bar{X}_1) \ .... \ (X_p - \bar{X}_p)]$, and $var(\gamma)$ is the variance-covariance matrice of the HLM estimates. Unfortunately $var(\gamma)$ is not part of the regular output from the version of HLM used in this study. This is not a problem for the situation with two classes in a qualitative variable, however, this is a problem for the multiple class situation. The following sections discuss the steps used to calculate the adjusted means and their variances in these situations.

2.1 Situation with Two Classes

Using the example in 1.1, X_1 is a dummy variable where 1 = "female" and 0 = "male", and if $\bar{X}_1$ = **0.495** (the proportion of female in sample), the adjusted mean deviation from the grand mean, $\hat{Y} - \bar{Y}$, for females and males are:

$$(\hat{Y} - \bar{Y})_{Female} = \gamma_1(X_1 - \bar{X}_1) = 0.505\gamma_1 \ ,$$

$$(\hat{Y} - \bar{Y})_{Male} = \gamma_1(X_1 - \bar{X}_1) = -0.495\gamma_1 \ .$$

The other variables in the equation drop out because the remaining X variables are evaluated at the mean deviation, which is zero in each case. The variances of $\hat{Y} - \bar{Y}$ for the two classes are:

$$var(\hat{y} - \bar{y})_{Female} = (X_1 - \bar{X}_1)^2 var(\gamma_1) = (0.505)^2 var(\gamma_1) \ ,$$

$$var(\hat{y} - \bar{y})_{Male} = (X_1 - \bar{X}_1)^2 var(\gamma_1) = (-0.495)^2 var(\gamma_1) \ .$$

The standard error (square root of the variance) of γ_1 is part of the standard output from the HLM, therefore these variance can be easily calculated.

2.3 Situation with Multiple Classes

Using the example in 1.2, where X_2, X_3, and X_4 are dummy variables corresponding to the classes where mother's education is less than high school, high school, and some college, let $\bar{X}_2 = 0.09$, $\bar{X}_3 = 0.26$, $\bar{X}_4 = 0.20$, then the corresponding adjusted means for the classes are:

$$(\hat{Y} - \bar{Y})_{<H.S.} = \gamma_2(1 - 0.09) + \gamma_3(0 - 0.26) + \gamma_4(0 - 0.20) \ ,$$

$$(\hat{Y} - \bar{Y})_{H.S.} = \gamma_2(0 - 0.09) + \gamma_3(1 - 0.26) + \gamma_4(0 - 0.20) \ ,$$

$$(\hat{Y} - \bar{Y})_{Some\ College} = \gamma_2(0 - 0.09) + \gamma_3(0 - 0.26) + \gamma_4(1 - 0.20) \ ,$$

$$(\hat{Y} - \bar{Y})_{College} = \gamma_2(0 - 0.09) + \gamma_3(0 - 0.26) + \gamma_4(0 - 0.20) \ .$$

The variance of these means, however, cannot be computed without $var(\gamma)$. To circumvent this problem, the contrast statement in HLM was used to provide a chi-square test of specify linear functions of the $\gamma's$. Then the known relationship that:

$$\chi^2 = \frac{(Y - \bar{Y})^2}{var(Y - \bar{Y})} \qquad (4)$$

is used to calculate the variance.

For example, to calculate the variances of the above adjusted mean deviations, the following contrast statements were specified:

		γ_0	γ_1	γ_2	γ_3	γ_4	...	γ_p
Contrast 1	*<H.S.*	0	0	0.91	-0.26	-0.20	...	0
Contrast 2	*H.S.*	0	0	-0.09	0.74	-0.20	...	0
Contrast 3	*Some College*	0	0	-0.09	-0.26	0.80	...	0
Contrast 4	*College*	0	0	-0.09	-0.26	-0.20	...	0

The chi-square values for each of these contrasts were then used in equation (4) to compute the variance. For example, for the class mother's education less than high school,

$$var(\hat{Y} - \bar{Y})_{<H.S.} = \frac{(0.91\gamma_2 - 0.26\gamma_3 - 0.20\gamma_4)^2}{\chi^2}$$

A HLM input program, and the output generated from running this program, are included in the appendix to illustrate how one can specify tests of contrasts.

A HLM Input Program and Output Generated From the Program

```
_$1$DIA1:[IRLSIMP.HLM PGMS9.TWNEW.MW]S1_02.HLM;7

WGTVAR=cls_wgt
RESFIL=n
BHCNT=10
SEEIT=n
NOBASE=n
outvar=naru
indvar=base/gmdminor/
indvar=gage.f//
indvar=dminor.c//
indvar=gfei1.f//
indvar=gfei2.f//
indvar=gfei3.f//
indvar=gfei4.f//
indvar=gmei1.f//
indvar=gmei2.f//
indvar=gmei3.f//
indvar=gfwlth1.f//
indvar=gfwlth2.f//
indvar=gfwlth3.f//
indvar=gf2comp1.f//
indvar=gf2comp2.f//
indvar=gf2comp3.f//
indvar=gxtnd.f//
indvar=gdlang.f//
indvar=gsex.f//
numit=100
seeit=n
resfil=n
fixtau=3
homvar=y
title=(s1_02.hlm): dummy vars for model1
output=d61:[irlsimp.hlm.out4]model1b.out
FIXTAU=3
HYPOTH=y
GAMMA1=0,0,0,0,0,0,0,0,-0.092,-0.261,0.796,0,0,0,0,0,0,0,0,0
GAMMA2=0,0,0,0,0,0,0,0,-0.092,-0.261,-0.204,0,0,0,0,0,0,0,0,0
GAMMA3=0,0,0,0,0,0,0,0,0.908,-0.261,-0.204,0,0,0,0,0,0,0,0,0
GAMMA4=0,0,0,0,0,0,0,0,-0.092,0.739,-0.204,0,0,0,0,0,0,0,0,0
STOPVAL=0.00005
MODEL=m
```

BEST COPY AVAILABLE

```
_$1$DIA1:[IRLSIMP.HLM.OUT4]MODEL1I.OUT;1

THE OUTCOME VARIABLE IS NARU

THE GAMMA(*)-STANDARD ERROR-T STATISTIC TABLE:

                        GAMMA(*)       STANDARD ERROR     T STATISTIC    p-VALUE
                     --------------   ---------------    ------------   ---------
FOR       BASE COEF.
          BÅSE        556.132128          1.990632          279.375       0.000
    GMDMINOR          -63.839188          6.687907           -9.545       0.000
FOR GAGE     *SLOPE
          BASE         -1.934409          0.179078          -10.802       0.000
FOR DMINOR    SLOPE
          BASE        -15.915781          3.974746           -4.004       0.000
FOR GFEI1    *SLOPE
          BASE        -17.927472          7.856794           -2.282       0.022
FOR GFEI2    *SLOPE
          BASE        -12.200569          7.333343           -1.664       0.096
FOR GFEI3    *SLOPE
          BASE         -7.612623          7.430194           -1.025       0.306
FOR GFEI4    *SLOPE
          BASE         -1.267357          7.133322           -0.178       0.859
FOR GMEI1    *SLOPE
          BASE        -14.603661          4.808601           -3.037       0.003
FOR GMEI2    *SLOPE
          BASE         -0.976529          3.260502           -0.300       0.764
FOR GMEI3    *SLOPE
          BASE         -0.874805          3.231145           -0.271       0.787
FOR GFWLTH1 *SLOPE
          BASE        -14.078470          3.405608           -4.134       0.000
FOR GFWLTH2 *SLOPE
          BASE         -5.136486          3.188910           -1.611       0.107
FOR GFWLTH3 *SLOPE
          BASE         -2 734418          3.193162           -0.856       0.392
FOR GF2COMP1*SLOPE
          BASE        -14.825203          3.548685           -4.178       0.000
FOR GF2COMP2*SLOPE
          BASE         -6.986993          3.595966           -1.943       0.052
FOR GF2COMP3*SLOPE
          BASE        -22.591328          3.372631           -6.698       0.000
FOR GXTND    *SLOPE
          BASE        -19.608713          2.462039           -7.964       0.000
FOR GDLANG   *SLOPE
          BASE         -8.849293          2.796374           -3.165       0.002
FOR GSEX     *SLOPE
          BASE         14.648574          2.258468            6.486       0.000

* - THE RESIDUAL VARIANCE FOR THIS PARAMETER HAS BEEN SET TO ZERO.

THE PRECEEDING GAMMA(*) TABLE REFLECTS THE SPECIFIED WEIGHTING


THIS ANALYSIS WAS WEIGHTED USING CLS_WGT
```

```
_$1$DIA1:[IRLSIMP.HLM.OUT4]MODEL1G.OUT;1

          RESULTS OF GENERAL LINEAR HYPOTHESIS TESTING
CONTRASTS SPECIFIED:
                   , GAMMAS            CONTRAST COEFFICIENTS
                  -------------      -------------------------
FOR      BASE
         BASE      556.132128         0.000
     GMDMINOR      -63.839188         0.000
FOR GAGE
         BASE       -1.934409         0.000
FOR DMINOR
         BASE      -15.915781         0.000
FOR GFEI1
         BASE      -17.927472         0.000
FOR GFEI2
         BASE      -12.200569         0.000
FOR GFEI3
         BASE       -7.612623         0.000
FOR GFEI4
         BASE       -1.267357         0.000
FOR GMEI1
         BASE      -14.603661         0.908
FOR GMEI2
         BASE       -0.976529        -0.261
FOR GMEI3
         BASE       -0.874805        -0.204
FOR GFWLTH1
         BASE      -14.078470         0.000
FOR GFWLTH2
         BASE       -5.136486         0.000
FOR GFWLTH3
         BASE       -2.734418         0.000
FOR GF2COMP1
         BASE      -14.825203         0.000
FOR GF2COMP2
         BASE       -6.986993         0.000
FOR GF2COMP3
         BASE      -22.591328         0.000
FOR GXTND
         BASE      -19.608713         0.000
FOR GDLANG
         BASE       -8.849293         0.000
FOR GSEX
         BASE       14.648574         0.000

CHI-SQUARE STATISTIC                =        9.988737

NUMBER OF DEGREES OF FREEDOM =        1

P-VALUE
                                    =        0.001992
```

```
_$1$DIA1: [IRLSIMP. HLM. OUT4]MODEL1G. OUT; 1

          RESULTS OF GENERAL LINEAR HYPOTHESIS TESTING
CONTRASTS SPECIFIED:
                        GAMMAS           CONTRAST COEFFICIENTS
                    -------------      --------------------------

FOR       BASE
          BASE        556. 132128       0. 000
    GMDMINOR          -63. 839188       0. 000
FOR GAGE
          BASE         -1. 934409       0. 000
FOR DMINOR
          BASE        -15. 915781       0. 000
FOR GFEI1
          BASE        -17. 927472       0. 000
FOR GFEI2
          BASE        -12. 200569       0. 000
FOR GFEI3
          BASE         -7. 612623       0. 000
FOR GFEI4
          BASE         -1. 267357       0. 000
FOR GMEI1
          BASE        -14. 603661      -0. 092
FOR GMEI2
          BASE         -0. 976529       0. 739
FOR GMEI3
          BASE         -0. 874805      -0. 204
FOR GFWLTH1
          BASE        -14. 078470       0. 000
FOR GFWLTH2
          BASE         -5. 136486       0. 000
FOR GFWLTH3
          BASE         -2. 734418       0. 000
FOR GF2COMP1
          BASE        -14. 825203       0. 000
FOR GF2COMP2
          BASE         -6. 986993       0. 000
FOR GF2COMP3
          BASE        -22. 591328       0. 000
FOR GXTND
          BASE        -19. 608713       0. 000
FOR GDLANG
          BASE         -8. 849293       0. 000
FOR GSEX
          BASE         14. 648574       0. 000


CHI-SQUARE STATISTIC              =        0. 140361

NUMBER OF DEGREES OF FREEDOM =    1

P-VALUE                           = <. 500




********    END OF OUTPUT    ********
```

```
_$1$DIA1:[IRLSIMP.HLM.OUT4]MODEL1H.OUT;1

         RESULTS OF GENERAL LINEAR HYPOTHESIS TESTING
CONTRASTS SPECIFIED:
                         GAMMAS        CONTRAST COEFFICIENTS
                      ------------   ------------------------
FOR     BASE
        BASE          556.132128       0.000
    GMDMINOR          -63.839188       0.000
FOR GAGE
        BASE           -1.934409       0.000
FOR DMINOR
        BASE          -15.915781       0.000
FOR GFEI1
        BASE          -17.927472       0.000
FOR GFEI2
        BASE          -12.200569       0.000
FOR GFEI3
        BASE           -7.612623       0.000
FOR GFEI4
        BASE           -1.267357       0.000
FOR GMEI1
        BASE          -14.603661      -0.092
FOR GMEI2
        BASE           -0.976529      -0.261
FOR GMEI3
        BASE           -0.874805       0.796
FOR GFWLTH1
        BASE          -14.078470       0.000
FOR GFWLTH2
        BASE           -5.136486       0.000
FOR GFWLTH3
        BASE           -2.734418       0.000
FOR GF2COMP1
        BASE          -14.825203       0.000
FOR GF2COMP2
        BASE           -6.986993       0.000
FOR GF2COMP3
        BASE          -22.591328       0.000
FOR GXTND
        BASE          -19.608713       0.000
FOR GDLANG
        BASE           -8.849293       0.000
FOR GSEX
        BASE           14.648574       0.000


CHI-SQUARE STATISTIC          =       0.150016

NUMBER OF DEGREES OF FREEDOM =    1

P-VALUE                       = <.500
```

```
_$1$DIA1:[IRLSIMP.HLM.OUT4]MODEL1H.OUT;1

          RESULTS OF GENERAL LINEAR HYPOTHESIS TESTING
CONTRASTS SPECIFIED:
                           GAMMAS          CONTRAST COEFFICIENTS
                        -------------    -------------------------
FOR      BASE
         BASE            556.132128        0.000
    GMDMINOR             -63.839188        0.000
FOR GAGE
         BASE             -1.934409        0.000
FOR DMINOR
         BASE            -15.915781        0.000
FOR GFEI1
         BASE            -17.927472        0.000
FOR GFEI2
         BASE            -12.200569        0.000
FOR GFEI3
         BASE             -7.612623        0.000
FOR GFEI4
         BASE             -1.267357        0.000
FOR GMEI1
         BASE            -14.603661       -0.092
FOR GMEI2
         BASE             -0.976529       -0.261
FOR GMEI3
         BASE             -0.874805       -0.204
FOR GFWLTH1
         BASE            -14.078470        0.000
FOR GFWLTH2
         BASE             -5.136486        0.000
FOR GFWLTH3
         BASE             -2.734418        0.000
FOR GF2COMP1
         BASE            -14.825203        0.000
FOR GF2COMP2
         BASE             -6.986993        0.000
FOR GF2COMP3
         BASE            -22.591328        0.000
FOR GXTND
         BASE            -19.608713        0.000
FOR GDLANG
         BASE             -8.849293        0.000
FOR GSEX
         BASE             14.648574        0.000


CHI-SQUARE STATISTIC               =        1.415427

NUMBER OF DEGREES OF FREEDOM =     1

P-VALUE                            =        0.232102
```

4TH GRADE NARRATIVE: OBSERVED AND ADJUSTED MEANS

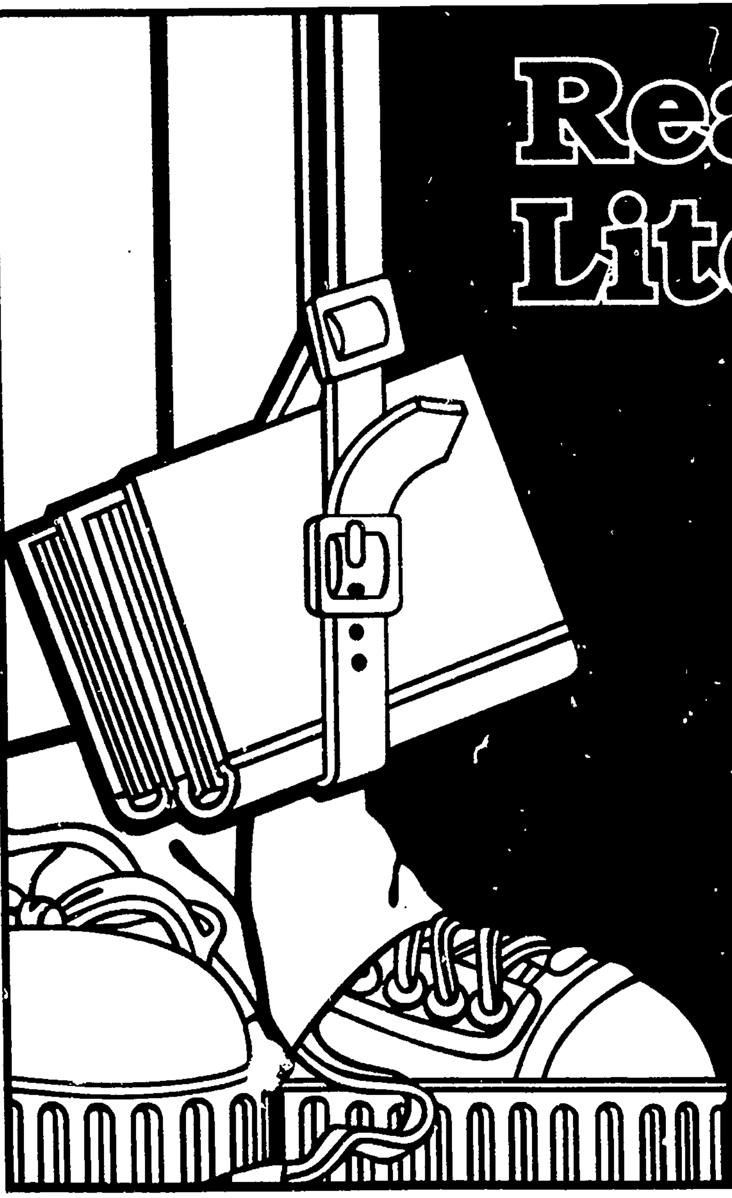
	observed deviation from grand mean	adjusted deviation from grand mean	standard error of adjusted deviation	statistically significant difference	observed mean	adjusted mean
MINORITY STATUS						
white	15	4.05	1.1	black	670	559.85
black	-50	-18.77	3.7		505	536.23
Hispanic	-27	-7.04	4.3		528	547.90
Indian	-7	0.30	6.3		548	554.70
Asian	12	10.57	6.1		567	565.57
FATHER'S EDUCATION						
none	-12	8.28	6.5		543	563.28
less than HS	35	-14.39	3.6	college	520	540.61
HS	-9	-3.75	2.3		546	551.26
HS+	2	-1.57	2.5		557	663.43
College	12	4.74	1.4		567	559.74
FAMILY WEALTH						
poor	-27	-8.74	2.1	Q3, rich	528	546.26
Q2	-1	-1.63	1.9		554	553.37
Q3	14	3.85	1.9		569	558.85
rich	15	6.20	2		570	561.20
FAMILY COMPOSITION						
2 parent	14	5.91	1	other	569	560.91
Mo only	1	9.14	3.4	Step, other	556	564.14
Step	8	-4.48	3.5	other	547	550.53
other	-34	-18.01	2.8		521	536.99
INSTRUCTIONAL HOURS						
<=25	-3	-11.09	3.2	<=30, >30	554	545.91
<=30	3	3.47	1.6		560	560.47
>30	16	9.41	5.4		573	566.41
CLASS SIZE						
<=15	-2	-1.69	6.6		555	555.31
<=20	21	7.56	3.2	<=30	578	564.56
<=25	-3	-0.05	2.2		554	556.95
<=30	-14	-5.68	2.8		543	551.32
30+	6	1.32	5		563	558.32
REGION						
N.E.	8	14.82	4.1	S.E., Central	565	571.82
S.E.	25	-19.16	3.7	Central, West	532	537.84
Central	14	-1.57	3		571	555.43
West	2	5.66	3.1		559	562.66
PARENT COOPERATION						
<average	-43	-26.41	5.2	av,>av,>>av	514	530.59
average	-5	-5.42	2.7	>>av	552	551.58
>average	9	5.70	2.8		566	562.70
>>average	31	17.13	2.6		588	574.13

Attachment A-1

Reading Literacy Tests

IEA

Reading Literacy



Fourth Grade

Developed by: International Assocation for the Evaluation of Educational Achievement
(IEA)
1990-91

DIRECTIONS

This reading test will be given in two sessions. Each session includes several kinds of questions. During each session you will be given practice questions so that you will understand exactly what you are expected to do.

There are a few directions that you should follow throughout the test:

- Read the directions carefully.
- Try to answer each question.
- Choose **only one** answer for each question.
- Always choose the **BEST** answer. Even if you are not sure of the answer to a question, choose what you think is the **BEST** answer and go on to the next question.
- If you change an answer, be sure to erase the first answer completely.
- **Look only at your own work.**

SESSION 1

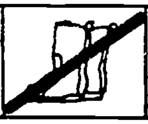
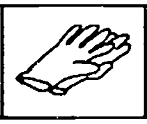
Part I

For each of the questions in this part of the test, you will see a word and four pictures.

- First, read the word.
- Then, look at all four pictures.
- Draw a line through the picture that matches the word. Draw a line through **only** one picture for each question.
- If you want to change an answer, be sure to erase the first line completely.

Now let's try two practice questions. Look at Practice Question 1. The word in Practice Question 1 is "boots." Look at the four pictures beside the word "boots." You will see that a line has been drawn through the picture of a pair of boots. This is how you should answer these questions.

Practice Question 1

boots				

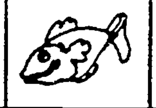
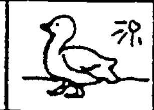
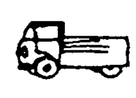
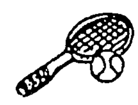
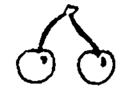
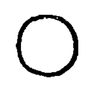
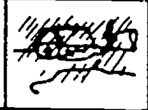
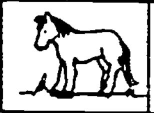
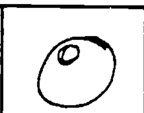
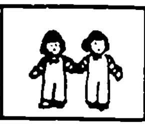
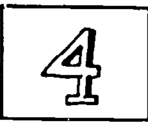
Now, look at Practice Question 2. The word in this question is "whale." Draw a line on the picture that shows a whale.

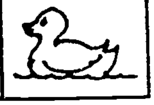
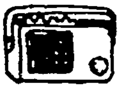
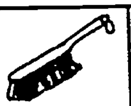
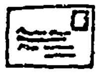
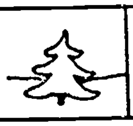
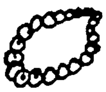
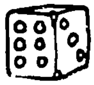
Practice Question 2

whale				

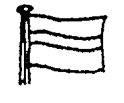
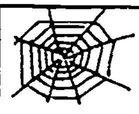
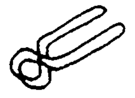
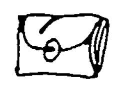
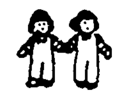
You will have only one and a half minutes to do this part of the test. So **work quickly.** Answer as many questions as you can. Look only at your own work. Your teacher will tell you when to start. Then keep working until you come to the page that says "STOP! PUT DOWN YOUR PENCIL," or until your teacher tells you to stop.

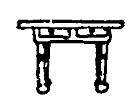
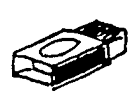
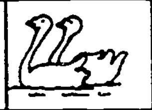
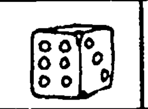
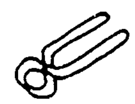
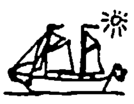
DO NOT TURN THIS PAGE
UNTIL YOU ARE TOLD TO DO SO

1.	bird				
2.	head				
3.	child				
4.	shoe				
5.	foot				
6.	coat				
7.	circle				
8.	boy				
9.	snake				
10.	fingers				

11.	flag				
12.	duck				
13.	bell				
14.	arrow				
15.	mouse				
16.	brush				
17.	orange				
18.	star				
19.	hammer				
20.	key				

GO ON TO THE NEXT PAGE

21.	lamp				
22.	woman				
23.	fish				
24.	horse				
25.	letter				
26.	drum				
27.	twins				
28.	comb				
29.	man				
30.	five				

31.	knife				
32.	hands				
33.	flower				
34.	heart				
35.	bed				
36.	cup				
37.	plant				
38.	cap				
39.	butterfly				
40.	boat				

STOP!
PUT DOWN YOUR PENCIL

Part II

In this part you will have several passages to read. Some are short; others are long. Some have pictures and diagrams; others have only words. Some are easy; others are hard.

Let's try this practice exercise so you will understand what you are supposed to do.

First, read this passage about "The Birthday Present."

Then, answer the four practice questions that follow it.

The Birthday Present

Dan was given a new bike for his birthday. He likes to ride it with his head in the air. Sometimes he takes his hands off the handlebars. His parents told him to be careful, but he did not listen. One day the front wheel of Dan's bike hit a large stone which he had not seen. Crash. The bike stopped quickly, but Dan kept going. He finished up in a ditch full of water.

Practice Question 1. What did Dan get for his birthday?

A. A stone
[✓ B.] A bike
C. A wheel
D. A car

Answer B, "A bike," is the right answer. You will see that a circle has been drawn around the letter "B." In each multiple choice question, you should draw a circle around the letter in front of the right answer.

Here is another practice question.

Practice Question 2. How does Dan ride his bike?

A. He watches the road carefully.
B. He looks at the handlebars.
C. He does not ride carefully.
D. He rides the way his parents told him.

Which is the correct answer? The **BEST** answer is "He does not ride carefully." Did you draw a circle around the letter "C"?

Now, try Practice Question 3.

Practice Question 3. How do you think Dan would feel at the end of this story?

A. Proud of his bike.
B. Wet and unhappy.
C. Pleased with himself.
D. Eager to do it again.

Which is the right answer? The **BEST** answer is "Wet and unhappy." Did you draw a circle around the letter "B"?

Now, try Practice Question 4. This time you should write your answer on a line under the question.

Practice Question 4. What did Dan's bike hit?

____

____

What is the right answer? You should have written a sentence that said something like "Dan's bike hit a large stone."

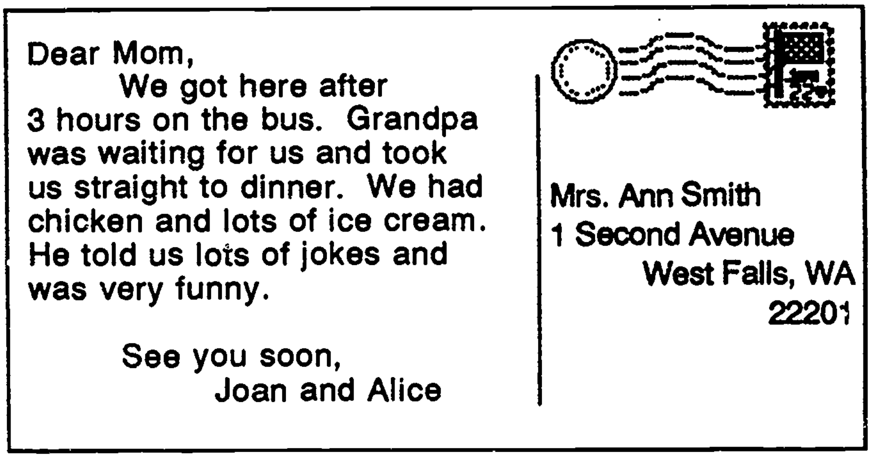
In the next few pages, you will read more passages and questions like these. **Remember:**

- Work as quickly as you can. When you finish each page, move on to the next until you come to the page that says "STOP! PUT DOWN YOUR PENCIL," or until your teacher tells you to stop.
- Try to answer all questions, but do not spend too much time on questions you cannot do. You will have **35 minutes** to finish the questions. If a question is too difficult for you to answer, choose what you think is the **BEST** answer and move on to the next question.
- Wait until your teacher tells you to start.

**DO NOT TURN THIS PAGE
UNTIL YOU ARE TOLD TO DO SO**

POSTCARD

Read this postcard that Joan and Alice wrote to their mother. Then answer questions 41-42.



Dear Mom,

We got here after 3 hours on the bus. Grandpa was waiting for us and took us straight to dinner. We had chicken and lots of ice cream. He told us lots of jokes and was very funny.

See you soon,
Joan and Alice

Mrs. Ann Smith
1 Second Avenue
West Falls, WA
22201

41. Who met the children when they arrived?

A. Their grandfather
B. Their parents
C. Their friend
D. Their mother

42. How did the children travel?

A. By bus
B. By car
C. By train
D. By airplane

GO ON TO THE NEXT PAGE

Read the story and answer questions 43-47.

THE BIRD AND THE ELEPHANT

A large tree grew in the middle of the jungle. At the top, a small bird had made a nest for her family of three baby birds. One day, an elephant came by. He leaned against the trunk, and scratched his back. The tree started to crack and sway. The baby birds, full of fear, huddled against their mother. She stuck the tip of her beak out of the nest, and said, "Hey, big animal, there are many trees around here! Why shake this one? My children are afraid, and could fall out of their nest."

The elephant said nothing, but he looked at the bird with his small eye, flapped his large ears in the wind, and left.

The next day, the elephant returned and scratched against the trunk once more. The tree began to sway. The frightened baby birds once again huddled against their mother's wings. Now Mother Bird was angry. "I order you to stop shaking our tree," she cried, "or I will teach you a lesson!"

"What could you do to a giant like me?" laughed the elephant. "If I wanted to, I could give this tree such a push that your nest and your children would be flung far and wide."

The mother bird said nothing.

The next day, the elephant returned and scratched again. Quick as a flash, the mother bird flew into one of the elephant's enormous ears, and there, tickled the elephant by scratching him with her feet. The elephant shook his head ... nothing happened. So he begged the bird to leave and promised to stop scratching against the trunk.

The bird then left the elephant's ear and returned to her nest, beside her children.

The elephant never again returned to scratch his back.

43. What does the passage tell us?

A. When you're strong, you can bother others.
B. Elephants shouldn't shake trees.
C. The weak can sometimes overcome the strong.
D. Always face danger head-on.

44. Which of these things happened <u>first</u>? The mother bird

A. warned her children.
B. threatened the elephant.
C. told him to scratch somewhere else.
D. won the contest.

45. Which sentence in the story tells us that the elephant thinks he is the strongest? It starts with these underlined words:

A. "Hey, big animal,..."
B. The elephant said nothing,...
C. The next day, the elephant returned...
D. "What could you do..."

46. What did the mother bird do to stop the elephant from returning to that tree?

A. She ordered him to stop.
B. She scratched his back.
C. She tickled his ear.
D. She stuck her beak into him.

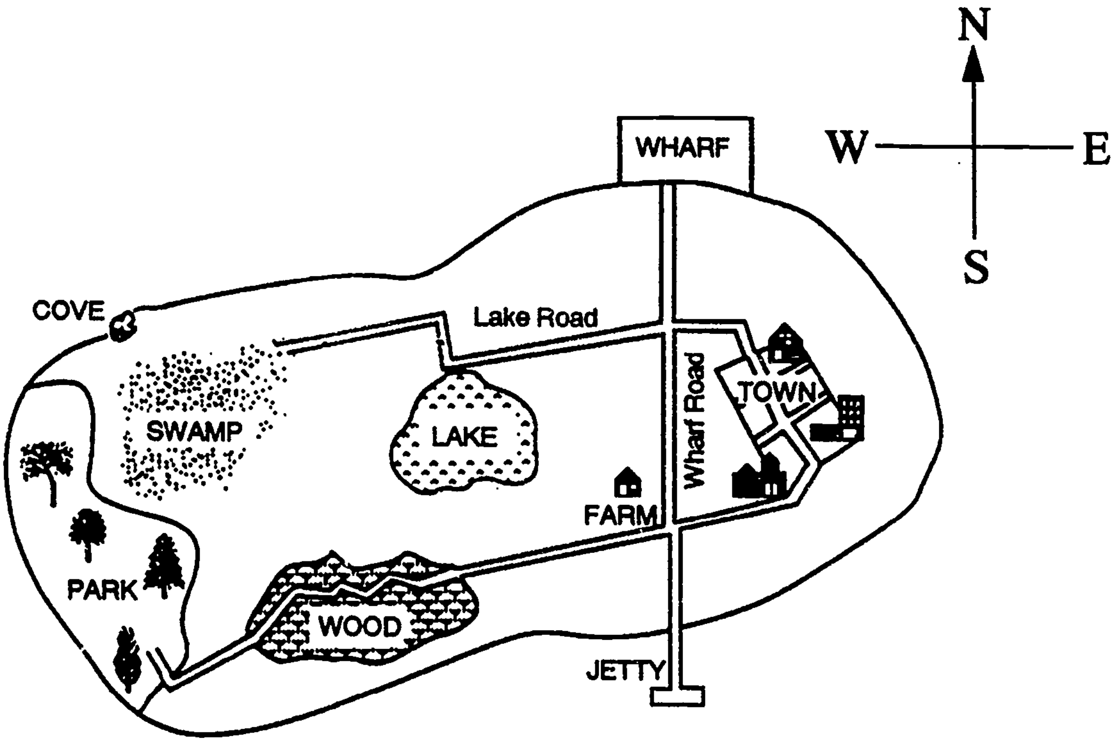
47. The story ends happily because

A. the elephant died.
B. the elephant did not come back.
C. the tree was strong enough.
D. the birds learned to fly.

GO ON TO THE NEXT PAGE

ISLAND

Use this map of ONE-LAKE ISLAND to answer questions 48-51.



48. Which of these is south of the farm?

A. Jetty
B. Park
C. Town
D. Wood

49. Which of these is <u>closest</u> to the swamp?

A. Town
B. Jetty
C. Lake
D. Farm

50. Follow these instructions:

1 —Start at the farm.
2 —Go north to first cross-roads.
3 —Turn left.
4 —Go to the first corner.

Now, where are you?

A. Town
B. Lake
C. Wharf
D. Swamp

51. If you want to go from the town to the park, you will go

A. north.
B. south.
C. east.
D. west.

GO ON TO THE NEXT PAGE

MARIA'S TIMETABLE

Here is Maria's school timetable. Use it to answer questions 52-54.

Lesson	Monday	Tuesday	Wednesday	Thursday	Friday
1st	English	History	Science	English	History
2nd	English	Math	English	History	English
3rd	Science	English	English	Math	Math
4th	Math	The world about us	Math	The world about us	Science
5th	Music	Art	The world about us	Music	Physical education
6th	-	Physical education	Physical education	-	Art

52. What is the third lesson for Maria on Thursdays?

A. Math
B. Science
C. English
D. History

53. On how many days does she have six lessons?

A. 2
B. 3
C. 4
D. 5

54. Which subject is Maria taught most often?

A. Math
B. Science
C. The world about us
D. English

GO ON TO THE NEXT PAGE

Read the passage and answer questions 55-61.

THE WALRUS

The walrus is easy to recognize because it has two large teeth sticking out of its mouth. These teeth are called eye teeth.

The walrus lives in cold seas. If the water freezes over, the walrus keeps a hole free of ice either by swimming round and round in the water, or by hacking off the edge of the ice with its eye teeth. The walrus can also use its skull to knock a hole in the ice.

The walrus depends on its eye teeth for many things. For example, when looking for food a walrus dives to the bottom of the sea and uses its eye teeth to scrape off clams. The walrus also uses its eye teeth to pull itself on the ice. It needs its eye teeth to attack or kill a seal and eat it, or to defend itself if attacked by a polar bear.

The walrus may grow very big and very old. A full-grown male is almost 13 feet long and weighs more than 2200 pounds. It may reach an age of 30 years.

The walrus sleeps on the ice or on a piece of rock sticking out of the water, but it is also able to sleep in the water.

55. Where does the walrus live?

A. In very cold water
B. In tropical countries
C. On the bottom of lakes
D. In cold forest country

56. How long can a walrus live?

A. 2 years
B. 4 years
C. 30 years
D. 100 years

57. What does a walrus eat?

A. Ice
B. Seals
C. Seaweed
D. Birds

58. We can tell that the walrus has to protect itself from

A. seals.
B. bears.
C. eagles.
D. lions.

59. What does a walrus do when it wants to get up on the ice?

A. It jumps up.
B. It cries for help.
C. It uses its eye teeth.
D. It uses its skull.

GO ON TO THE NEXT PAGE

60. How does the walrus get its food?

A. It catches fish with its eye teeth.
B. It scrapes clams off the bottom of the sea.
C. It knocks a hole in the ice with its skull.
D. It attacks polar bears.

61. What problems would the walrus have if it lost its eye teeth?

Write your answer on the lines below. Make sure you write enough to make your answer clear. You may want to use examples from the story to help explain your answer.

Do not write in this area.

____ 1 ____ 2 ____ 3

GO ON TO THE NEXT PAGE

Read the story and answer questions 62-67.

No Dogs Is Not Enough

By Linda Leopold Strauss

Reprinted by permission
of the author.

Tony wanted a dog.
Tony's mother said no.
Mrs. Lawlor who lived up the street promised Tony one of Snuffy's puppies - if Tony's mother said yes.

Tony's mother said no.
"I have enough to do already," she told Tony, "without taking care of a dog."

The Lawlors were going to put an ad in the paper to sell Snuffy's puppies, even the brown puppy with the white nose. Tony had to find a way to change his mother's mind.

The next morning, after he got dressed, Tony found a rope. He made a loop at one end and a smaller loop at the other end that he held in his hand. Then he ran downstairs, trailing the rope behind him.

"I'm going to walk Nosey before breakfast," Tony told his mother.

His mother stared at him.

"Nosey. My dog," said Tony. "Isn't she cute?" And he went out the back door, pulling the rope behind him.

"Now Tony...," began his mother, but Tony was gone. When he came back into the kitchen, he looped the rope over the back of his chair.

"Down, Nosey," he said. "Down, girl. Sit."

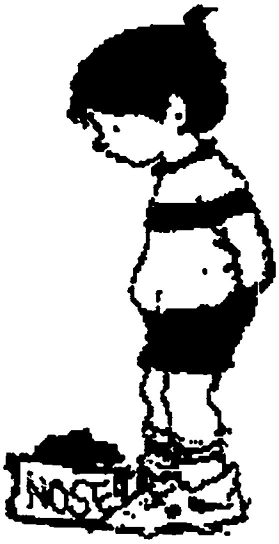
Tony's father leaned over to look at the floor next to Tony's chair. He looked at Tony. "Are you feeling all right?" he asked.

"Fine," said Tony. "Nosey's fine, too. She's a good dog, isn't she, Dad?"

"You'd never even know she's there," said Tony's father.

Tony was very busy after school. He took the money he had been saving from his allowance and walked Nosey to the corner store. He bought dog food and a red dish to put it in, a leash, and a real leather collar. When he got home, he put the dog food and a bowl of water in the kitchen.

GO ON TO THE NEXT PAGE



"Mom?" said Tony. "Nosey's such a good dog. Can she sleep in my room tonight?"

"Absolutely not," said Tony's mother firmly. "No dogs in the bedroom." Then she laughed out loud. "You and your imagination!"

After a few days, the neighbors got used to seeing Tony walk around the block with a leash and a real leather dog collar. Tony walked Nosey twice a day, rain or shine, and every morning he put fresh food in her new red bowl. Tony's mother had to admit that Tony took good care of Nosey.

"If I can take care of Nosey," Tony told his mother, "I can take care of any puppy."

"Perhaps," said his mother, "but we have Nosey now. One dog is enough."

Early the next morning, Tony came down to the kitchen. "Have you seen Nosey?" he asked his mother. "I can't find her anywhere." He walked over to the red dish. "Nosey hasn't touched her food," he said in a worried voice. "She must have gotten out."

"But how?" asked Tony's mother. "You had her with you at bedtime. I saw her myself." She turned to Tony's father. "Didn't you, dear?" she asked.

Tony's father looked at her and shook his head. "You and your imagination!" he laughed. "You and that dog!"

Right after breakfast, Tony went out to search for Nosey. He walked up driveways and behind garages and down the hill to the playground.

"Nosey," he called, but Nosey didn't come.

"Maybe we should put an ad in the paper," suggested Tony's mother at lunchtime.

"What would we say Nosey looked like?" Tony wanted to know.

There was no doubt about it. Nosey was going to be hard to find.

"We won't find her," said Tony, and he was right.

"I hate to admit it," said Tony's mother at dinner, "but I think I miss Nosey." "We could get another dog," said Tony quickly.

"What if Nosey comes back?" asked his father. "Your mother says one dog is enough."

"Nosey is not coming back," said Tony. "And no dogs is not enough."

"Not coming back?" said his mother. "That's a different story." She looked at Tony. "I see in the paper that Lawlors' puppies are still for sale."

"Brown ones," said Tony's father. "One with a white nose."

<u>Tony held his breath.</u>

"It's a nice night," said Tony's father. "Let's take a walk to the Lawlors'."

"And don't forget the leash and the collar," said Tony's mother.

GO ON TO THE NEXT PAGE

62. At the beginning of the story, Tony's mother thought dogs were

A. too much work.
B. too costly.
C. too noisy.
D. too dangerous.

63. His mother first found out about Nosey when Tony

A. went out to look for a lost dog.
B. walked him on a rope.
C. found an ad for a dog in the paper.
D. put fresh dog food in a bowl.

64. Why did Tony pretend to have a dog?

A. He was trying to make his mother angry.
B. It was easier than taking care of a real dog.
C. He wanted to prove that he could feed and walk a dog.
D. He thought a make-believe dog was better than no dog at all.

65. Why did Tony tell his parents that Nosey ran away?

A. He really could not find Nosey anywhere.
B. He was tired of taking care of Nosey.
C. He wanted his parents to replace Nosey with a real dog.
D. He decided that he did not want a dog after all.

66. Why did the author write, "Tony held his breath," near the end of the story?

A. To give a clue about how Tony felt at the moment
B. To tell how Tony felt about his parents
C. To show where the story took place
D. To remind you that Nosey was lost

67. What do you think would happen after Tony got his puppy from the Lawlors?

A. Tony would still try to get Nosey back.
B. The puppy would get lost just as Nosey did.
C. Tony would take good care of the puppy.
D. Tony's parents would be angry.

STOP! PUT DOWN YOUR PENCIL
END OF SESSION 1

Session 2

In this part you will have several passages to read and questions to answer. Let's do four practice questions.

Look at the table below that lists magazine prices.

MAGAZINE PRICES

Adventure Stories	$2.50
New Comics	$1.00
Today's Best Jokes	$3.00
Children's Sports	$2.50

Practice Question 1. How much does Adventure Stories cost?

A. $1.00
[✓ B.] $2.50
C. $3.00
D. $5.00

Which is the correct answer? The magazine Adventure Stories costs $2.50. This is the right answer. You will see that a circle has been drawn around the letter "B."

Now let's do Practice Question 2.

Practice Question 2. Which magazine costs the most?

A. Adventure Stories
B. Children's Sports
C. New Comics
D. Today's Best Jokes

Which is the correct answer? The correct answer is Today's Best Jokes. Did you draw a circle around the letter "D"?

Now do Practice Question 3.

Practice Question 3. Mary has only $1.00. She wants to buy a magazine. Which one can she buy?

A. Adventure Stories
B. Today's Best Jokes
C. Children's Sports
D. New Comics

Which is the correct answer? The correct answer is New Comics. Did you draw a circle around the letter "D"?

Now do Practice Question 4.

Practice Question 4. How much does Today's Best Jokes cost?

____

What is the correct answer? The correct answer is $3.00. Did you write "$3.00" on the line under Practice Question 4?

In the next few pages, you will see more questions like the practice questions you just completed. **Remember:**

- Work as quickly as you can. When you finish one page, move on to the next page. Keep working until you come to the page that says "STOP! PUT DOWN YOUR PENCIL." You will have **35 minutes** to finish all the questions. Try to answer all questions, but do not spend too much time on questions that you cannot do. If a question is too difficult for you to answer, choose what you think is the **BEST** answer and move on to the next question.
- Wait until your teacher tells you to start.

DO NOT TURN THIS PAGE
UNTIL YOU ARE TOLD TO DO SO

Read the passage and answer questions 1-3.

WHAT IS QUICKSAND?

Quicksand is a special kind of sand. Quicksand can swallow a pig, or a human, or even an elephant.

Quicksand often looks like plain wet sand. But it is really a soupy sand with so much water between the grains that you can't stand on it.

If you step into quicksand, you will slowly sink up to your knees. If you thrash and squirm, you will sink deeper and deeper. But if you lie flat on your back with your arms stretched out, you can float on the sand, as you can float in water.

Watch out for quicksand on sand bars, on the bottom of streams, or along sandy seacoasts.

You can test for quicksand by poking it with a long stick or pole. If the sand shakes and quakes, don't try to walk on it! It may be quicksand.

1. What is the main purpose of the article?

   A. To tell people how to avoid the dangers of quicksand
   B. To encourage people to protect the beauty of nature
   C. To describe how people and animals have been swallowed by quicksand
   D. To explain how quicksand got its name

2. According to the article, what should you do if you step into quicksand?

   A. Thrash your arms and try to shake yourself out
   B. Grab a stick and try to pull yourself out
   C. Stand still and yell for somebody to help you
   D. Lie on your back with your arms stretched out

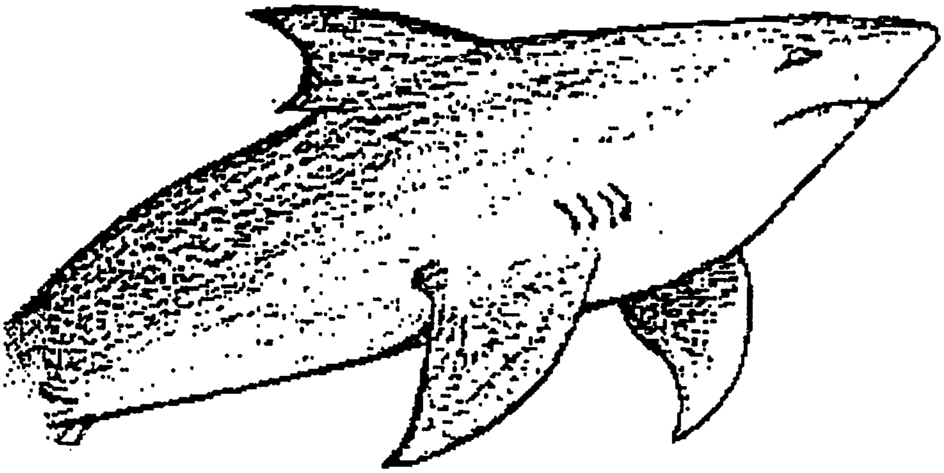
3. According to the article, how can you test to see if sand is really quicksand?

   A. Stick your hand into it.
   B. Step lightly on it.
   C. Poke it with a stick.
   D. Look at it.

GO ON TO THE NEXT PAGE

Read the story and answer questions 4-8.

A SHARK MAKES FRIENDS



a fable by GEORGE CIANTAR

The SHARK glided forward, too full to chase the few that had escaped. What a dinner that had been! Beautiful silver mackerel! Dozens and dozens of them - swallowed whole and thrashing. But now that the shark felt satisfied, his thoughts turned away from the food.

"I always swim alone," he said. "I have no friends. Nobody trusts me; nobody loves me. They all fear me, even my own kind. And yet ... and yet ... I know I could be so nice to my friends if only I had any. Oh, it's a lonely life being a shark!" And with that he began to cry, softly at first, then in loud racking howls of heartbreak.

A sardine heard his grief. From a safe distance, she stopped to watch and listen. "You sound unhappy," she said to the shark.

"I am. Oh, how I am!" he said between sobs. "I'm only a young shark, yet nobody ever wants to play with me or be my friend."

"That's because you eat other fish when they get too close," she said.

"I used to," said the shark, "but I never will again. I'm not interested in food any more. What I need is a friend."

The sardine was filled with pity for the shark and decided to trust him. They became good friends very quickly.

Later she took him to meet the family. The other sardines were terrified when they saw him, but she told them that he was her very best friend, and they, too, agreed to trust him.

They all traveled together in and out of the reefs, the sardines going about their sardine business and the shark, like a gentle giant, gliding along with them. These were happy days.

But the shark was beginning to get an uncomfortable feeling - and the longer he kept company with his new friends, the more uncomfortable the feeling became. He tried to control it but...when a shark is hungry, a shark must eat!

He stopped and let the sardines pass him. Their scales flashed as their bodies caught the light. He felt the vibrations of the water fanned by their tails.

With a powerful thrust from his tail, he swept through them from behind, then from the front, and again from the back - filling his fast-working mouth with fish. Soon he was satisfied, and far too full to chase the few who escaped. He began to think about what he had done.

"I've eaten all my friends!" he said. "Now I must swim alone again. Now I must suffer loneliness again." He sobbed loudly for many minutes.

A young pilchard heard him. He swam in a little closer to see why the shark was so unhappy.

4. Why was the shark swimming alone at the beginning of the story?

   A. He was lost in the reefs.
   B. He was too large and clumsy to swim with the other fish.
   C. Other fish didn't trust him.
   D. He didn't really like his friends.

5. The shark was unhappy when the sardine heard him because

   A. he was hungry.
   B. he was lonely.
   C. he was missing his family.
   D. he had just eaten too many mackerel.

6. The sardine became friends with the shark because in the beginning she

   A. felt sorry for him.
   B. loved him.
   C. was lonely herself.
   D. wanted to change him.

7. How did her family feel when the sardine took the shark home?

   A. Pleased
   B. Frightened
   C. Friendly
   D. Surprised

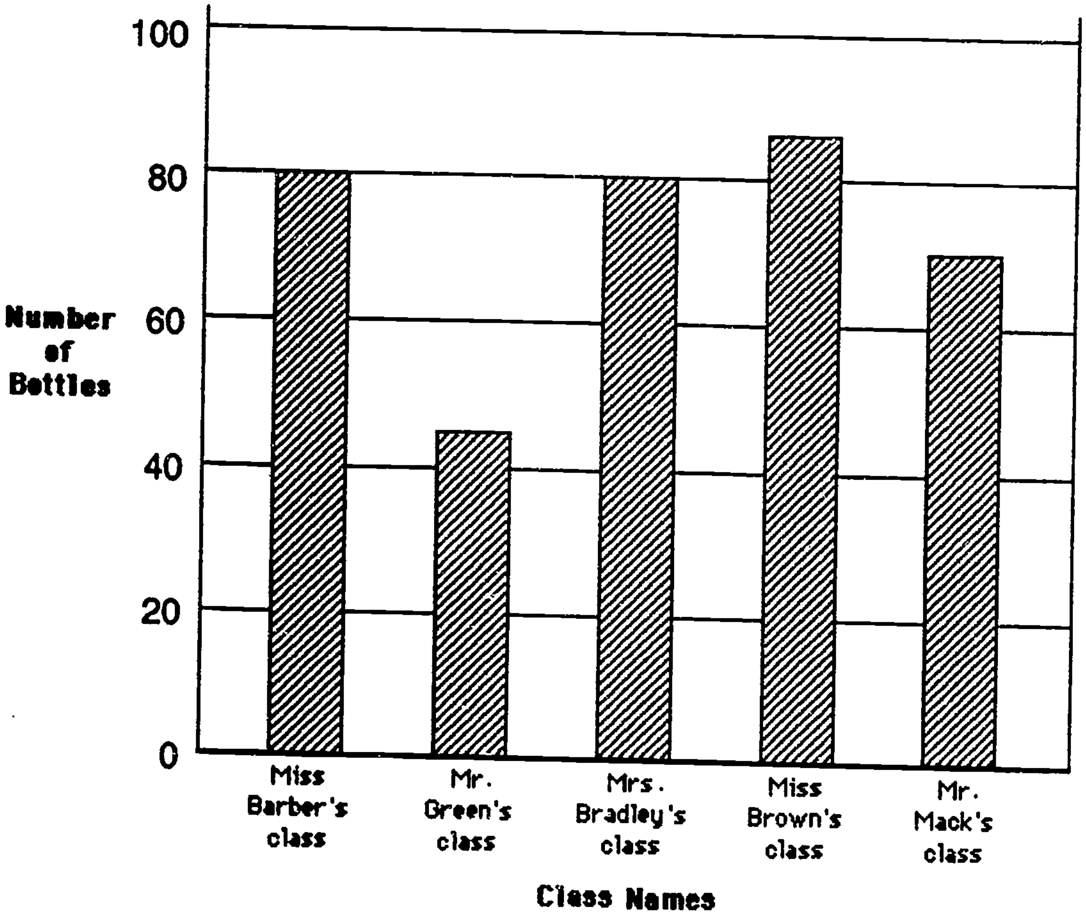
8. What was the reason for the shark's <u>uncomfortable feeling</u>?

A. He was not used to his new way of life.
B. He was getting hungry.
C. He felt crowded by the family of sardines.
D. He had eaten too much.

GO ON TO THE NEXT PAGE

EMPTY BOTTLES

Mid-Town School had a bottle collection.
Children in each class brought empty bottles to school.
The principal made a bar graph of the number of bottles from five classes. Use the graph to answer questions 9-12.



9. Which class brought 45 bottles?

A. Miss Barber's class
B. Miss Brown's class
C. Mrs. Bradley's class
D. Mr. Green's class

10. The principal asked each class to collect at least 50 bottles. How many classes have collected that many?

A. 2
B. 3
C. 4
D. 5

11. Which class got the prize for the most bottles?

A. Mr. Green's class
B. Mr. Mack's class
C. Miss Barber's class
D. Miss Brown's class

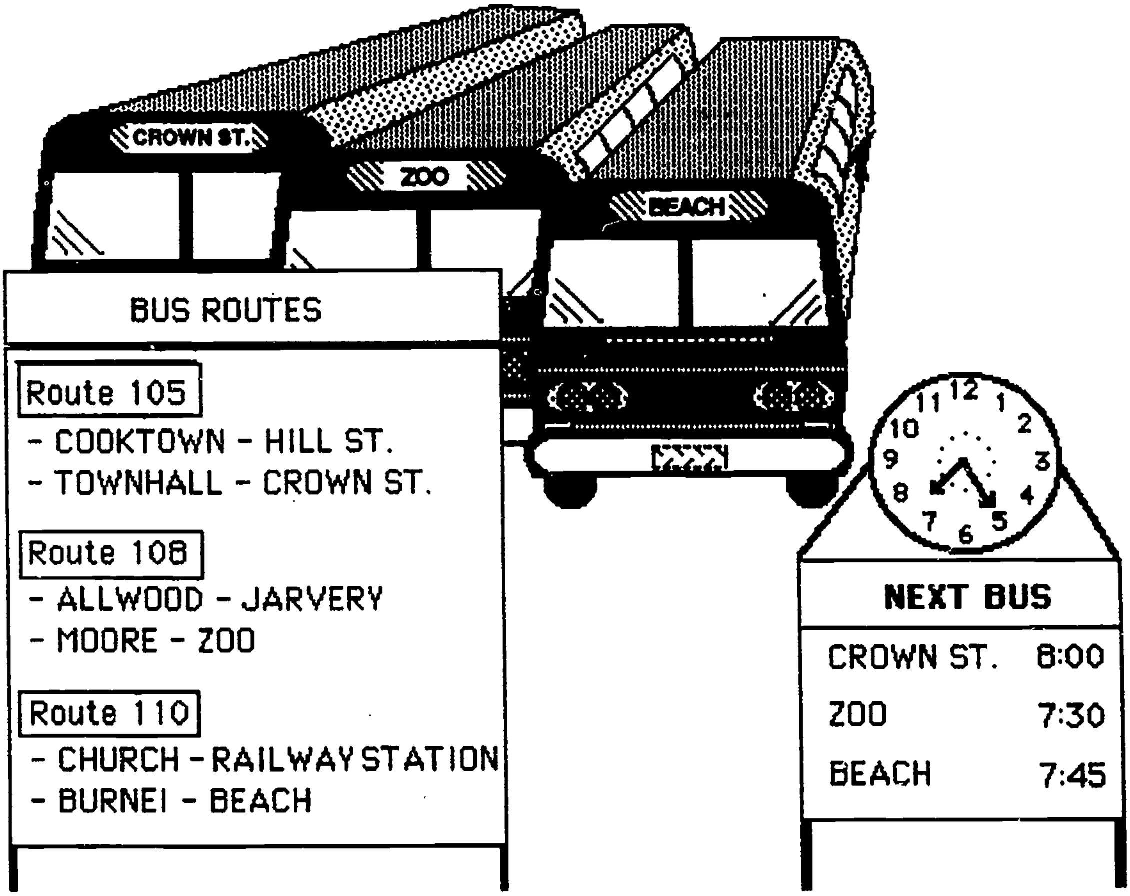
12. Which two classes collected exactly 80 bottles?

A. Miss Barber's class and Mrs. Bradley's class
B. Miss Barber's class and Mr. Mack's class
C. Miss Brown's class and Mrs. Bradley's class
D. Miss Brown's class and Mr. Mack's class

GO ON TO THE NEXT PAGE

BUSES

This sign shows the bus routes. Use it to answer questions 13-16.



BUS ROUTES

Route 105
- COOKTOWN - HILL ST.
- TOWNHALL - CROWN ST.

Route 108
- ALLWOOD - JARVERY
- MOORE - ZOO

Route 110
- CHURCH - RAILWAY STATION
- BURNEI - BEACH

NEXT BUS

CROWN ST.	8:00
ZOO	7:30
BEACH	7:45

13. Anne wants to go to the railway station. Which route number should she choose?

Route: ______________________________

14. Where do you think the bus stops first on Anne's way to the railway station?

____

15. How long will it be before the next bus leaves for the zoo?

____

16. What is the name of the place where buses stop just before the zoo?

____

GO ON TO THE NEXT PAGE

TABLE OF CONTENTS

This is a part of the table of contents in a book. Read it, and use it to answer questions 17-19.

Table of Contents

Title	Author	Pages
Let's Look at Tracks	Ann Zim	3 - 9
What's Inside of Me?	Herb Martin	10 - 15
The Rice Bowl	Les Jones	16 - 19
The Ant and the Elephant	Bill Guthrie	20 - 24
Ring of Roses	Pat Brooke	25 - 32
Making Model Airplanes	Andy Purves	33 - 35

17. On what page does Pat Brooke's story start?

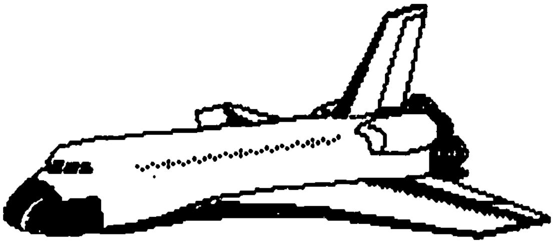
A. 3
B. 20
C. 25
D. 33

18. Who wrote the passage "The Rice Bowl"?

A. Les Jones
B. Ann Zim
C. Herb Martin
D. Bill Guthrie

19. Where in the book do you think you could find a picture like this?

A. Pages 16-19
B. Pages 20-24
C. Pages 25-32
D. Pages 33-35



TEMPERATURE

The chart below shows some temperature readings made at different times on four days. Use the chart to answer questions 20-24.

	6 a.m.	9 a.m.	12 Noon	3 p.m.	8 p.m.
Monday	15°F	17°F	20°F	21°F	19°F
Tuesday	15°F	15°F	15°F	10°F	9°F
Wednesday	8°F	10°F	14°F	13°F	15°F
Thursday	8°F	11°F	14°F	17°F	20°F

20. When was the highest temperature recorded?

A. Noon on Monday
B. 3 p.m. on Monday
C. Noon on Tuesday
D. Noon on Wednesday

21. On one day the temperature dropped quickly. When do you think this happened?

A. Monday morning
B. Tuesday afternoon
C. Wednesday afternoon
D. Thursday morning

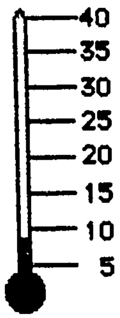
22. On how many days was it colder at 8 p.m. than at 12 noon?

A. None
B. 1
C. 2
D. 3

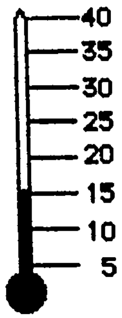
23. On which day did the temperature go on rising steadily from 6 a.m. to 8 p.m.?

A. Monday
B. Tuesday
C. Wednesday
D. Thursday

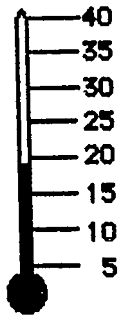
24. Which of these thermometers show the temperature at 6 a.m. on Wednesday?



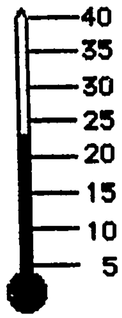
A



B



C



D

GO ON TO THE NEXT PAGE

Read the passage and answer questions 25-28.

MARMOTS

For three or four thousand years a family of marmots had been settled in a grassy little valley under the cliffs of a jagged peak whose present name is the Rock of Wonders. The maps specify that the peak is about seven thousand two hundred feet in altitude.

The spot was a remarkably suitable one for all sorts of good reasons. In the first place, the slope faced south. The sun shone on it from dawn until dark, and in the spring the snow melted there faster than anywhere else. One could warm oneself as much as one wished, or again one could sit in the shade of huge rocks fallen from the heights of the mountain. A little way off a tiny spring fed a little lake. Thanks to this fresh spring, which never dried up, the grasses round about grew thick and strong, even in the month of August.

The rocks provided many sitting places and perfectly safe holes where one could take refuge in an emergency. As for the Family Cave, hollowed out many years ago by the grandfather of the present inhabitants, it opened out pleasantly from under a flat slab between two clumps of arnica. Since it had already been improved by two generations, it would have been difficult to find a drier and more comfortable apartment.

25. Why had the marmots lived so long in one place?

A. They did not like to travel.
B. They could not climb down the cliffs.
C. They came there long ago.
D. It was a very good place.

26. What kept the grasses in the valley green and healthy all summer?

A. The warm sun
B. The melting snow
C. The high peak
D. A spring

27. Why was it a good thing that the valley faced to the south?

A. The summer days were warmer.
B. Winter snows melted early.
C. The best view was to the south.
D. Marmots need a lot of sun.

28. The last sentence of the story makes us think that the marmots will

A. make many new homes.
B. continue to live in the Family Cave.
C. have many babies.
D. work hard to make the cave comfortable.

GO ON TO THE NEXT PAGE

Read the passage and answer questions 29-34.

HOW TO READ THE AGE OF A TREE

If you can find a tree which has been cut down, you will see many rings on the base of the trunk. By learning to read these rings, you can find out about the tree's life.

The number of rings tells you how old the tree is. Each year, new wood is formed on the outside of the tree. This new wood is light in color when the tree is growing in the spring and summer, and dark in winter when the tree is not growing much. So, if you count the rings of dark-or-light-colored wood, you can often find out how old the tree is.

You can also tell which years have been good years and which years have been bad years. When the light-colored rings are very wide, it means that the tree has been growing quickly that year. If the light rings are narrow, it has been growing slowly.

If the rings on a tree trunk were greatly magnified, you would be able to see why the rings are light-colored when the tree is growing quickly, and dark-colored when the tree is growing slowly. The tree trunk is made up of microscopic tubes, like long pipes, carrying water and minerals from the soil, through the trunk, and up to the leaves. They are wide and thin-walled when the tree is growing quickly and they are carrying a lot of water. They are narrow and bunched together when the tree is not growing so quickly.

When a tree is old, the tubes in the center of the tree don't carry water. The walls of the tubes have become thick with materials which have stuck along them over the years forming a special kind of wood called "heartwood." This kind of wood is darker in color than the young, growing wood on the outside of the tree.

You don't very often see whole tree trunks which have been cut across. But once you learn to read a cross section of the wood, you can see much more in wood which has been used to make boxes, furniture, houses, and other things.

In most wood, instead of seeing the trunk cut across, you are seeing it cut along its length. Because you don't see the cross section, you can't tell how old it was.

29. The writer says you can tell the age of a tree by

A. the number of rings in its trunk.
B. the size of the base of its trunk.
C. its height.
D. the rings on its outside bark.

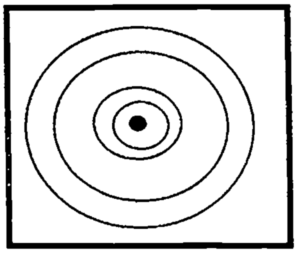
30. When the wood of a tree is mostly light in color, this means that the tree

A. grew quickly.
B. grew slowly.
C. only grew in winter.
D. only grew in summer.

31. Heartwood is wood which is

A. older and darker.
B. fast-growing.
C. younger and lighter.
D. slow -growing.

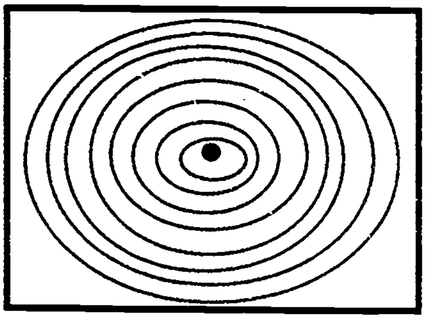
32. In the cross section of the tree trunk shown in Box 1, all the rings are wide and about the same width. This shows that the tree


BOX 1

A. grew quickly all its life.
B. grew slowly all its life.
C. grew quickly when it was young and more slowly later.
D. grew slowly when it was young and more quickly later.

GO ON TO THE NEXT PAGE

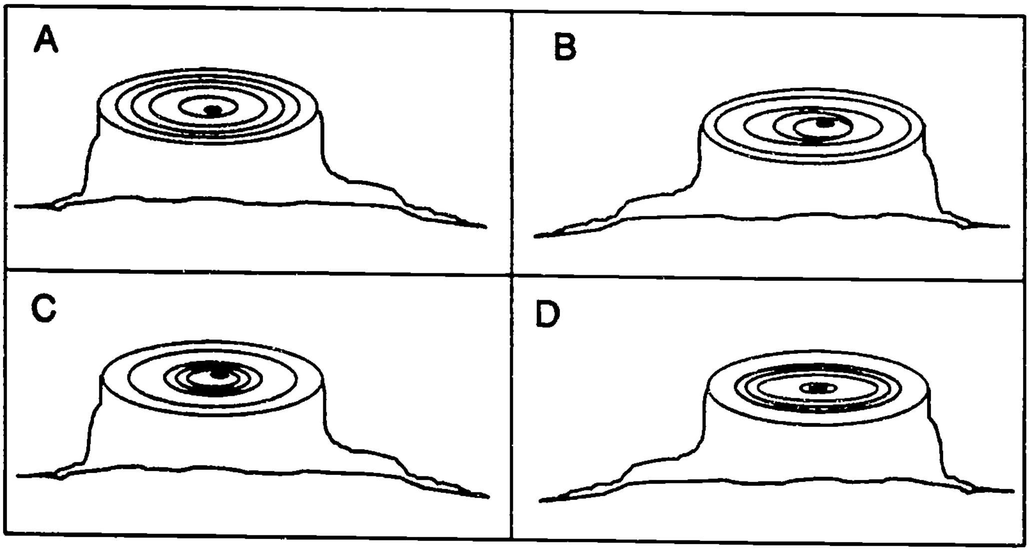
33. How many years old is the tree shown in Box 2?



BOX 2

A. Less than 6
B. 9
C. 12
D. More than 12

34. In a country which has a dry climate, it rains heavily every third year. Which drawing shows a tree trunk from this country?



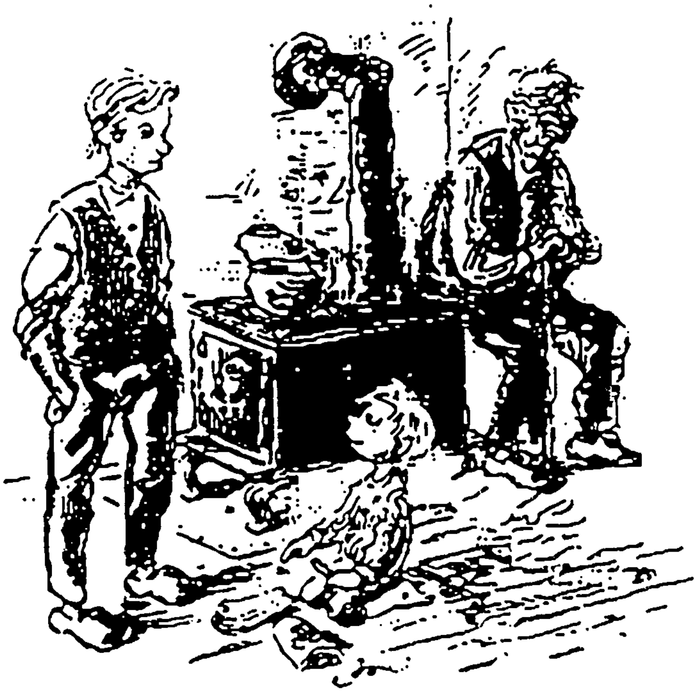
Read the story and answer questions 35-41.

GRANDPA

Once upon a time, there was a very old man. His eyes had become weak. His ears were deaf, and his knees would shake. When he sat at the table, he was hardly able to hold the spoon. He spilled soup on the tablecloth, and he often slobbered.

He lived with his son and daughter-in-law. They also had a small boy who was four years old, so the old man was a grandfather.

His son and his son's wife found it disgusting to see him spilling food at the table. And so they finally ordered him to sit in a corner behind the stove. Here, they served him his food on a small earthenware plate. Now, Grandpa didn't even get enough to satisfy his hunger. He sat there feeling sad. He looked at the table, where the others were eating, and his eyes filled with tears.



GO ON TO THE NEXT PAGE

Then, one day his shaking hands could not even hold the plate. It fell to the floor and was broken into many pieces. The young wife scolded him. But the old grandfather said nothing. He just sighed. Then the young wife bought him a very cheap wooden bowl. Now he had to eat from that.

One day, while they were having dinner, the grandchild sat on the floor, and was very busy with some small pieces of wood.

"What are you doing?" asked his father.

"I am making a bowl," the boy answered.

"What is it for?"

"It is for my father and mother to eat from when I grow up."

The man and his wife looked at each other for a long time. Then, they started crying. At once, they asked the old grandpa back to the table, and from then on he always ate with them. After that, even if he sometimes spilled his food, they never said a word about it.

35. What happened when Grandpa sat at the table?

A. He always had a good meal.
B. His feet would shake.
C. He spilled his soup.
D. He dropped his plate.

36. The son and his wife asked Grandpa to sit behind the stove because

A. it was warmer there.
B. the table was not big enough for everyone.
C. he could not see or hear.
D. they did not like to see him eat.

37. Why did the son's wife scold Grandpa?

A. He spilled his soup.
B. He broke his plate.
C. He looked so sad.
D. He showed bad manners.

38. Grandpa was given a new bowl made of wood because

A. he had wanted such a bowl.
B. the family had no more earthenware plates.
C. a wooden bowl does not break so easily.
D. the boy had made one for him.

39. How did Grandpa feel when he sat by the stove?

A. Bored
B. Tired
C. Pleased
D. Unhappy

40. The son and his wife cried because

A. the boy wanted to make a wooden bowl.
B. their old father could not eat properly.
C. they understood that they too would grow old.
D. the wooden bowl was also broken.

GO ON TO THE NEXT PAGE

41. Why did the parents decide to ask Grandpa back to the table?

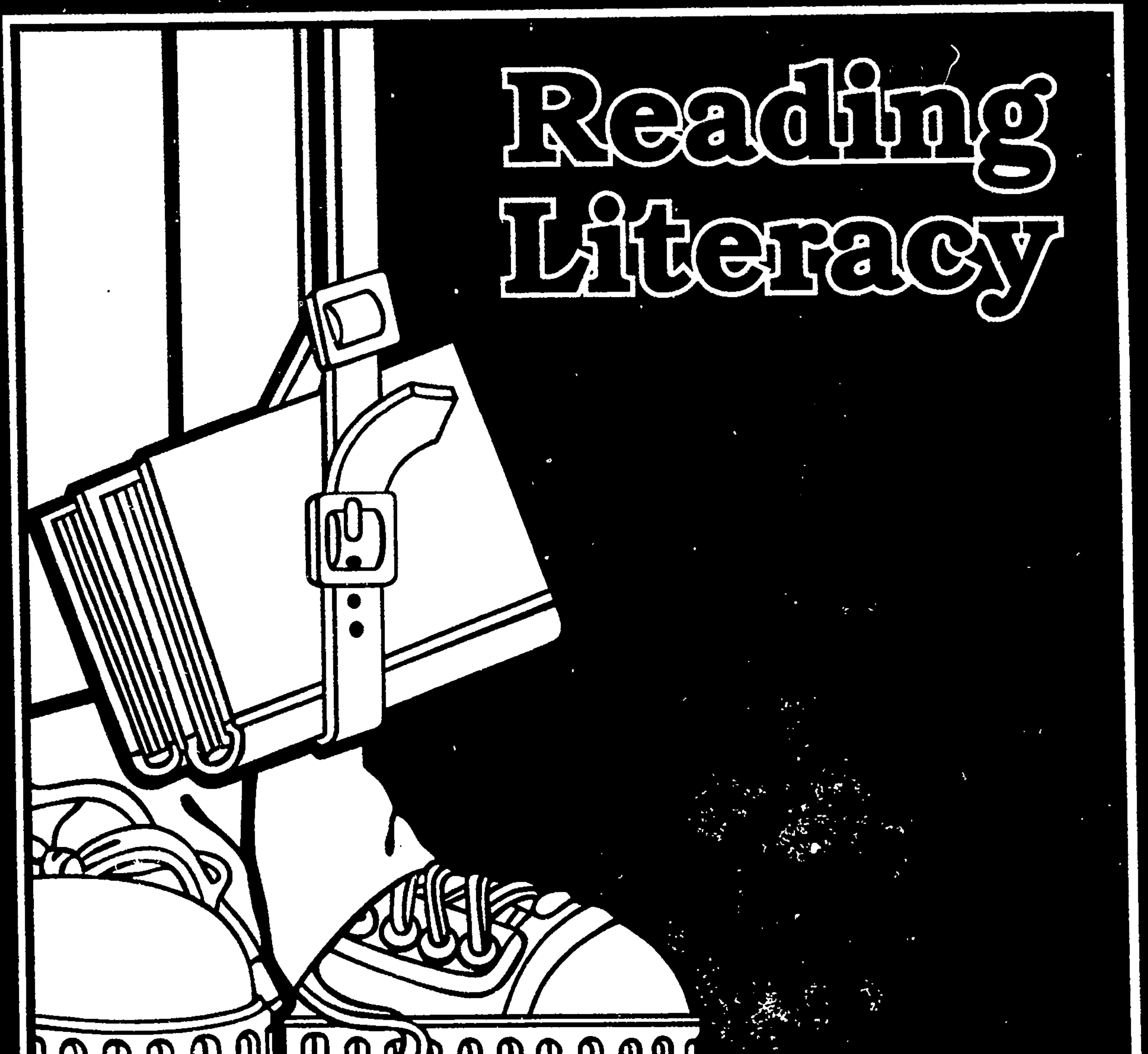
Write your answer on the lines below. Make sure you write enough to make your answer clear. You may want to use examples from the story to help explain your answer.

Do not write in this area

1 2 3

STOP!
END OF SESSION 2

Reading Literacy



Ninth Grade

Developed by: International Association for the Evaluation of Educational Achievement (IEA)
1990-91

DIRECTIONS

This reading test will be given in two sessions. Each session includes several kinds of questions. During each session you will be given practice questions so that you will understand exactly what you are expected to do.

There are a few directions that you should follow throughout the test:

- Read the directions carefully.
- Try to answer each question.
- Choose **only one** answer for each question.
- Always choose the **BEST** answer. Even if you are not sure of the answer to a question, choose what you think is the **BEST** answer and go on to the next question.
- If you change an answer, be sure to erase the first answer completely.
- **Look only at your own work.**

GO ON TO THE NEXT PAGE

SESSION I

In this session you will have several passages to read. Some are short; others are long. Some have pictures and diagrams; others have only words. Some are easy; others are hard.

Let's try this practice exercise so you will understand what you are supposed to do.

First, read this passage about Jackie.

Then, answer the four practice questions that follow it.

JACKIE

Jackie was riding her bicycle home after the storm. The river was running very high. Suddenly she realized that the railway bridge had been washed away. The train was due in 10 minutes, and the nearest station was a long way off. What could she do?

Then she remembered the red towel in her bag. Quickly she turned around, and rode back towards the oncoming train. When it came into sight, she stopped and waved her towel furiously.

That day, Jackie saved 200 passengers.

Practice Question 1.

What was Jackie doing after the storm?

[✓ A.] Riding her bike.

B. Walking home.

C. Swimming in the river.

D. Traveling on the train.

Answer A, "Riding her bike," is the right answer. You will see that a circle has been drawn around the letter "A." In each multiple choice question, you should draw a circle around the letter in front of the right answer.

Here is another practice question:

Practice Question 2.

Why did Jackie turn around and ride the other way?

A. She forgot her red towel.

B. She decided to go home.

C. She wanted to save the train.

D. She rode to the station.

Which is the correct answer? The **BEST** answer is "She wanted to save the train." Did you draw a circle around the letter "C"?

Now try Practice Question 3.

Practice Question 3.

How did she stop the train?

A. She shouted at it.
B. She waved her towel.
C. She went to the station.
D. She used the telephone.

Which is the correct answer? The **BEST** answer is "She waved her towel." Did you draw a circle around the letter "B"?

Now try Practice Question 4.

Practice Question 4.

How many passengers were saved?

____

____

What is the right answer? You should have written a sentence that said something like "200 passengers were saved."

In the next few pages, you will read more passages and do more questions like these.

Remember:

- Work as quickly as you can. When you finish each page, move on to the next until you come to the page that says **"STOP! PUT DOWN YOUR PENCIL,"** or until your teacher tells you to stop.
- Try to answer all questions, but do not spend too much time on questions you cannot do. You will have **40 minutes** to finish the questions. If a question is too difficult for you to answer, choose what you think is the **BEST** answer and move on to the next question.
- Wait until your teacher tells you to start.

DO NOT TURN THIS PAGE
UNTIL YOU ARE TOLD TO DO SO.

Read the story and answer questions 1-5.

KILLING THE FOX

(1) I killed the fox, because I had a gun in my hand when I met it. It seemed to me a matter of course that I should kill a fox if I met it in the woods and carried a gun in my hand.

(2) It was during the winter time. Snow was falling every day, and every day I walked around in the wood with a funny old gun and a black dog named Gustav. I did not hunt. Sometimes I aimed and shot at spruce cones to entertain myself and to amuse Gustav, who at every shot, jumped and barked loudly out of delight at the bang. It did not frighten him, for he had not yet learned that a gun is a deadly weapon.

(3) One day, when it was already getting dark, I met a little fox. He had been down to the village on business, and was on his way home with a hen in his mouth. I was hidden behind a juniper bush, and he ran close by me without seeing me. I aimed and shot. Why? I don't really know. I suppose this is what one does with a gun.

(4) The fox ran another few steps forward, as if nothing had happened. Then he suddenly stopped as if surprised and dropped the hen. And with a weak anxious sound he stretched out on the snow and died. Gustav, the black dog, rushed forward in wild delight with his most cheerful bark and playfully snapped at his ear. But the next moment he realized that the unknown animal was dead. There was an indescribably shy and perplexed look in his black, shining eyes. After a while he crept up to me with a whimper, his tail dragging.

(5) I left the fox there and went home, for I was suddenly cold.

(6) Next day I returned along the same path, as it was my favorite route. Whistling softly, I followed the path without thinking about what had happened the day before. Suddenly, I winced and stopped dead. On the ground before my feet lay that dead fox. The crows had picked the bloodshot, upturned eye.

(7) I stood for a while, looking at the corpse, listening to the sound of two tree branches rubbing against one another by the wind.

(8) A live fox is more beautiful than a dead one, I said to myself. And then I looked for other roads.

1. What was the fox's "business" (in Paragraph 3)?

   A. To go to the village
   B. To steal fowls and other such food
   C. To patrol the paths in the wood
   D. To meet other foxes near the village

2. Why did the author shoot the fox?

   A. He wanted to punish the fox.
   B. He was an experienced and skillful hunter.
   C. He did it without thinking.
   D. He was frightened by the fox.

3. What did Gustav do when he understood that the fox was dead?

   A. He ran home and hid himself.
   B. He crawled up to his master.
   C. He took another road.
   D. He snapped at the ear of the fox.

4. Why do you think the writer suddenly began to feel cold (in Paragraph 5)?

   A. It was starting to get dark.
   B. He was sorry about what he had done.
   C. Gustav began to whimper.
   D. It began to snow.

5. What message is the writer trying to convey?

   A. One should not kill animals without reason.
   B. One should not leave a dead fox on a path.
   C. One should bury animals so that crows can't eat their eyes.
   D. One should not follow the same path twice.

GO ON TO THE NEXT PAGE

FILL IN THE TRAVELER'S CARD

Anna checked her luggage once more. Everything was in order. Tickets, money and a brand new passport (No. 0399426, issued in Suva on 01-14-89). The photograph showed a mass of curly black hair, serious eyes and a wide nose. Anna Teresa Rama: born in Nadi on July 18th, 1974; identification number 180774-018 W; citizen of Fiji.

That was correct. She was going to be a fine representative of Fiji. The whole of Suva would be proud of her. To think that, at 16 Telau Street, Suva, there lived a girl who was going to New Caledonia to represent her country at the South Pacific Games in the high jump. Hopefully she would do well.

The journey to Noumea, New Caledonia, would start on July 15. Anna pushed her passport into her handbag and began to think of her big adventure.

Anna was given this traveler's card in Noumea. Fill it in for her using the information above.

PLEASE PRINT

6. Last Name __________ 7. First Name __________

8. Place of Birth __________ 9. Date of Birth __________

10. Home Address ______________________________
____

11. Reason For Trip (Check One)

____Business ____Visiting Relatives
____Vacation ____Other

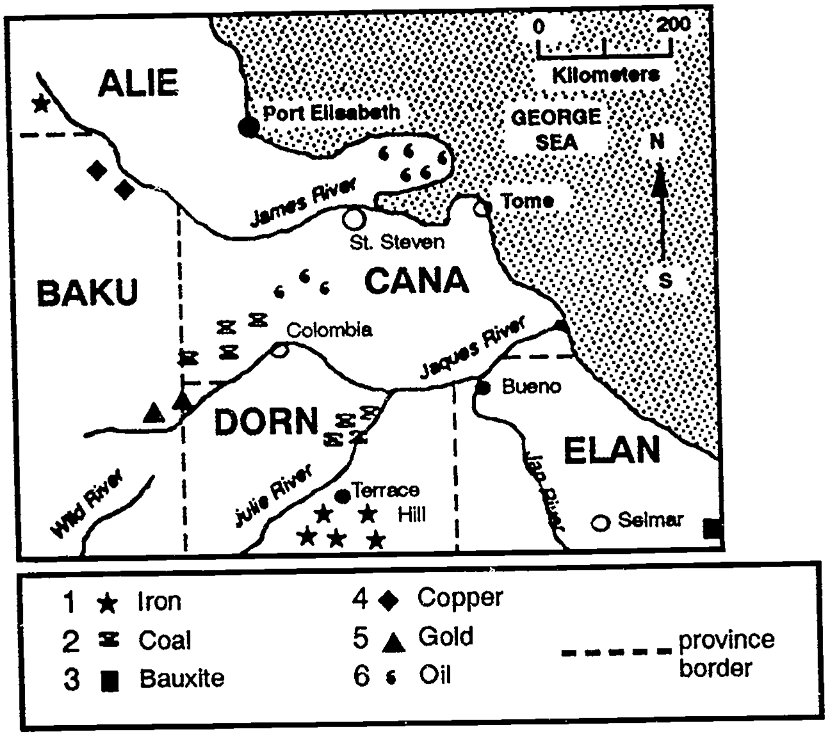
12. Passport No.: __________________

SIGNATURE: Anna Rama

OFFICIAL: (Leave Blank)

Use the map below to answer questions 13-15.

RESOURCES



13. What resource is there in the province of ELAN?

____

14. Which province has the largest amount of iron?

____

15. What resource is mined southwest of Colombia?

____

GO ON TO THE NEXT PAGE

Read the passage and answer questions 16-19.

MARMOTS

For three or four thousand years a family of marmots had been settled in a grassy little valley under the cliffs of a jagged peak whose present name is the Rock of Wonders. The maps specify that the peak is about seven thousand two hundred feet in altitude.

The spot was a remarkably suitable one for all sorts of good reasons. In the first place, the slope faced south. The sun shone on it from dawn until dark, and in the spring the snow melted there faster than anywhere else. One could warm oneself as much as one wished, or again one could sit in the shade of huge rocks fallen from the heights of the mountain. A little way off a tiny spring fed a little lake. Thanks to this fresh spring, which never dried up, the grasses round about grew thick and strong, even in the month of August.

The rocks provided many sitting places and perfectly safe holes where one could take refuge in an emergency. As for the Family Cave, hollowed out many years ago by the grandfather of the present inhabitants, it opened out pleasantly from under a flat slab between two clumps of arnica. Since it had already been improved by two generations, it would have been difficult to find a drier and more comfortable apartment.

16. Why had the marmots lived so long in one place?

    A. They did not like to travel.
    B. They could not climb down the cliffs.
    C. They came there long ago.
    D. It was a very good place.

17. What kept the grasses in the valley green and healthy all summer?

    A. The warm sun
    B. The melting snow
    C. The high peak
    D. A spring

18. Why was it a good thing that the valley faced to the south?

    A. The summer days were warmer.
    B. Winter snows melted early.
    C. The best view was to the south.
    D. Marmots need a lot of sun.

19. The last sentence of the story makes us think that the marmots will

    A. make many new homes.
    B. continue to live in the Family Cave.
    C. have many babies.
    D. work hard to make the cave comfortable.

GO ON TO THE NEXT PAGE

Use the job advertisements below to answer questions 20-22.

JOB VACANCIES

A. DELIVERY PERSON
A growing business seeks strong young person to help in stockroom, to manage and deliver stock. Driver's license required. If interested, call Mr. Boss. Ph. 123-4567.

B. TELEPHONE OPERATOR
Are you young and interested in working in a lively business? Do you like to work with many people in the same room? Then join our team as a telephone operator. We will help you learn foreign languages and extend your general knowledge. Phone the Central Exchange 456-7890 if you would like to be considered.

C. CASHIER
Do you like numbers? Do you take accuracy and confidentiality for granted? We are looking for a person who will work at the cash register as well as in the bookkeeping department. Morning hours only. Apply in writing to Moneymates Ltd., Box 12, Lincoln, NE 68218.

D. WAITRESS
Popular restaurant on outskirts of city seeks assistant waitress to work, 4 p.m. to 11 p.m. Experience required. Phone 145-9870 if interested.

Which job advertisement would you read and act on if you wanted a job

20. where you will work with large amounts of money?

A. Delivery Person
B. Telephone Operator
C. Cashier
D. Waitress

21. in which you are surrounded by many fellow workers?

A. Delivery Person
B. Telephone Operator
C. Cashier
D. Waitress

22. in which you worked part-time in the daytime?

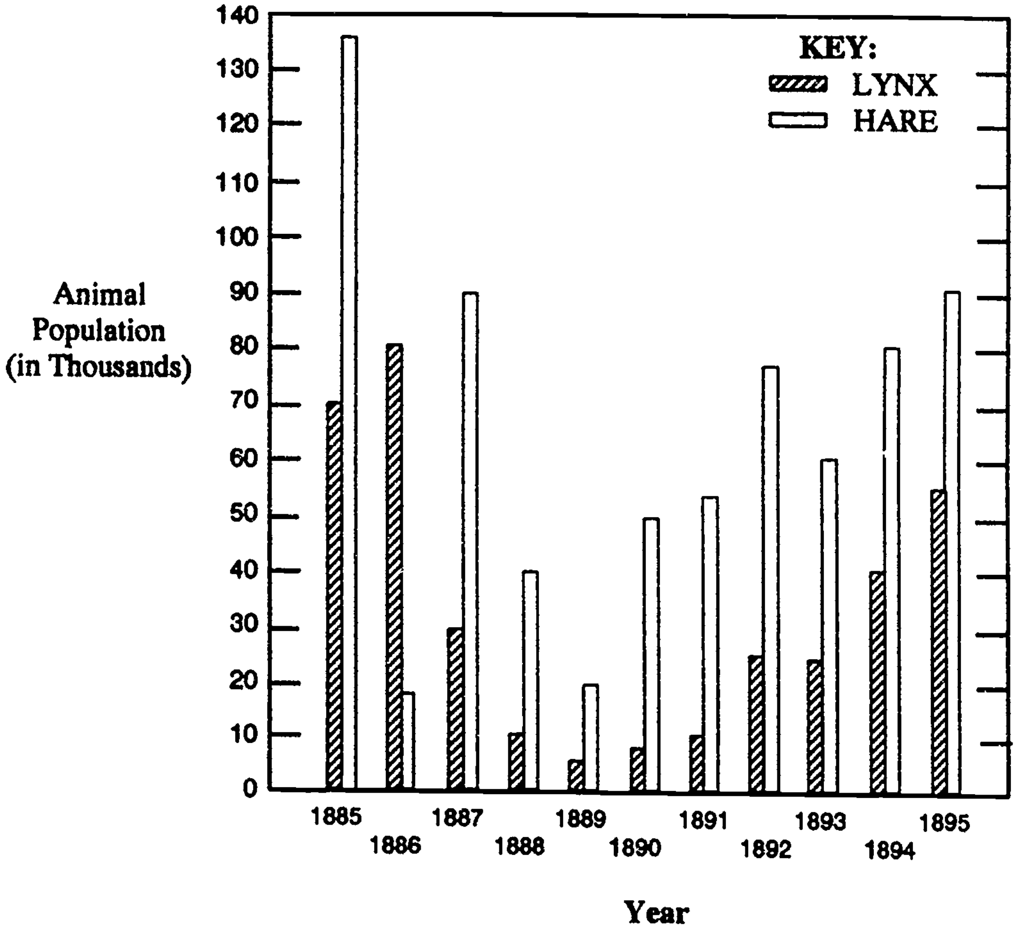
A. Delivery Person
B. Telephone Operator
C. Cashier
D. Waitress

GO ON TO THE NEXT PAGE

Use the bar graph below to answer questions 23-25.

THE SNOWSHOE HARE AND THE CANADIAN LYNX

(A Prey And Predator Cycle)



23. In which two years was the number of snowshoe hares almost the same?

A. 1888 and 1891
B. 1886 and 1894
C. 1887 and 1895
D. 1887 and 1888

24. Which year is not consistent with the general pattern shown by the graph?

A. 1885
B. 1886
C. 1889
D. 1891

25. If there was a sudden decline in the hare population, due to disease or overcrowding, which of the following would also probably occur?

A. The lynx would suddenly increase in numbers.
B. The hare population would take a decade to recover.
C. The lynx population would also decline soon after.
D. The hares would double their number the following year.

GO ON TO THE NEXT PAGE

Read the passage and answer questions 26-31.

THE PROMISE OF LASER

The Early Years

The laser is one of the most dramatic developments of the 20th century. It is a device that emits a narrow beam of light. But this is no ordinary light; it has very special properties, and lasers can be used to perform a wide range of different tasks. They can be used to cut metals at high speed and to carry out delicate machining and drilling work on hard and brittle materials. Finely controlled lasers have been used to perform delicate surgery and dentistry. They have been used to measure the speed of light with incredible accuracy, and to detect tiny amounts of distant pollutants in the atmosphere. The list of uses is increasing all the time.

All these and many other developments have taken place in little more than 20 years. The first laser was built in 1960 by Theodore Maiman. It contained a special man-made ruby rod, with a flash tube coiled around it. When the ruby rod was subjected to intense flashes of ordinary white light, it produced pulses of red laser light.

This breakthrough did not come as a complete surprise. The basic theory of the laser had been put forward by Albert Einstein in 1917. What was missing was the equipment to build lasers. Only with the rapid development of technology after World War II could the necessary equipment be built.

Once the initial discovery had been made, the flood-gates opened. Within a few years a whole host of different lasers had been produced, using solids, liquids and gases. The brilliant colors of laser light, combined with claims for its amazing properties, attracted widespread attention.

The possibility of a laser "death ray," already popular in science fiction, was immediately taken up in films. In the film "Goldfinger," James Bond was threatened with a terrible end as an "industrial" laser sliced its way through a slab of gold towards him. This image of the laser has grown and lasers have come to be regarded as the weapons of the future. The visual appeal of laser light was also exploited; pop groups and open air art shows were soon using brilliant colored laser displays.

The Laser Bears Fruit

In fact, laser developments did not quite live up to this publicity. Scientists began to discover more and more uses for lasers. But for the first ten years or so, the pace of discovery was so fast that the technologists could not keep up. Then, during the 1970s some of the early discoveries began to bear fruit.

This progress has been marked by many spectacular successes. The power of an early laser was used to project a visible spot of light onto the Moon. When the Apollo II astronauts landed on the Moon in 1969, they left behind a special mirror which has since been used to measure the distance between Earth and the Moon to within a few centimeters. Immediate use was made of high power lasers in medicine and dentistry. And as power levels increased, heavier tasks were taken on in cutting metals. In the 1980s the possibility of a real "death ray" began to emerge again.

The more gentle measuring properties of laser light have also been exploited. The properties of atoms and molecules can now be measured with an accuracy undreamed of before the advent of lasers. The speed of light has been measured with a tremendous increase in accuracy.

Less complicated uses for lasers have been found in construction work. A pencil-thin laser beam can be used to align parts of bridges and huge buildings with great precision. Even simple tasks, such as digging trenches and laying pipelines, are made easier. Lasers are also becoming an important part of new telecommunication systems. They are being used not only to transmit data but also to read it and write it.

All these developments make use of the special properties of laser light and the wide variety of materials that can be used in lasers. The fact that a large number of gases, liquids and solids can be made to produce laser light is important. Each different material produces a different form of light. So the best laser for any given task can be found from the wide range available. Lasers also come in many sizes. Huge systems are being built to see if they offer a way of producing limitless power by atomic fusion. At the other extreme, solid lasers no larger than a grain of salt have been made for telephone systems. The age of the laser is now upon us.

26. When was the first laser built according to this passage?

A. In 1917
B. During World War II
C. In 1960
D. In 1970

27. The development of lasers was <u>not</u> a complete surprise to scientists because

A. they are not complicated to build.
B. they can be used for many purposes.
C. they were often described in fiction.
D. the theory had been known for some time.

GO ON TO THE NEXT PAGE

28. What was the central component of the <u>first</u> laser?

A. Ruby rod
B. Gas rod
C. Liquid
D. Light rod

29. Which achievement was regarded by the author as a spectacular success?

A. The use of lasers in the film "Goldfinger."
B. The projection of a visible spot of light on the Moon.
C. The use of laser lights by pop groups.
D. The development of a death ray.

30. Why does the author think the age of the laser is upon us?

A. The more gentle properties of lasers have also been exploited.
B. The possibility of a real death ray has begun to emerge.
C. They are now available for a wide variety of uses.
D. There will be many developments in the next 10 years.

31. Apollo astronauts who landed on the Moon left behind a mirror which was used with lasers to

A. measure the speed of light to the Moon.
B. explore ways of developing limitless power.
C. carry out space exploration.
D. measure the distance between the Earth and Moon.

Read the story and answer questions 32-36.

MUTE

Let me tell you about the time when the repairs to the school turned out to be very different from what was expected.

One day, good old Cosme, who was the mute of the village, came running uphill through the path that led to South Nutsville. He panted, his big body staggering with the effort, his round face reddened, his shining bald head dripping with sweat, and screaming.

Screaming? But he was mute! Yes, he was the mute man of the village. Nevertheless, he was a chatter-box. He was an engaging and communicative fellow, with a chattering and cheerful nature, who was always starting conversations with whomever was about. But he had the bad fortune to be mute.

It had to be him! With so many people in the world who hardly talk and for whom hardly anything would change if they were to be mute... But no, it had to happen to him. Confusions of our chromosomes decide, before we are born, how we will be, from tip to toe. Two chromosomes that didn't get along well must have fallen to his lot. Indeed, one of them probably said, "He will be a great babbler." And the adjacent one said, "He will be mute."

Nevertheless, he had overcome his problem. He had lots of spirit and morale. With the movements of his hands, with every facial expression, and with the contortions of his whole body, he carried on non-stop chattering with everybody. When the old men of the village gathered in the square to sit in the sun, he was always the one who cheered up the conversation, since most of them were quite moody.

Anyway, he came running, shouting with his arms, and claiming everyone's attention as he pointed down-hill, with his eyes like balls and a terrible frightened look. With all those gestures he wanted to say, "They're coming! They're coming!" What they couldn't understand was who were coming. While they tried to read his gestures which seemed to indicate that they were very big and scary things, they asked him, "Are they elephants?" He shook his head.

"Are they giants?" Good old Cosme continued to say no with his head.

Suddenly, they could be heard, and indeed they did sound like giants. They roared up with a great clatter of chains and motors. When they turned up at last around a bend, the village saw what they were. Huge machines, red and yellow, with a monstrous appearance: bulldozers, dredges, and mechanical shovels, endowed with powerful jaws and claws. They seemed like giant scorpions, spiders or diplodocus, but each was made of iron and with a man in the cabin.

GO ON TO THE NEXT PAGE

Behind them followed a truck from which many workers and a foreman jumped nimbly with a roll of designs under one arm.

"Is this South Nutsville?" he asked the wondering group of open-mouthed bystanders.

"Yes."

"And the school?"

"Over there. On the outskirts of the village," they pointed.

"Come on chaps! Let's get to work! If we finish just one day late we'll get a penalty that will hurt!"

From that moment, the peaceful village was shaken by the quivers and vibrations of the earth works, in particular, during the first days, when the monstrous machines snapped at the ground. Right beside the old school the ground was levelled. They dug out the rocks in order to place a circular platform and then left. In time, many trucks loaded with huge concrete beams and queer pieces arrived, while the whole village watched, unable to believe their eyes. The teacher, most surprised of all, exclaimed, "But it isn't a repair job. It's a completely different building....!"

And they all commented, very intrigued:

"But it's round! Will it be a baseball stadium?"

"It will be huge! And we are so few!"

"Have you seen it? It has no stairs! Just curving slopes!"

"It doesn't have a single window!"

And when finally the teacher dared to show his surprise to the foreman, the foreman simply answered with a shrug, "Designs are designs."

In the village, everybody was continually astonished. They spoke of nothing else, especially the mute.

32. Near the beginning the text says that "the mute of the village arrived screaming." This means that

A. he had recovered his speech because of fright.

B. he uttered sounds similar to screams.

C. he was named "the mute" because he was talkative.

D. he expressed himself eagerly through body movements.

33. What did the writer mean by the statement that old Cosme's chromosomes didn't get along well (line 14)?

A. Fate had made a communicative man mute.
B. Old Cosme's chromosomes were mute.
C. Old Cosme's parents didn't get along well when discussing their son's fate.
D. The lonely life in the village made old Cosme mute.

34. The workers who came to build the school were in a hurry because

A. they wanted to return home quickly.
B. they would have to pay a fine if they didn't finish on time.
C. they would be rewarded if they finished before the expected date.
D. they didn't like working in front of the gaping villagers.

35. The people of South Nutsville were surprised because

A. they thought giants were coming to attack them.
B. they had never seen machines before.
C. they didn't know anything about a building job on the school.
D. the work on the school was different from what they expected.

36. What did the foreman mean with the words, "Designs are designs" (line 58)?

A. He did not understand designs.
B. He was simply following orders from his boss.
C. Designs always have a competent designer.
D. This design will produce a fine school.

GO ON TO THE NEXT PAGE

Read the passage and answer questions 37-41.

A WOMAN LEARNS TO READ

Ndugu Rukia Okashi is a 53-year old farmer living in Arusha, Tanzania. She grows maize, beans, and vegetables, has seven children, and she learned to read about ten years ago. She says:

"There is a great difference in my present situation when compared with the old days. A lot of changes have taken place. When I had to sign papers and documents, I could only use the thumb-print and I never knew exactly what I was signing. So I was sometimes cheated. Now that I can read and write no one can ask me to sign just blindly. I first have to ask myself, and it is only after I am satisfied that I agree to sign. If I don't agree with the contents of the documents, I just don't sign.

Now that I can read, I know which food is good to make me strong, which keep me well, and so on. I now can give my children a balanced diet.

In the old days, when one walked through the streets one couldn't read any signs. You may come across a 'Danger' signboard but you continue to walk ahead until someone shouts, 'Mama, mama, mama, mama, stop!' But these days, I can read all the sign-posts such as 'Don't pass here; Keep out.' In traveling also, I used to ask the driver to let me get off at a certain place, but sometimes the driver would take you much further beyond your destination. If such an incident occurs now, I shout and protest.

So now I feel great and self-confident. Now I can refuse or disagree where formerly I used to be the victim of other people because I was illiterate."

37. Which of these phrases best expresses the underlying theme of this passage?

A. The benefits of becoming literate.
B. The way in which one person became literate.
C. The problems of being an illiterate Tanzanian farmer.
D. The difficulties of coping in a literate world.

38. What is the theme of the <u>third</u> paragraph of this text?

A. Mrs. Okashi had become a better farmer.
B. She began to eat a greater range of foods.
C. Her understanding of nutrition improved.
D. Literacy is necessary in today's world.

39. What advantage does she see in learning to read signs?

- A. She can avoid trouble.
- B. She can improve her income.
- C. She is not so likely to be cheated.
- D. She doesn't make her children so ashamed.

40. Mrs. Okashi often protests loudly if she is

- A. charged too much by drivers.
- B. taken too far by drivers.
- C. mistakenly gets into danger.
- D. misled by public signs.

41. What do you think would be the disadvantages in your country for an adult who could not read or write?

Write your answer on the lines below. Make sure you give enough information to make your answer clear. You may want to use references from the passage to explain your ideas.

Do not write in this area

1 2 3

STOP!
END OF SESSION 1.

SESSION 2

In this part you will have several pages to read and questions to answer. Let's do four practice questions.

The chart below shows the final results of an international athletic contest. Use it to answer the four practice questions.

	Gold Medals	Silver Medals	Bronze Medals
Country P	4	3	2
Country Q	2	2	2
Country R	1	2	-
Country S	-	-	3

Practice Question 1.

How many gold medals has Country R won?

[✓ A.] 1 gold medal

B. 2 gold medals

C. 3 gold medals

D. 4 gold medals

Which is the correct answer? Country R has won one gold medal. You will see that a circle has been drawn around the letter "A."

Now let's do Practice Question 2.

Practice Question 2.

Which country has won 3 bronze medals?

A. Country P

B. Country Q

C. Country R

D. Country S

Which is the correct answer? Country S has won three bronze medals. Did you draw a circle around the letter "D"?

Now let's do Practice Question 3.

Practice Question 3.

Which country won the most medals altogether?

A. Country P

B. Country Q

C. Country R

D. Country S

Which is the correct answer? Country P has won nine medals. Did you draw a circle around the letter "A"?

Now let's do Practice Question 4.

Practice Question 4.

How many gold medals were awarded altogether?

____

What is the correct answer? The correct answer is "7." Did you write "7 gold medals" on the line under Practice Question 4?

In the next few pages, you will read more passages and do more questions like these.

Remember:

- Work as quickly as you can. When you finish each page, move on to the next until you come to the page that says **"STOP! PUT DOWN YOUR PENCIL,"** or until your teacher tells you to stop.
- Try to answer all questions, but do not spend too much time on questions you cannot do. You will have **40 minutes** to finish the questions. If a question is too difficult for you to answer, choose what you think is the **BEST** answer and move on to the next question.
- Wait until your teacher tells you to start.

DO NOT TURN THIS PAGE
UNTIL YOU ARE TOLD TO DO SO.

Read the passage and answer questions 1-6.

PARACUTIN

Paracutin was born in Mexico in February, 1943. At the end of one week, Paracutin was 500 feet high, and it is now over 9,000 feet high. Today Paracutin is asleep.

What is Paracutin? It is the first volcano in the world which was seen from its birth right up to the present day. On February 20, 1943, a peasant and his wife set out to work in their corn fields from the Mexican village of Paracutin. They were surprised to find the earth warm under their feet. Suddenly they heard noises deep in the earth, and a small hollow appeared in their field. In the afternoon there was a sudden loud noise and stones were flung high in the air. The peasants ran from the field and turned to watch. They saw the birth of a volcano.

There were great bursts of stone and lava, and a little hill began to form. By evening this hill was 100 feet high and hot ashes were falling on the village. At night the glare of the hot lava lit up the countryside. The trees near the village were killed and the villagers had to leave their houses. When the village was abandoned, its name was given to the volcano. The news quickly spread to Mexico City, far to the east. Many sightseers and scientists flocked to the scene. The volcano grew and grew for ten years and hundreds of square miles of forest were destroyed. Then Paracutin went to sleep. In spite of all the explosions, not one person was killed.

1. Paracutin was once the name of

   A. a peasant.
   B. a village.
   C. an old mountain.
   D. a Mexican.

2. What was destroyed in the eruption?

   A. Only a village
   B. The villagers living close by
   C. The forests and fields around Paracutin
   D. Two peasants

3. When the writer says that Paracutin "went to sleep," he means that it

   A. flattened out.
   B. stopped sending out ashes and lava.
   C. will never be a volcano again.
   D. got covered with grass and trees.

4. In this passage the author is trying to

   A. describe an interesting happening.
   B. explain a scientific theory.
   C. make us believe something.
   D. build up suspense.

5. Paracutin is now

   A. erupting.
   B. temporarily inactive.
   C. permanently dead.
   D. flattened.

6. What can we learn about volcanoes from this passage?

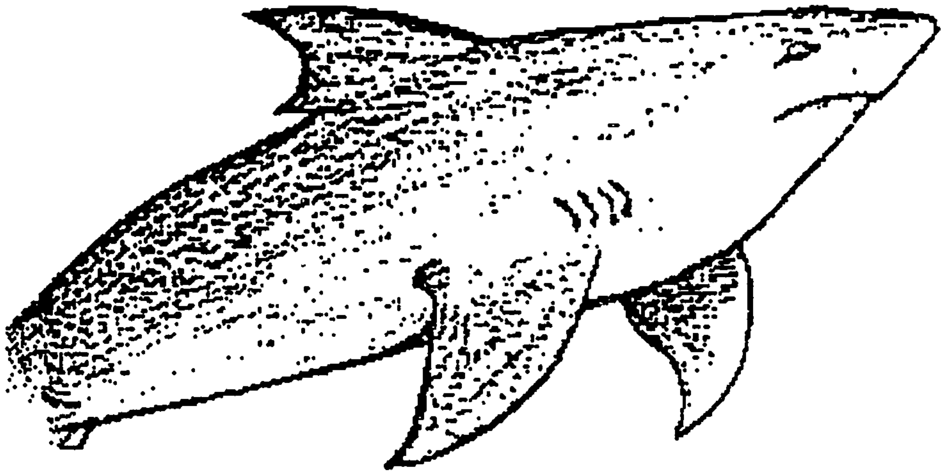
   A. New volcanoes may appear in unexpected places.
   B. There have always been volcanoes on the earth.
   C. Volcanoes are active from time to time.
   D. Volcanoes are active for only a few months.

GO ON TO THE NEXT PAGE

Read the story and answer questions 7-12.

A SHARK MAKES FRIENDS

a fable by GEORGE CIANTAR



The SHARK glided forward, too full to chase the few that had escaped. What a dinner that had been! Beautiful silver mackerel! Dozens and dozens of them - swallowed whole and thrashing. But now that the shark felt satisfied, his thoughts turned away from the food.

"I always swim alone," he said. "I have no friends. Nobody trusts me; nobody loves me. They all fear me, even my own kind. And yet ... and yet ... I know I could be so nice to my friends if only I had any. Oh, it's a lonely life being a shark!" And with that he began to cry, softly at first, then in loud racking howls of heartbreak.

A sardine heard his grief. From a safe distance, she stopped to watch and listen. "You sound unhappy," she said to the shark.

"I am. Oh, how I am!" he said between sobs. "I'm only a young shark, yet nobody ever wants to play with me or be my friend."

"That's because you eat other fish when they get too close," she said.

"I used to," said the shark, "but I never will again. I'm not interested in food any more. What I need is a friend."

The sardine was filled with pity for the shark and decided to trust him. They became good friends very quickly.

Later she took him to meet the family. The other sardines were terrified when they saw him, but she told them that he was her very best friend, and they too agreed to trust him.

They all traveled together in and out of the reefs, the sardines going about their sardine business and the shark, like a gentle giant, gliding along with them. These were happy days.

But the shark was beginning to get an uncomfortable feeling - and the longer he kept company with his new friends, the more uncomfortable the feeling became. He tried to control it but...when a shark is hungry, a shark must eat!

He stopped and let the sardines pass him. Their scales flashed as their bodies caught the light. He felt the vibrations of the water fanned by their tails.

With a powerful thrust from his tail, he swept through them from behind, then from the front, and again from the back - filling his fast-working mouth with fish. Soon he was satisfied, and far too full to chase the few who escaped. He began to think about what he had done.

"I've eaten all my friends!" he said. "Now I must swim alone again. Now I must suffer loneliness again." He sobbed loudly for many minutes.

A young pilchard heard him. He swam in a little closer to see why the shark was so unhappy.

7. Why was the shark swimming alone at the beginning of the story?

   A. He was lost in the reefs.
   B. He was too large and clumsy to swim with the other fish.
   C. Other fish didn't trust him.
   D. He didn't really like his friends.

8. The shark was unhappy when the sardine heard him because

   A. he was hungry.
   B. he was lonely.
   C. he was missing his family.
   D. he had just eaten too many mackerel.

9. The sardine became friends with the shark because in the beginning she

   A. felt sorry for him.
   B. loved him.
   C. was lonely herself.
   D. wanted to change him.

GO ON TO THE NEXT PAGE

10. How did her family feel when the sardine took the shark home?

- A. Pleased
- B. Frightened
- C. Friendly
- D. Surprised

11. What was the reason for the shark's <u>uncomfortable feeling</u>?

- A. He was not used to his new way of life.
- B. He was getting hungry.
- C. He felt crowded by the family of sardines.
- D. He had eaten too much.

12. If the writer had made this story longer, what do you think the pilchard would do next?

Write your answer on the lines below. Make sure you give enough information to make your answer clear. You may want to use references from the passage to explain your ideas.

Do not write in this area

1 2 3

This is a part of a bus schedule; use it to answer questions 13-15.

BUS SCHEDULE

Route - Weston to City / City to Weston

INWARD - TO CITY | **OUTWARD - FROM CITY**

Leaves Weston	Leaves Trump St.	Leaves Monument	Leaves Hilltop	Arrives City	Leaves City	Leaves Hilltop	Leaves Monument	Leaves Trump St.	Arrives Weston
					5:20	5:24	5:30	5:45	5:55
					5:50	5:54	6:00	6:15	6:25
					6:20	6:24	6:30	6:45	6:44
6:00	6:10	6:25	6:31	6:35	6:40	6:44	6:50	7:05	7:15
6:30	6:40	6:55	7:01	7:05	7:10	7:14	7:20	7:35	7:45
7:00	7:10	7:25	7:31	7:35	7:40	7:44	7:50	8:05	8:15
7:20	7:30	7:45	7:51	7:55	8:00	8:04	8:10	8:25	8:35
7:50	8:00	8:15	8:21	8:25	8:30	8:34	8:40	8:55	9:05
8:20	8:30	8:45	8:51	8:55	9:00	9:04	9:10	9:25	9:35
8:50	9:00	9:15	9:21	9:25	9:30	9:34	9:40	9:55	10:05
9:20	9:30	9:45	9:51	9:55	10:00	10:04	10:10	10:25	10:30
10:00	10:10	10:35	10:41	10:45	10:50	10:54	11:00	11:15	11:25
10:30	10:40	10:55	11:01	11:05	11:10	11:14	11:20	11:35	11:45
11:30	11:40	11:55	12:01	12:05	12:10	12:14	12:20	12:35	12:45

13. When does the <u>first</u> bus from Weston to City leave Monument each day?

____

14. If you miss the 8:21 bus from Hilltop to City, what time would you arrive at City if you took the next bus?

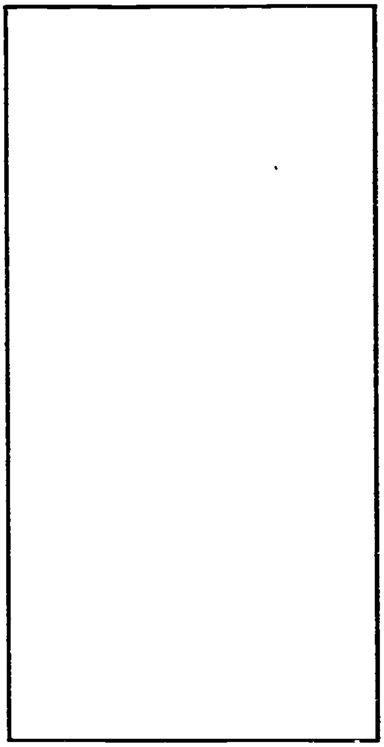
____

15. Which is the <u>latest</u> bus you can catch from Monument to arrive at Weston before 11 o'clock?

____

GO ON TO THE NEXT PAGE

Below is a rectangle. Follow the directions and then answer questions 16-18.



DIRECTIONS

1. Using 3 horizontal lines, divide the rectangle above into 4 equal sections.
2. Draw a circle in the top section.
3. Draw a square in each of the next two sections.
4. Draw a triangle standing on 1 point in the lowest section.
5. Write the number 9 in the triangle and 1 in the circle.
6. Write an 8 in the lowest square and a 6 in the top square.

16. What number is in the top section?

____

17. What is the <u>sum</u> of the numbers in the <u>top two</u> sections?

____

18. What is the <u>difference</u> between the numbers in the <u>bottom two</u> sections?

____

GO ON TO THE NEXT PAGE

GLOBAL WEATHER

The chart below gives the low and high temperatures and the weather conditions for certain cities around the world. Use this information to answer questions 19-22.

Adelaide	75/102 F	Cloudy	Los Angeles	45/66 F	Clear
Amsterdam	46/55 F	Cloudy	Madrid	37/66 F	Clear
Athens	50/61 F	Cloudy	Manila	75/88 F	Cloudy
Auckland	68/75 F	Clear	Melbourne	no	report
Bahrain	63/68 F	Clear	Moscow	32/39 F	Showers
Bangkok	81/91 F	Cloudy	New Delhi	55/81 F	Clear
Beijing	28/45 F	Cloudy	New York	14/34 F	Rain
Bonn	45/54 F	Cloudy	Paris	43/61 F	Clear
Brisbane	68/81 F	Fine	Perth	72/88 F	Fine
Brussels	37/55 F	Cloudy	Rome	36/63 F	Clear
Cairo	48/72 F	Clear	San Francisco	48/72 F	Cloudy
Copenhagen	39/41 F	Cloudy	Seoul	25/43 F	Cloudy
Darwin	72/81 F	Rain	Sydney	66/82 F	Fine
Frankfurt	39/52 F	Cloudy	Taipei	46/55 F	Cloudy
Geneva	41/54 F	Clear	Tokyo	37/43 F	Rain
Hong Kong	55/61 F	Cloudy	Toronto	25/30 F	Rain
Honolulu	68/82 F	Clear	Vancouver	21/27 F	Cloudy
Jakarta	75/91 F	Sunny	Vienna	41/54 F	Cloudy
London	48/57 F	Rain	Washington	32/37 F	Rain

19. What was the low temperature in Beijing?

____

20. In how many places were there rain or showers?

____

21. Which place reported the highest temperature?

____

22. For how many places did the temperature not rise above freezing point (32° F) all day?

____

GO ON TO THE NEXT PAGE

The chart below shows some temperature readings made at different times on four days. Use the chart to answer questions 23-27.

TEMPERATURE

	6 a.m.	9 a.m.	12 Noon	3 p.m.	8 p.m.
Monday	15°F	17°F	20°F	21°F	19°F
Tuesday	15°F	15°F	15°F	10°F	9°F
Wednesday	8°F	10°F	14°F	14°F	15°F
Thursday	8°F	11°F	14°F	17°F	20°F

23. When was the highest temperature recorded?

A. Noon on Monday
B. 3 p.m. on Monday
C. Noon on Tuesday
D. Noon on Wednesday

24. On one day the temperature dropped quickly. When do you think this happened?

A. Monday morning
B. Tuesday afternoon
C. Wednesday afternoon
D. Thursday morning

25. On how many days was it colder at 8 p.m. than at 12 noon?

A. None
B. 1
C. 2
D. 3

26. On which day did the temperature go on rising steadily from 6 a.m. to 8 p.m.?

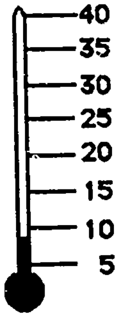
A. Monday

B. Tuesday

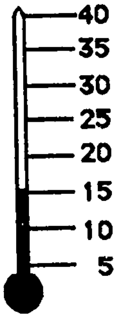
C. Wednesday

D. Thursday

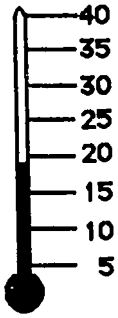
27. Which of these thermometers shows the temperature at 6 a.m. on Wednesday?



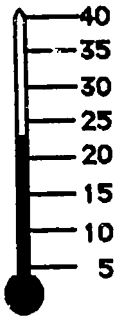
A.



B.



C.



D.

GO ON TO THE NEXT PAGE

Below is a label you would find on a container of Aspirol. Use it to answer questions 28-30.

SOLUBLE ASPIROL WITH VITAMIN C

COMPOSITION:

Each tablet contains acetyl-salicylic acid (0.4g) and ascorbic acid (240mg).

USE:

Aspirol "C" has the ability to reduce symptoms associated with ailments such as the common cold. It has the classic advantages of Aspirol in relieving pain and reducing fevers, and the therapeutic effects of Vitamin C in stimulating the body's natural defenses. When dissolved in water, it makes a very pleasant fizzy drink. The active ingredients are absorbed rapidly, and the therapeutic action is very fast.

Aspirol "C" is particularly suitable for the relief of rheumatic pains, lumbago and sciatica, headache, toothache, neuralgia, period pains, and the symptoms of influenza.

DOSE:

Adults:	1-2 tablets, 3 to 4 times per day	
Children:	- 4 to 6 years:	1/2 tablet up to 3 times per day
	- 6 to 12 years:	1 tablet up to 3 times per day

DIRECTIONS:

Soluble Aspirol "C" tablets must always be dissolved in water before use (1 tablet in 1/2 glass of water) and be taken after meals.

WARNING:

Do not give to children under 4 years of age.

Do not exceed the stated dose.

These tablets should not be taken by people with stomach ailments, such as gastro-duodenal ulcer, or by those who have an allergy to acetyl-salicylic acid or salicilates, or by those with a tendency to asthma or hemorrhages.

Consult your doctor, after 3 days of use at maximum dose level, or after 5-7 days of continual use. Consult your doctor, also, before giving Aspirol "C" to children, or young people, with viral infections like influenza or chicken-pox.

If, while taking Aspirol "C," there are continual signs of vomiting or drowsiness, stop at once.

For use in pregnancy, consult your doctor.

28. We learn from the label that to take Aspirol "C" you must

A. swallow it whole.
B. dissolve it in water.
C. chew it slowly.
D. suck it.

29. Aspirol "C" is not suitable for those who suffer from

A. stomach pains.
B. headaches.
C. period pains.
D. toothaches.

30. In one day, an adult should take no more than

A. three tablets.
B. four tablets.
C. six tablets.
D. eight tablets.

GO ON TO THE NEXT PAGE

Read the story and answer questions 31-37.

THE MAGICIAN'S REVENGE

"Now, ladies and gentlemen," said the magician, "having shown you that the cloth is absolutely empty, I will proceed to take from it a bowl of goldfish. Presto!"

All around the hall, people were saying, "Oh, how wonderful! How does he do it?" But the Quick Man in the front row said, in a big whisper, to the people near him, "He had it up his sleeve."

Then the people nodded brightly at the Quick Man and said, "Oh, of course!" And everybody whispered round the hall, "He had it up his sleeve."

"My next trick," said the magician, "is the famous Hindustani rings. You will notice that the rings are apparently separate; at a blow, they all join (clang, clang, clang). Presto!"

There was a general buzz of stupefaction until the Quick Man was heard to whisper, "He must have had another bunch up his sleeve."

Again, everybody nodded and whispered, "The rings were up his sleeve."

The brow of the magician was clouding with a gathering frown.. "I will now," he continued, "show you a most amusing trick, by which I am able to take any number of eggs from a hat. Will some kind gentleman kindly lend me his hat? Ah, thank you. Presto!"

He extracted seventeen eggs, and for thirty-five seconds the audience thought that he was wonderful. Then, the Quick Man whispered along the front bench, "He has a hen up his sleeve." And all the people whispered, "He has a lot of hens up his sleeve." The egg trick was ruined.

The show continued in this way. It transpired, from the whispers of the Quick Man, that the magician must have concealed up his sleeve, in addition to the rings, hens, and fish, several packs of cards, a loaf of bread, a doll's cradle, a live guinea pig, a fifty-cent piece and a rocking chair.

The reputation of the magician was rapidly sinking below zero. At the close of the evening, he rallied for a final effort. "Ladies and gentlemen," he said, "I will present to you, in conclusion, the famous Japanese trick recently invented by the natives of Tipperary. Will you, sir," he continued, turning toward the Quick Man, "kindly hand me your gold watch?" It was passed to him.

"Have I your permission to put it into this mortar and pound it to pieces?" he asked. The Quick Man nodded and smiled. The magician threw the watch into the mortar and grasped a sledge hammer from the table. There was a sound of violent smashing.

"He's slipped it up his sleeve," whispered the Quick Man.

"Now, sir," continued the magician, "will you allow me to take your handkerchief and punch holes in it? Thank you. You see, ladies and gentlemen, there is no deception; the holes are visible to the eye." The face of the Quick Man beamed. This time the real mystery of the thing fascinated him. "And now, sir, will you kindly pass me your silk hat and allow me to dance on it? Thank you."

The magician made a few rapid passes with his feet and exhibited the hat crushed beyond recognition. "And will you now, sir, take off your celluloid collar and permit me to burn it in the candle? Thank you, sir. And will you allow me to smash your spectacles for you with my hammer? Thank you."

By this time, the features of the Quick Man were assuming a puzzled expression. "This thing beats me," he whispered, "I don't see through it a bit."

There was a great hush upon the audience. Then the magician drew himself up to his full height and, with a withering look at the Quick Man, he concluded, "Ladies and gentlemen, you will observe that I have, with this gentleman's permission, broken his watch, burned his collar, smashed his spectacles, and danced on his hat. If he will give me further permission to paint green stripes on his overcoat, or to tie his suspenders in a knot, I shall be delighted to entertain you by so doing. If not, the performance is at an end."

Amid a glorious burst of music from the orchestra, the magician bowed, the curtain fell, and the audience, convinced that there were at least some tricks that were not up the magician's sleeve, dispersed.

31. Why was the brow of the magician <u>clouding</u> (line 17)?

A. He could not concentrate on his tricks.

B. He was running out of tricks.

C. He did not like his audience.

D. He was annoyed with the Quick Man.

GO ON TO THE NEXT PAGE

32. In lines 28-30, the author listed all the things the magician was supposed to have had up his sleeve. He did this

A. to show how clever the magician was.
B. to explain how many tricks he did.
C. to show how silly the Quick Man's claim was.
D. to show why the audience did not think the magician was very good.

33. Why was the magician's reputation sinking below zero (line 31)?

A. The audience was influenced by the Quick Man.
B. The Quick Man thought the magician was a fake.
C. The magician's skill was deserting him.
D. The magician was unkind to the Quick Man.

34. What was the "final effort" the magician "rallied for" (line 32)?

A. The hens-up-his-sleeve trick
B. His speech to the audience
C. His punishment to the Quick Man
D. A trick on the Quick Man's watch

35. Why was the Quick Man puzzled after his spectacles were broken?

A. He could not see the magician so clearly.
B. He could not explain the new trick.
C. The audience no longer agreed with him.
D. The magician was using better tricks.

36. What was the last thing the magician did?

A. Conducted the orchestra
B. Painted the Quick Man's overcoat
C. Bowed to his audience
D. Pulled down the curtain

37. How do you suppose the Quick Man would feel the next day?

A. Delighted with the evening
B. Proud of himself
C. Sorry for the magician
D. Angry with the magician

GO ON TO THE NEXT PAGE

Read the passage and answer questions 38-43.

SMOKE

The relationship between smoking and cancer, smoking and heart attacks and many other serious diseases is undeniable. Convincing evidence comes from many statistical studies that show the close relationship between the number of cigarettes smoked daily and the probability of dying of cancer or a heart attack.

The explanation for this terrible phenomenon comes from research laboratories. It has been shown that a single puff of smoke can break down the DNA in human cells, this being the long molecule which contains the cell's genetic and metabolic information. What destroys the genetic code are some tar-like substances produced by the process of combustion. In chemical terms, these are oxidizing molecules, but one can also accurately describe them as little ravenous monsters that tear apart the bonds that keep the DNA together. After each poisonous whiff, the DNA patiently reconstructs itself again, but clearly at each restoration the probability of errors increases, and in the end some malignant genes (which are always present in unstressed DNA) manage to get the upper hand and thus stimulate cancer. This is the destructive process that the cells of the organs which carry the smoke to the lungs have to undergo every time. It is not surprising that the mouth, tongue, larynx, windpipe and bronchi in smokers are more often affected by malignant tumors.

The smoke's final destination is in the lungs where, besides tar, it deposits natural radioactive substances concentrated by combustion. Each day a heavy smoker, one who smokes more than 20 cigarettes a day, absorbs the same amount of radiation which he would receive when having a chest X-ray. Nicotine, on the other hand, goes straight into the blood stream and has a strong constrictive action on the arteries. This way the circulation of blood to all the tissues diminishes. That is why skin temperature decreases, sexual organs produce fewer hormones and nervous metabolism slows down. The brain becomes less efficient and dizziness and giddiness appear, but such sensations are barely perceived by the heavy smoker. On the contrary, these are very strong sensations in those who smoke for the first time and they constitute the "drug effect" that has led many towards becoming habitual smokers.

38. The passage says that we can be sure that smoking is dangerous on the basis of

- A. the studies on the effect of smoke on the nervous system.
- B. the results of various statistical studies on smoking.
- C. the enormous number of people who die of cancer of the lungs.
- D. the change in the smoke's color and density.

39. The destruction of DNA caused by smoke is dangerous because

A. it leads to the deposit of tar-like substances in the lungs.
B. it causes cancer of the blood.
C. it may bring about error in the reconstitution of DNA.
D. it prevents DNA from reconstructing.

40. Smoke is dangerous for the lungs because

A. nicotine and tar accumulate there.
B. it causes a greater predisposition to cancer there.
C. stronger bonds form between DNA and malignant genes.
D. tar and radioactive substances are deposited there.

41. The result of nicotine's constrictive action on arteries is that

A. the arteries become larger.
B. the blood circulation slows down.
C. the temperature of the body rises.
D. the nervous metabolism speeds up.

42. The main aim of the text is to

A. describe the action of smoke on DNA.
B. show the relationship between smoking and drugs.
C. give a scientific explanation of the causes of cancer.
D. warn the readers about the dangers of smoking.

43. Which of these phrases best indicates the writer's attitude toward smoking?

A. "patiently reconstructs..."
B. "ravenous monsters..."
C. "constrictive action..."
D. "habitual smokers..."

GO ON TO THE NEXT PAGE

Read the story and answer questions 44-50.

LISTEN TO THE ANGLES LAUGHING

© by Thomaline Aguallo
illustrated by Troy Howell

Antonia is always happy in her grandmother's house. Grandmother has a lion's-paw table, and a bathtub that squats on fat white feet, and there is a stained-glass angel on her front door. Grandmother's piano is an old fat lady visiting on a Sunday afternoon. It sits in the parlor with a wide yellow smile, and it is wearing a Spanish silk shawl.

Grandmother's porch is a palace of flowers, a jungle of color for Antonia's eyes. Antonia lies in a Mexican hammock and fluffs the fur of the marmalade cat. Antonia is seven.

In Grandmother's house it is sunny and warm, and it smells of cinnamon and apples and hot herb tea. It smells of chocolate and chili and tortillas wrapped warm and waiting in a snow white cloth. It smells of the peppermint plant growing outside the kitchen door.

Grandmother is in the kitchen when Antonia comes in. She is wearing her ruby earrings and a straw sun hat. A big apron covers her dress of bright orange and yellow poppies. Grandmother's hands are white with flour, and she is singing, singing, singing Spanish songs. Antonia doesn't always understand the words, but her grandmother always sings the songs, and they make her feel good inside.

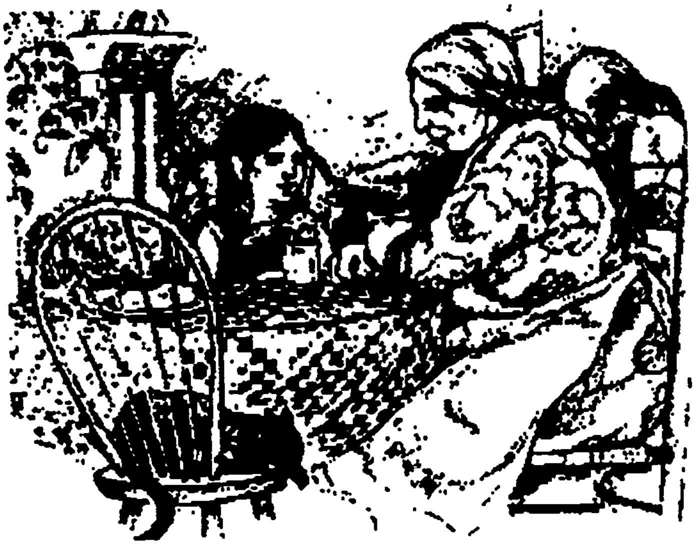
Grandmother's hair is thick and braided, and her lap is soft and warm. Grandmother's arms wrap all around Antonia as they sit in the rocking chair in the bright sunny kitchen of Antonia's grandmother's house. Grandmother has a full box of jewelry: filigree earrings and turquoise beads, an amber necklace that catches the light. Grandmother has a black veiled hat trimmed with velvet roses and a taffeta coat the color of the evening sky.

But best of all, Antonia loves the fragile old music box that Grandfather bought in Portugal. "You must open it very gently," Grandmother tells Antonia, "and then the music comes like summer rain."

When Antonia stays at Grandmother's house, she sleeps in Grandmother's own soft bed. Sometimes the dark feels too big, too quiet - and Antonia feels very small. Sometimes, then, Grandmother brings in the music box. *"Escucha los angeles riendo, mi nina,"* Grandmother says. Listen to the angels laughing, my little one. She opens the box, and Antonia hears the music. Grandmother kisses her and smoothes her long dark hair. Antonia is peaceful, and she can close her eyes.

In the summer Antonia helps her grandmother in the garden. She weeds and waters. She picks peas and lettuce and carrots and piles them in her grandmother's wicker basket. She eats fresh corn right off the stalks and tomatoes that hang like Christmas balls on tall green vines.

In the winter the wind comes. It blows the branches of the pepper tree hard against the windows of Grandmother's house. Sometimes in the evening Grandmother sews, and Antonia straightens up the old sewing box, folding the fabric, winding the thread, wrapping balls of ribbon, untangling the lace.



"Please tell me a story," Antonia says.

"I have told you all the stories I will ever know," says Grandmother, her fingers flashing like silver in the evening light. But she tells Antonia about when she was a little girl growing up in Texas.

And somehow the days at Grandmother's house always end the same way: Antonia and Grandmother sitting at the kitchen table, playing checkers with lemon drops on the red-checkered cloth. They drink hot foamy chocolate out of thick white mugs and eat hot tortillas with butter and jam oozing at the corners. They laugh together, and Grandmother sings a Spanish song.

Then one spring day the drapes are drawn, and Antonia's grandmother's house is dark and quiet. Grandmother is sick, and no one sings anymore in Grandmother's house. Antonia comes to visit as usual, but her mother is with her now. The house smells of medicine and someone else's perfume, and the air feels heavy to Antonia, as if she were wearing someone else's winter coat. Antonia knows her mother is worried because her mouth smiles for Antonia, but her eyes are old and sad. Sometimes she speaks crossly to Antonia and looks at her as though she takes up too much space in her grandmother's house. Antonia's aunts flutter like baby birds; they speak rapidly; they put their hands to their abundant hearts. They say, "SHHH!" very loudly whenever her shoes squeak on Grandmother's polished floor.

Antonia is told to sit on the brocade couch. She is given a coloring book and a box of crayons, all sharp and pointy the way she likes them. But it is lonely to color without Grandmother sitting near and telling her how well she uses colors, how nicely she keeps inside the lines.

"I want to see my grandmother," Antonia says, pulling at the jacket of her Aunt Mercedes's dress.

"Ay, que nina!" Aunt Mercedes scolds. "Your grandmother needs a lot of rest. She must have quiet now."

GO ON TO THE NEXT PAGE

"But I want to help my grandmother!" Antonia says. "Mama, can Grandmother see that the orange trees are blooming?"

"Oh, Antonia," says her mother. "Please be still. If you want to help your grandmother, just sit down on the couch. Be quiet as a mouse. That's all you can do."

But Antonia knows they are wrong. She knows she can do more.

She pictures her grandmother lying in a room where it is always night. What does she think about? Does she know when the sun is shining? Does she know the strawberries are flowering on the berry patch by the side of the house? Antonia remembers how it feels to be in the dark when it seems that you are all alone.

So one day when there is no one to notice her or to tell her to stop, Antonia goes up the stairs to her grandmother's room. In her arms she carries something wrapped up like a baby. Grandmother is lying quietly in her big soft bed. The room is shadowy and hot, and the only light comes from a candle in a small red glass.

As softly as she can, Antonia tiptoes to Grandmother's window. She draws back the curtain and pulls up the shade. She opens the window. She lets in the light and the smell of wet grass and sweet peas and ripening strawberries.

Antonia sits on her grandmother's bed. "Grandmother," Antonia says. "Listen, Grandmother." She unwraps her grandmother's music box. She opens it with fingers made of feathers. The music box sings like summer rain.

"Oh, Grandmother," Antonia says. *"Yo te amo, abuelita."* I love you, little grandmother.

Grandmother's eyes are open and smiling. *"Yo te amo también, mijita,"* Grandmother says. I love you also, little daughter.

Antonia holds her grandmother's hand. She sings a bit of a Spanish song. They listen to the angels laughing. They both feel good inside.

44. What is the major theme of the story?

A. Antonia understood that sick people usually enjoy hearing music.
B. Antonia learned that if Grandmother rested, she would get well.
C. Antonia understood what Grandmother needed and she supplied it.
D. Antonia learned that disobeying Grandmother was sometimes necessary.

45. One reason Antonia enjoyed spending time at Grandmother's house was that she

A. did not have to do homework there.
B. had interesting things to do there.
C. could go to bed as late as she wanted there.
D. liked to listen to the radio with Grandmother.

46. Antonia helped her grandmother by

A. sharing with her the things that she liked.
B. sneaking in to play checkers with Grandmother.
C. bringing in the cat to keep Grandmother company.
D. bringing Grandmother tortillas dripping with jam and butter.

47. When Antonia tiptoes up to see her grandmother, it tells us that Antonia

A. is often a problem child for her parents.
B. loves Grandmother enough to find a way to help her.
C. is selfish to the point of risking her grandmother's health.
D. will do anything to please her mother and aunts.

GO ON TO THE NEXT PAGE

48. Which words best describe Grandmother's relationship with Antonia?

A. Strict but loving
B. Happy and silly
C. Worried and careful
D. Warm and understanding

49. At the end of this story, if Mother found Antonia in Grandmother's room, Mother would probably

A. believe that Grandmother was well.
B. know that Grandmother was angry.
C. feel sorry for Antonia and Grandmother.
D. realize that Antonia made Grandmother feel better.

50. Which sentence is most important to include in a summary of this story?

A. Antonia's aunts talked very loudly.
B. Antonia's mother gave her a coloring book.
C. Antonia carried the music box upstairs to her sick grandmother.
D. Antonia did not understand her grandmother's Spanish songs.

STOP!
END OF SESSION 2

Attachment A-2

Student Questionnaires

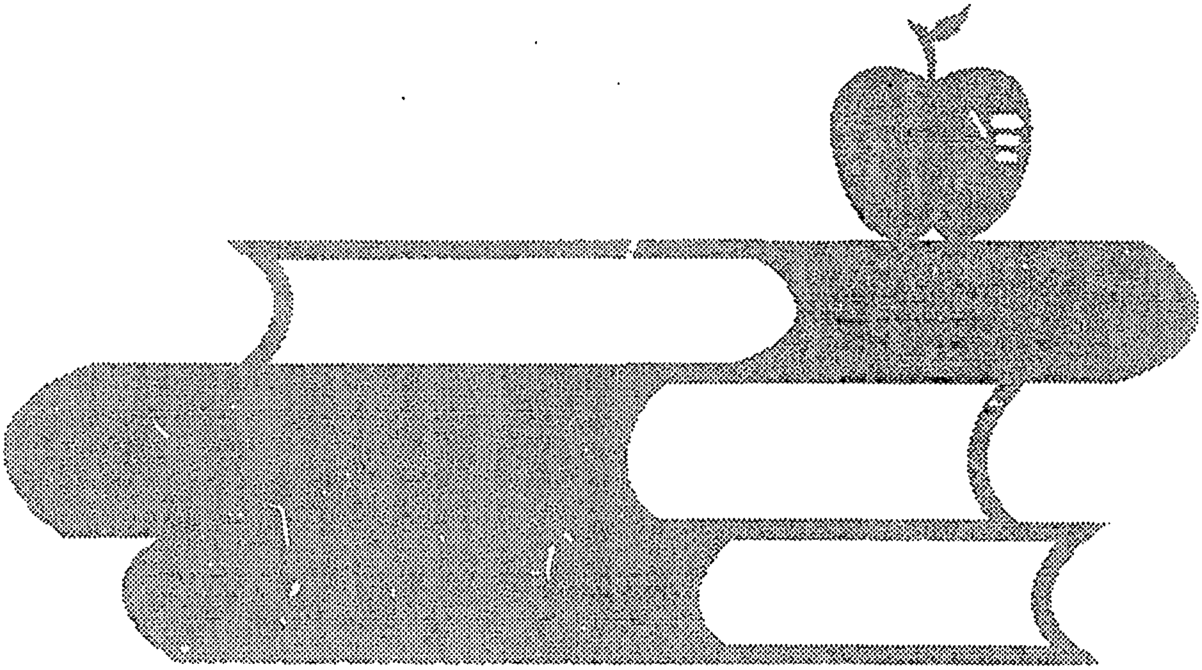
OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92

ID: ____________________

READING LITERACY

STUDENT QUESTIONNAIRE

Fourth Grade



CODE: ________

Directions:

Please answer all of the following questions as best as you can. If you need help, ask me.

A. Questions 1 to 14 are about you and your home.

1. How old were you on your last birthday?

___ ______ Years old

2. When were you born?

|__________||___|___||___|___|
Month Day Year

3. Are you a boy or a girl? *(Circle only one.)*

Boy 1
Girl 2

4. What is your race/ethnicity? *(Circle only one.)*

a. Asian or Pacific Islander (including any of the following) 1
Chinese, Filipino, Japanese, Korean, Vietnamese, Laotian, Cambodian/ Kampuchean, Thai, Samoan, Guamanian, Asian Indian, Pakistani, Bangladeshi, Sri Lankan, Iranian, Afghan, Turkish, Iraqi, Israeli, Lebanese, and Other Asian
b. American Indian or Alaskan native 2
c. Hispanic (including any of the following)...... 3
Mexican, Mexican American, Chicano, Cuban, Puerto Rican, and Other Hispanic
d. White (non-Hispanic), or 4
e. Black (non-Hispanic) 5

5. **What is the highest level of schooling your father or male guardian has completed?** *(Circle only one.)*

Elementary school	1
Junior high school	2
Some high school	3
High school	4
Some college	5
College or university	6

6. **What is the highest level of schooling your mother or female guardian has completed?** *(Circle only one.)*

Elementary school	1
Junior high school	2
Some high school	3
High school	4
Some college	5
College or university	6

7. **How often do you eat each of the following meals?** *(Circle one number on each line.)*

	Never or almost never	1 or 2 times a week	3 or 4 times a week	Every day
Breakfast	1	2	3	4
Lunch	1	2	3	4
Dinner or supper	1	2	3	4

8. **Does your family regularly get or see a newspaper at home?** *(Circle one only.)*

No	1
Yes	2

9. **On a school day, about how many hours do you usually watch TV or video outside of school hours?**

________ Hours (per day)

10. **About how many books are there in your home?** *(Do not count newspapers, comic books, or magazines; circle one only.)*

None	1
1 - 10	2
11 - 50	3
51 - 100	4
101 - 200	5
More than 200	6

11. How many of the following things can be found at your home?
(Circle one number on each line. Be sure to circle "0" if you have none.)

		None	One	More than one
a.	Automobile that runs	0	1	2
b.	Bicycle	0	1	2
c.	Cassette recorder (or walkman)	0	1	2
d.	Clothes-dryer	0	1	2
e.	Color TV set	0	1	2
f.	Computer	0	1	2
g.	Dish-washer	0	1	2
h.	Microwave oven	0	1	2
i.	Recreational vehicle (e.g., camper, mobile home, etc.)	0	1	2
j.	Refrigerator	0	1	2
k.	Stereo set	0	1	2
l.	Telephone	0	1	2
m.	Video recorder (VCR)	0	1	2
n.	Atlas	0	1	2
o.	Dictionary	0	1	2
p.	Encyclopedia	0	1	2
q.	Typewriter	0	1	2
r.	Pocket calculator	0	1	2

12. Do you have a specific place to study in your home?

No	1
Yes	2

13. How many people including yourself live in your home?

_______ People

14. Which of the following people live in the same household with you? *(Circle one number on each line.)*

	No	Yes
a. Father	1	2
b. Other male guardian (stepfather or foster father)	1	2
c. Mother	1	2
d. Other female guardian (stepmother or foster mother)	1	2
e. Brother(s) (including step- or half-)	1	2
f. Sister(s) (including step- or half-)	1	2
g. Grandparent(s)	1	2
h. Other relative(s) (children or adults)	1	2
i. Non-relative(s) (children or adults)	1	2

581

B. Questions 15 and 16 are about your use of a language other than English.

15. Do you speak any language other than English at home? *(Circle only one.)*

No	1
Yes	2

16. What was the first language you learned to speak when you were a child? *(Circle only one.)*

English	1	(Go to Q.31)
Spanish	2	
Chinese	3	
Japanese	4	
Korean	5	
A Filipino language	6	
Italian	7	
French	8	
German	9	
Greek	10	
Polish	11	
Portuguese	12	
Other (Specify) ______________	13	

C. Questions 17 to 24 are about how well you use a language other than English.

17. How well do you understand that language when it is spoken to you? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

18. How well do you speak that language? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

19. How well do you read that language? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

20. How well do you write in that language? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

21. Do you attend classes (in or out of school) to study that language? *(Circle only one.)*

No	1
Yes	2

22. **How often do you use that language in each situation listed below?** *(Circle one on each line.)*

	Never	Some-times	About half of the time	Always or most of the time	Does not apply
a. With your parents or guardians?	0	1	2	3	4
b. With your brothers or sisters?	0	1	2	3	4
c. With your friends?	0	1	2	3	4
d. With your relatives?	0	1	2	3	4

23. **How often do people <u>at home</u> read to you in <u>that language</u>?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

24. **Is there any other place <u>outside of school and your home</u> where someone reads to you in <u>that language</u>?** *(Circle one only.)*

No	1
Yes	2

584

D. Questions 25 to 32 are about how well you use English.

25. How long have you been using English to do school work? *(Circle one only.)*

Less than 1 year	1
1 year	2
2 years	3
3 years	4
4 years or more	5

26. How often do you speak English at home? *(Circle only one.)*

Always	1
Almost always	2
Sometimes	3
Hardly ever	4
Never	5

27. How well do you understand spoken English? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

28. How well do you speak English? *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

29. **How well do you read in English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

30. **How well do you write in English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

31. **How often do people at home read to you in English?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

32. **Is there any other place outside of school and your home where someone reads to you in English?** *(Circle one only.)*

No	1
Yes	2

E. Questions 33 to 45 are about your reading.

33. Is there a public library in your neighborhood?

No	1
Yes	2
Don't know	3

34. How often do you borrow books from a school or public library? *(Circle one only.)*

Never	1
Hardly ever	2
Once a month	3
Once a week	4
More than once a week	5

35. How good are you at reading? *(Circle one only.)*

Very poor	1
Poor	2
Average	3
Good	4
Very good	5

36. **Please tell us what you think are the three most important ways to become a good reader?** *(Choose three only.)*

	Most important
a. Liking it	2
b. Spending lots of time reading	2
c. Being able to concentrate well	2
d. Knowing how to sound out words	2
e. Learning the meaning of lots of words	2
f. Having many good books around	2
g. Having a lively imagination	2
h. Having lots of reading for homework	2
i. Having lots of drill at hard things	2
j. Having lots of written exercises	2
k. Being told how to do it	2
l. Other (Specify) ____________________	2

____

37. **Are you in a special class to help you read at your grade level?**

No	1
Yes	2

38. **Do you read aloud at home?** *(Circle only one.)*

No	1	(Go to Q.42)
Yes	2	

39. **How often do you read aloud to someone at home?** *(Circle only one.)*

Less than once a week	1
1 to 3 times a week	2
Nearly every day	3

40. **To whom do you read aloud at home?** *(Circle only one.)*

Parent(s)	1
Brother(s) or sister(s)	2
Other person	3

41. **What do you most often read aloud at home?** *(Circle only one.)*

Newspapers	1
Magazines	2
Books	3
Textbooks	4
Comic books	5
Letters	6
Words on television screens	7
Other	8

42. **How often do your parents or other people at home ask you what you have been reading at school?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

43. How often do you do each of the following before you begin to read? *(Circle one number on each line.)*

	Never	Once in a while	Quite often	Most of the time
a. Look at the title, illustrations and headings to find out what it is likely to be about	1	2	3	4
b. Think about what you already know about the topic	1	2	3	4
c. Remember other selections about the same topic	1	2	3	4
d. Try to guess what will happen or what information you might learn	1	2	3	4
e. Talk to somebody else about it.....	1	2	3	4
f. Pinpoint issues that you are interested in exploring	1	2	3	4

44. How often do you do each of the following while you are reading something for the first time? *(Circle one number on each line.)*

	Never	Once in a while	Quite often	Most of the time
a. Picture in your mind what is happening	1	2	3	4
b. Make notes	1	2	3	4
c. Stop and think about what you have already read	1	2	3	4
d. Read over the materials again	1	2	3	4
e. Try to guess what will come next..	1	2	3	4

45. **How often do you do each of the following after you have finished reading?** *(Circle one number on each line.)*

	Never	Once in a while	Quite often	Most of the time
a. Look back over what you have read	1	2	3	4
b. Write down notes or ideas	1	2	3	4
c. Compare what you have read with things that have happened to you, feelings you have had, or things you have seen	1	2	3	4
d. Think about related selections you have read	1	2	3	4
e. Get new ideas about things to read or research	1	2	3	4
f. Talk to somebody else about it	1	2	3	4
g. Write something of your own on that topic	1	2	3	4

F. Questions 46 to 54 are about your reading homework.

46. **Do you get reading assignments to do by yourself?**

No	1	(Go to Q.55)
Yes	2	

47. **How often do you get reading homework?** *(Circle one only.)*

1 or 2 times a week	1
3 or 4 times a week	2
Nearly every day	3

48. **About how much time do you spend on your <u>reading homework</u>?** *(Circle one only.)*

Up to a quarter of an hour	1
Up to half an hour	2
More than half an hour	3

49. **How often are you asked questions in class about your <u>reading homework</u>?** *(Circle only one.)*

Always	1
Most of the time	2
Sometimes	3
Hardly ever	4
Never	5

50. **How often are you helped with your <u>reading homework</u>?** *(Circle only one.)*

I rarely get help	1
I sometimes get help	2
I get help most of the time	3

51. **If you don't finish your <u>reading work</u> at school, how often are you expected to finish it on your own time?** *(Circle only one.)*

Always	1
Most of the time	2
Sometimes	3
Hardly ever	4
Never	5

52. **How often are you given <u>written</u> assignments about the work in reading?** *(Circle only one.)*

Always	1
Most of the time	2
Sometimes	3
Hardly ever	4
Never	5

53. **Which kinds of reading work do you normally do for homework?** *(Circle only one on each line.)*

	No	Yes
a. We read and write answers to the teacher's questions	1	2
b. We read but do not have questions to answer	1	2
c. We choose what to read and report back to the teacher or class	1	2
d. We choose what to read but do not report back to the class	1	2
e. Other (Specify)________________________	1	2

54. **Which of the following people regularly help you with school work done at home?** *(Please circle one answer on each line.)*

	No	Yes
a. Mother	1	2
b. Father	1	2
c. Brother or sister	1	2
d. Someone paid to help you (a tutor)	1	2
e. Other person	1	2

G. Questions 55 to 63 are about reading for enjoyment.

55. Did you read a <u>book</u> for your own enjoyment last week? *(Circle only one.)*

No	1
Yes	2

(If "Yes," write in the title or author of the book.)

Book title: ______________________________

or

author: ______________________________

56. How often do you read <u>books</u> for your own enjoyment? *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

57. Did you read a <u>comic book</u> last week? *(Circle only one.)*

No	1
Yes	2

(If "Yes," write in the title or the person in the story.)

Comic book title or person: ______________________________

58. **How often do you read comic books?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

59. **Did you read a magazine last week?** *(Circle only one.)*

No	1
Yes	2

(If "Yes," write in the title of the magazine or the topic you read about.)

Magazine title or topic: ____________________

60. **How often do you read magazines?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

61. **Did you read a newspaper last week?** *(Circle only one.)*

No	1
Yes	2

(If "Yes," write in the name of the newspaper.)

Newspaper name: ____________________

62. **How often do you read <u>newspapers</u>?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

63. **Other than for school work, how often do you read written directions or instructions?** *(You might read them to put a toy together, to follow a recipe, to use a tool or to do something else. Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

H. Questions 64 to 69 are about reading in school.

64. **In school, how often do you read <u>textbooks</u> in reading or language class?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

65. **How often do you read <u>story books</u> in addition to your textbooks in reading or language class?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

66. **How often do you use workbooks or practice exercises in reading or language class?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

67. **In school, how often do you read textbooks or do practice exercises in science, geography or environmental studies?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

68. **How often do you look up information in books like encyclopedias, dictionaries, manuals or maps for school work?** *(Circle only one.)*

Almost never	1
About once a month	2
About once a week	3
Almost every day	4

69. **Do you have a favorite book?** *(If so, fill in the title below.)*

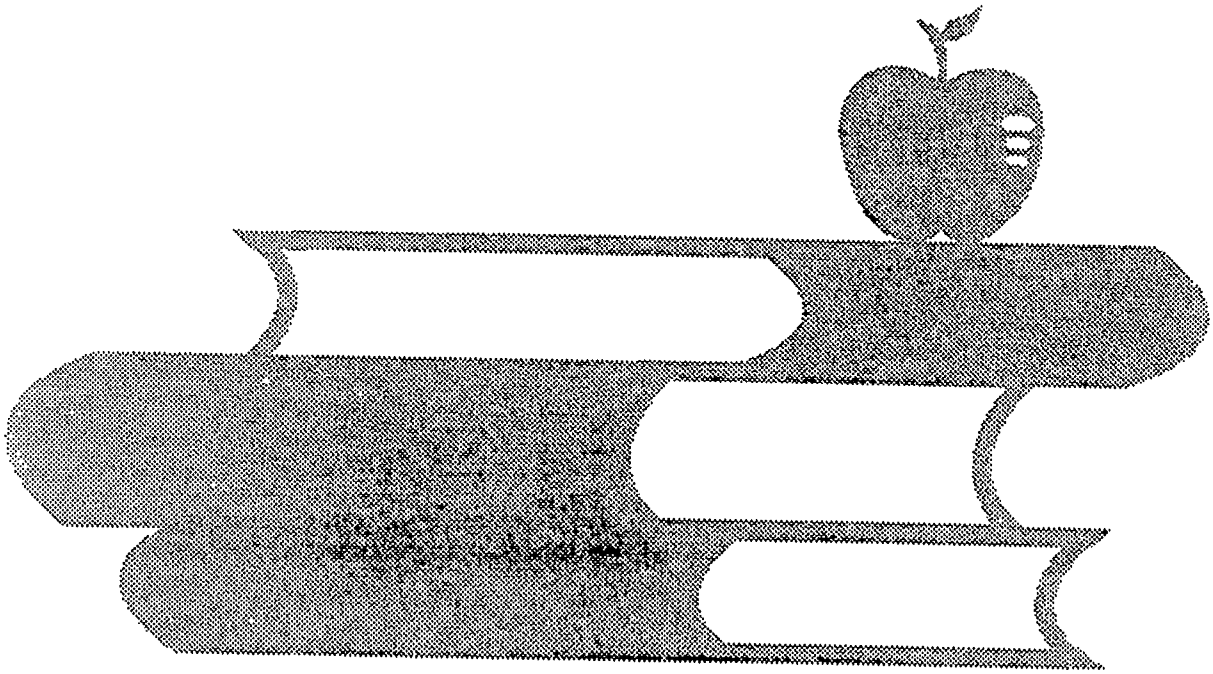
Thank you very much for your cooperation!

OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92
ID: ______________________

READING LITERACY

STUDENT QUESTIONNAIRE

Ninth Grade



CODE: ________

Directions:

Please answer all of the following questions as best as you can. If you need help, ask me.

A. Questions 1 to 20 are about you and your home.

1. **How old were you on your last birthday?**

_____ Years old

2. **When were you born?**

|__|__| Month |__|__| Day |__|__| Year

3. **What is your sex?** *(Circle only one.)*

Male	1
Female	2

4. **What is your race/ethnicity?** *(Circle only one.)*

a.	Asian or Pacific Islander (including any of the following) Chinese, Filipino, Japanese, Korean, Vietnamese, Laotian, Cambodian/Kampuchean, Thai, Samoan, Guamanian, Asian Indian, Pakistani, Bangladeshi, Sri Lankan, Iranian, Afghan, Turkish, Iraqi, Israeli, Lebanese, and Other Asian	1
b.	American Indian or Alaskan native	2
c.	Hispanic (including any of the following) Mexican, Mexican American, Chicano, Cuban, Puerto Rican, and Other Hispanic	3
d.	White (non-Hispanic), or	4
e.	Black (non-Hispanic)	5

5. **What is the highest level of schooling your father or male guardian has completed?** *(Circle only one.)*

Elementary school	1
Junior high school	2
Some high school	3
High school	4
Some college	5
College or university	6

6. **What is the highest level of schooling your mother or female guardian has completed?** *(Circle only one.)*

Elementary school	1
Junior high school	2
Some high school	3
High school	4
Some college	5
College or university	6

7. **How often do you eat each of the following meals?** *(Circle one number on each line.)*

	Never or almost never	1 or 2 times a week	3 or 4 times a week	Every day
Breakfast	1	2	3	4
Lunch	1	2	3	4
Dinner or supper	1	2	3	4

8. **Do you have regular (almost every day) responsibilities in your home (e.g., helping with housework or family business) before or after school?** *(Circle only one.)*

No	1	**(Go to Q.10)**
Yes	2	

9. **How much time each day do you spend on that regular responsibility in your home?** *(Circle only one.)*

Up to half an hour per day	1
Between half an hour and one hour per day	2
Between one and two hours per day	3
More than two hours per day	4

10. **Do you have a paid job outside your home?** *(Circle only one.)*

No	1	**(Go to Q.12)**
Yes	2	

11. **How many hours do you work a week for pay on your present job?** *(Circle only one.)*

Up to 4 hours a week	1
5 - 10 hours a week	2
11 - 20 hours a week	3
21 or more hours a week	4

12. **As things stand now, how far in school do you think you will get?** *(Circle only one.)*

Won't finish high school	1
Will graduate from high school, but won't go any further	2
Will go to vocational, trade, or business school after high school	3
Will attend college	4
Will graduate from college	5
Will attend a higher level of school after graduating from college	6

13. **Does your family regularly get or see a newspaper at home?** *(Circle one only.)*

No	1
Yes	2

14. **On a school day, about how many hours do you usually watch TV or video outside of school hours?**

________ Hours (per day)

15. **About how many books are there in your home?** *(Do not count newspapers, comic books, or magazines; circle one only.)*

None	1
1 - 10	2
11 - 50	3
51 - 100	4
101 - 200	5
More than 200	6

16. **How often do your parents or other people at home ask you about what you are reading?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

17. **How many of the following things can be found at your home?** *(Circle one number per line only. Be sure to circle "0" if you have none.)*

		None	One	More than One
a.	Automobile that runs	0	1	2
b.	Bicycle	0	1	2
c.	Cassette recorder (or walkman)	0	1	2
d.	Clothes-dryer	0	1	2
e.	Color TV set	0	1	2
f.	Computer	0	1	2
g.	Dish-washer	0	1	2
h.	Microwave oven	0	1	2
i.	Recreational vehicle (campers, mobile home, etc.)	0	1	2
j.	Refrigerator	0	1	2
k.	Stereo set	0	1	2
l.	Telephone	0	1	2
m.	Video recorder (VCR)	0	1	2
n.	Atlas	0	1	2
o.	Dictionary	0	1	2
p.	Encyclopedia	0	1	2
q.	Typewriter	0	1	2
r.	Pocket calculator	0	1	2

18. **Do you have a specific place to study in your home?**

No 1
Yes 2

19. **How many people including yourself live in your home?**

_____ People

20. **Which of the following people live in the same household with you?** *(Circle one number on each line.)*

	No	Yes
a. Father	1	2
b. Other male guardian (stepfather or foster father)	1	2
c. Mother	1	2
d. Other female guardian (stepmother or foster mother)	1	2
e. Brother(s) (including step- or half-)	1	2
f. Sister(s) (including step- or half-)	1	2
g. Grandparent(s)	1	2
h. Other relative(s) (children or adults)	1	2
i. Non-relative(s) (children or adults)	1	2

B. Questions 21 and 22 are about your use of a language other than English.

21. Do you speak any language other than English at home? (Circle one only.)

No	1
Yes	2

22. What was the first language you learned to speak when you were a child? (Circle only one.)

English	1	(Go to Q.37)
Spanish	2	
Chinese	3	
Japanese	4	
Korean	5	
A Filipino language	6	
Italian	7	
French	8	
German	9	
Greek	10	
Polish	11	
Portuguese	12	
Other (Specify) ________________	13	

605

C. Questions 23 to 30 are about how well you use a language other than English.

23. **How well do you understand that language when it is spoken to you?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

24. **How well do you speak that language?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

25. **How well do you read that language?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

26. **How well do you write in that language?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

27. **Do you attend classes (in or out of school) to study that language?** *(Circle only one.)*

No	1
Yes	2

28. **How often do you use that language in each situation listed below?** *(Circle one on each line.)*

	Never	Sometimes	About half of the time	Always or most of the time	Does not apply
a. With your parents or guardian?	0	1	2	3	4
b. With your brothers or sisters?	0	1	2	3	4
c. With your friends?	0	1	2	3	4
d. With your relatives?	0	1	2	3	4

29. **How often do people <u>at home</u> read to you in that language?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

30. **How often does someone read to you in that language <u>outside of school and your home</u>?** *(Circle one only.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

607

D. Questions 31 to 38 are about your use of English.

31. **How long have you been using English to do school work?** *(Circle only one.)*

Less than 1 year	1
1 year	2
2 years	3
3 years	4
4 or more years	5

32. **How often do you speak English at home?** *(Circle only one.)*

Always	1
Almost always	2
Sometimes	3
Hardly ever	4
Never	5

33. **How well do you understand spoken English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

34. **How well do you speak English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

35. **How well do you read in English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

36. **How well do you write in English?** *(Circle only one.)*

Not at all	1
Not well	2
Well	3
Very well	4

37. **How often do people at home read to you in English?** *(Circle only one.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

38. **How often does someone read to you in English outside of school and your home?** *(Circle only one.)*

Never	1
1 or 2 times a week	2
3 or 4 times a week	3
Nearly every day	4

E. Questions 39 to 47 are about your homework.

39. **How often are you given homework?** *(Circle one only.)*

Never	1
1 or 2 days a week	2
3 or 4 days a week	3
5 days or more a week	4

40. **About how much time do you spend on your homework when it is given?**

_______ Hours and _______ Minutes

41. **How often are you given English homework?** *(Circle one only.)*

Never	1
1 or 2 days a week	2
3 or 4 days a week	3
5 days or more a week	4

42. **About how much time do you spend on English homework when you get it?**

_______ Hours and _______ Minutes

43. **How often are you given <u>reading</u> to do at home by your English teacher?** *(Circle one only.)*

Never	1
1 or 2 days a week	2
3 or 4 days a week	3
5 days or more a week	4

44. **How often are you given <u>reading</u> to do at home by teachers other than your English teacher?** *(Circle one only.)*

Never	1
1 or 2 days a week	2
3 or 4 days a week	3
5 days or more a week	4

45. **How often are you asked questions in class about your <u>English</u> <u>homework</u>?** *(Circle one only.)*

I do not get English homework	1
Always	2
Most of the time	3
Sometimes	4
Hardly ever	5
Never	6

46. **Which kinds of reading work do you normally do for homework?** *(Circle only one number per line.)*

	No	Yes
a. We read and write answers to the teacher's questions	1	2
b. We read but do not have questions to answer	1	2
c. We choose what to read and report back to the teacher or class	1	2
d. We choose what to read but do not report back to the class	1	2
e. Other (Specify)________________	1	2

47. **Which of the following people help you with school work done at home?** *(Please circle one answer on each line.)*

	No	Yes
a. Mother	1	2
b. Father	1	2
c. Brother or sister	1	2
d. Someone paid to help you (a tutor)	1	2
e. Other person	1	2

F. Questions 48 to 58 are about your reading.

48. **Do you read aloud at home?** *(Circle only one.)*

No	1	**(Go to Q.52)**
Yes	2	

49. **How often do you read aloud to someone at home?** *(Circle only one.)*

Less than 1 time a week	1
1 to 3 times a week	2
Nearly every day	3

50. **To whom do you read aloud at home?** *(Circle only one.)*

Parent(s)	1
Brother(s) or sister(s)	2
Other person	3

51. **What do you read aloud at home?** *(Circle only one.)*

Newspapers	1
Magazines	2
Books	3
Textbooks	4
Comic books	5
Letters	6
Words on television screens	7

52. **Is there a public library in your neighborhood?** *(Circle only one.)*

No	1
Yes	2
Don't know	3

53. **How often do you borrow books from a school or public library?** *(Circle one only.)*

Never	1
Hardly ever	2
Once a month	3
Once a week	4
More than once a week	5

54. **How would you rate yourself as a reader?** *(Circle one only.)*

Very poor	1
Poor	2
Average	3
Good	4
Very good	5

55. **Please tell us what you think are the <u>three</u> most important <u>ways</u> to become a good reader?** *(Choose three only.)*

		Most important
a.	Liking it	2
b.	Spending lots of time reading	2
c.	Being able to concentrate well	2
d.	Knowing how to sound out words	2
e.	Learning the meaning of lots of words	2
f.	Having many good books around	2
g.	Having a lively imagination	2
h.	Having lots of reading for homework	2
i.	Having lots of drill at hard things	2
j.	Having lots of written exercises	2
k.	Being told how to do it	2
l.	Other (Specify)____________________	2

____

612

56. **How often do you do each of the following <u>before</u> you begin to read?** *(Circle one number per line.)*

	Never	Once in a while	Quite often	Most of the time
a. Look at the title, illustrations and headings to find out what it is likely to be about	1	2	3	4
b. Think about what you already know about the topic	1	2	3	4
c. Remember other selections about the same topic	1	2	3	4
d. Try to guess what will happen or what information you might learn	1	2	3	4
e. Talk to somebody else about it	1	2	3	4
f. Pinpoint issues that you are interested in exploring	1	2	3	4

57. **How often do you do each of the following while you are reading something for the first time?** *(Circle one number per line.)*

		Never	Once in a while	Quite often	Most of the time
a.	Picture in your mind what is happening	1	2	3	4
b.	Make notes	1	2	3	4
c.	Stop and think about what you have already read	1	2	3	4
d.	Read over the materials again	1	2	3	4
e.	Try to guess what will come next	1	2	3	4
f.	Imagine you are right there in the story	1	2	3	4
g.	Reflect upon the sounds and moods being suggested	1	2	3	4
h.	Reflect upon the feelings of characters	1	2	3	4
i.	Develop outlines with major and minor points	1	2	3	4
j.	Consider artwork and photos which might relate to what is written	1	2	3	4
k.	Consider graphics (charts and figures) which might relate to what is written	1	2	3	4
l.	Ask a teacher for help	1	2	3	4
m.	Ask a classmate for help	1	2	3	4
n.	Highlight important ideas	1	2	3	4
o.	Decide upon things for yourself	1	2	3	4

58. **How often do you do each of the following <u>after you have finished</u> reading?** *(Circle one number per line only.)*

		Never	Once in a while	Quite often	Most of the time
a.	Look back over what you have read	1	2	3	4
b.	Write down notes or ideas	1	2	3	4
c.	Compare what you have read with things that have happened to you, feelings you have had, or things you have seen	1	2	3	4
d.	Think about related selections you have read	1	2	3	4
e.	Get new ideas about things to read or research	1	2	3	4
f.	Talk to somebody else about it	1	2	3	4
g.	Write something of your own on that topic	1	2	3	4
h.	Refer to other books or materials	1	2	3	4
i.	Think about the theme or major point of the selection	1	2	3	4
j.	Try to figure out what the teacher wants	1	2	3	4
k.	Respond as closely as possible to the teacher's assignment	1	2	3	4
l.	Pass in your work to be checked	1	2	3	4
m.	Apply or use what you have done	1	2	3	4
n.	Develop illustrations which go along with the text	1	2	3	4

G. Questions 59 to 62 are about your school.

59. **How much time per week do you usually spend on reading silently in your English or Language Arts class?** *(Write your answer in hours and minutes; if you do not read silently in class, enter "0 hours and 0 minutes.")*

_______ Hours and _______ Minutes

60. **How often are you asked to write about what you have been reading in your English or Language Arts class?** *(Circle one only.)*

Never	1
Hardly ever	2
Sometimes	3
Most of the time	4
Always	5

61. **Which of the following do you write on your own outside of school?** *(Circle one on each line.)*

	No	Yes
Poetry	1	2
Diary	1	2
Letters	1	2
Messages	1	2
Stories	1	2
Computer programs	1	2
Other (Specify)________________ ________________________	1	2

62. **How often do you write on your own outside of school?** *(Circle only one.)*

Less than once a week	1
1 or 2 times a week	2
3 or 4 times a week	3
More than 4 times a week	4

Questions About Your Reading Activities

H. Question 63 is about school and homework.

63. **How often do you read or use books and textbooks for classes <u>in school and for homework</u>?** *(Circle only one per line.)*

		Almost never	Less than once a month	1 or 2 times a month	About once a week	2 or 3 times a week	Almost every day
a.	Stories/Literature/Fiction in English	1	2	3	4	5	6
b.	Science/Mathematics	1	2	3	4	5	6
c.	Foreign Language	1	2	3	4	5	6
d.	History/Geography/Civics/Economics	1	2	3	4	5	6
e.	Vocational-Technical/Home Economics	1	2	3	4	5	6
f.	Reference/Directory/Dictionary or Encyclopedia, in all subjects	1	2	3	4	5	6
g.	Workbooks and Exercises	1	2	3	4	5	6

I. Questions 64 to 67 are about personal interest and leisure.

64. **How often do you read books on these topics for personal interest and leisure time activity?** *(Circle only one per line.)*

		Almost never	Less than once a month	1 or 2 times a month	About once a week	2 or 3 times a week	Almost every day
a.	Mystery/Spy	1	2	3	4	5	6
b.	Romance	1	2	3	4	5	6
c.	Sports/Recreation	1	2	3	4	5	6
d.	Health	1	2	3	4	5	6
e.	Adventure	1	2	3	4	5	6
f.	Science Fiction	1	2	3	4	5	6
g.	Horror	1	2	3	4	5	6
h.	Poetry	1	2	3	4	5	6
i.	History/Politics	1	2	3	4	5	6
j.	Humor	1	2	3	4	5	6
k.	Science/Technology	1	2	3	4	5	6
l.	Travel	1	2	3	4	5	6
m.	Wildlife and Nature	1	2	3	4	5	6
n.	Music	1	2	3	4	5	6
o.	Classical Literature	1	2	3	4	5	6
p.	Biography/Autobiography	1	2	3	4	5	6
q.	Religion/Morality/Ethics	1	2	3	4	5	6
r.	Fashion	1	2	3	4	5	6
s.	Beauty	1	2	3	4	5	6
t.	Comic	1	2	3	4	5	6

65. **How often do you read these topics in magazines?** *(Circle only one per line.)*

		Almost never	Less than once a month	1 or 2 times a month	About once a week	2 or 3 times a week	Almost every day
a.	Sports/Recreation	1	2	3	4	5	6
b.	Health	1	2	3	4	5	6
c.	Movies/TV	1	2	3	4	5	6
d.	Romance	1	2	3	4	5	6
e.	Fashion	1	2	3	4	5	6
f.	News	1	2	3	4	5	6
g.	Politics	1	2	3	4	5	6
h.	Music	1	2	3	4	5	6
i.	Wildlife and Nature	1	2	3	4	5	6
j.	Computers/Technology	1	2	3	4	5	6
k.	Car/Motorcycle	1	2	3	4	5	6

66. **How often do you read these sections of newspapers?** *(Circle only one per line.)*

		Almost never	Less than once a month	1 or 2 times a month	About once a week	2 or 3 times a week	Almost every day
a.	Sports/Recreation	1	2	3	4	5	6
b.	Health	1	2	3	4	5	6
c.	Comic Strips	1	2	3	4	5	6
d.	Movies/TV	1	2	3	4	5	6
e.	News/Politics	1	2	3	4	5	6
f.	People/Romance/Fashion	1	2	3	4	5	6
g.	Classified Advertisements	1	2	3	4	5	6
h.	Business/Finance	1	2	3	4	5	6

67. **How often do you read or use these types of documents?** *(Circle only one per line.)*

		Almost never	Less than Once a month	1 or 2 times month	About once a week	2 or 3 times a week	Almost Every day
a.	Tables	1	2	3	4	5	6
b.	Charts/Graphs	1	2	3	4	5	6
c.	Maps	1	2	3	4	5	6
d.	Schedules/Timetables	1	2	3	4	5	6
e.	Directories	1	2	3	4	5	6
f.	Forms	1	2	3	4	5	6
g.	Diagrams	1	2	3	4	5	6
h.	Directions/Instructions	1	2	3	4	5	6

J. Question 68 is about your reaction to school.

68. *(Each item in this question says that school is a place where particular things happen to you or you feel in a particular way. We want to see whether you "definitely agree," "mostly agree," "mostly disagree," or "definitely disagree" with the items. Please read each item carefully and circle one number on each line which best describes how you feel.)*

School is a place where....

		Definitely disagree	Mostly disagree	Mostly agree	Definitely agree
a.	I know how to cope with the work	1	2	3	4
b.	Teachers are fair and just	1	2	3	4
c.	I really like to go	1	2	3	4
d.	Mixing with other people helps me understand myself	1	2	3	4
e.	I feel important	1	2	3	4
f.	I learn to get along with other people	1	2	3	4
g.	Teachers help me to do my best	1	2	3	4
h.	People have confidence in me	1	2	3	4
i.	Teachers treat me fairly in class	1	2	3	4
j.	People come to me for help	1	2	3	4
k.	I feel lonely	1	2	3	4
l.	I feel restless	1	2	3	4
m.	I know that people think a lot of me	1	2	3	4
n.	Teachers give me the marks I deserve	1	2	3	4
o.	People look up to me	1	2	3	4
p.	I feel depressed	1	2	3	4
q.	I know I can reach a satisfactory standard in my work	1	2	3	4
r.	I know the sorts of things I can do well	1	2	3	4
s.	I learn a lot about myself	1	2	3	4
t.	Teachers listen to what I say	1	2	3	4
u.	I feel happy	1	2	3	4
v.	I find that learning is a lot of fun	1	2	3	4
w.	I get enjoyment from being there	1	2	3	4
x.	I get satisfaction from the school work I do	1	2	3	4
y.	I get to know myself better	1	2	3	4
z.	I know I can do well enough to be successful	1	2	3	4
aa.	I get upset	1	2	3	4
bb.	I feel great	1	2	3	4
cc.	I have learned to accept other people as they are	1	2	3	4

Thank you very much for your cooperation!

Attachment A-3

Teacher Questionnaires

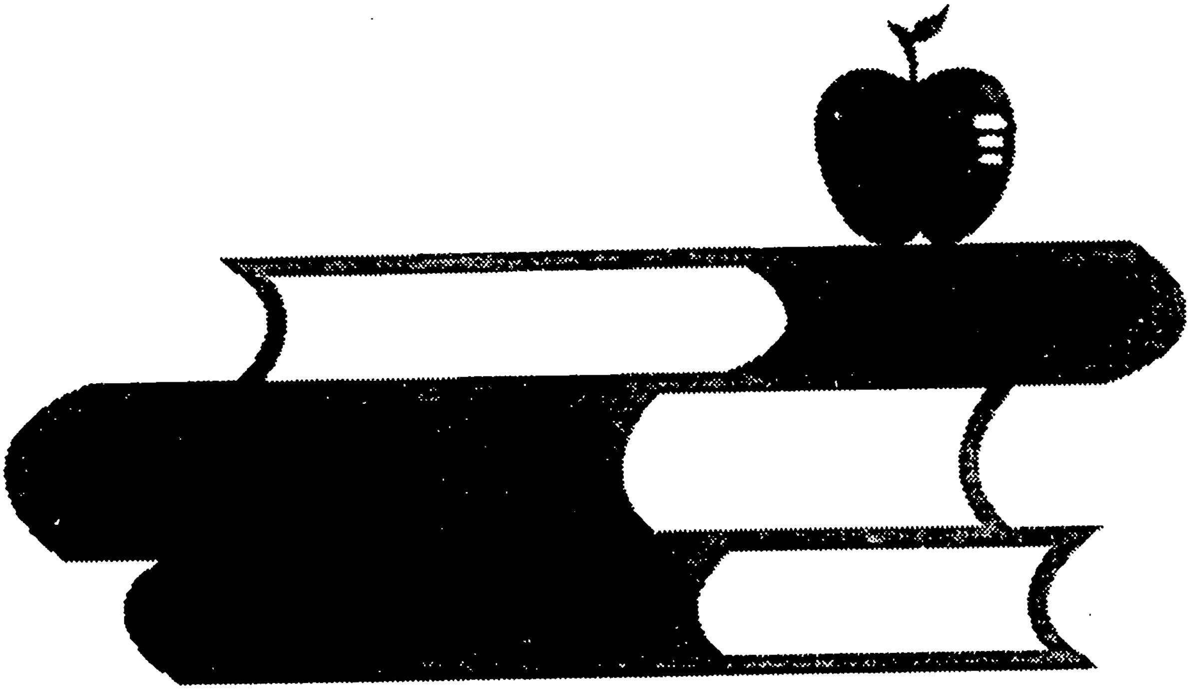
OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92

ID: ____________________

READING LITERACY

TEACHER QUESTIONNAIRE

Fourth Grade



CODE: ________

Directions:

The following questionnaire is part of an international study of Reading Literacy and attempts to identify differences in reading instruction. It is recognized that teachers around the world are likely to respond quite differently from one another.

Some questions are more relevant to particular countries. However, all teachers are asked to respond to every question so that international comparisons can be made.

Please answer all questions in such a way as to reflect most clearly your teaching practices. Most questions require you to circle your selected response. Others require you to write in a number. Where it is appropriate to enter "0" in the answer, please do so. Do not leave blanks.

We thank you for your effort.

A. Questions 1 to 17 have to do with you and your education.

1. **What is your sex?**

Male 1
Female 2

2. **What is your date of birth?**

|__|__| |__|__| |__|__|
Month Day Year

3. **What language was spoken in your home when you were a child?** *(Circle one number only.)*

English 1
Other (Specify) ____________________ 2
English and another language 3

4. **What is your ethnicity/race?** *(Circle only one.)*

Asian or Pacific Islander 1
American Indian or Alaskan Native 2
Hispanic 3
White (non-Hispanic) 4
Black (non-Hispanic) 5

5. **What is the highest level of education you have completed?** *(Circle only one.)*

High school 1
Some college 2
College graduate (Bachelor's degree) 3
Some post-baccalaureate 4
Master's degree 5
Education specialist degree 6
Doctorate 7

6. **Prior to becoming a teacher, did you attend an accredited teacher education program?**

No .... 1 **(Go to Q.9)**
Yes 2

7. **How many teacher education courses did you complete?**

________ Courses

8. **What percentage of that time was devoted to learning about the teaching of reading?** *(Circle only one.)*

0%	1
1-24%	2
25-49%	3
50-74%	4
75-100%	5

9. **What type of teacher certification do you hold?** *(Circle only one.)*

a. Regular or standard	1
b. Probationary	2
c. Temporary, provisional, or emergency	3

10. **Do you hold special certification as a reading teacher (including all types of certification)?**

No	1
Yes	2

11. **Approximately how many courses have you completed related to the teaching of reading since your initial teacher certification?** *(Circle only one.)*

None	1
One	2
Two	3
Three	4
Four or more	5

12. **How many times have you been to in-service teacher training courses in reading in the last three years?** *(Circle one number only.)*

None	1	(Go to Q.15)
Once	2	
Twice	3	
Three times	4	
Four or more times	5	

13. **What was the predominant mode of instruction used by the instructors in the last in-service teacher training course in reading you attended?** *(Circle only one.)*

Listening to presentations	1
Group discussions	2
Demonstration (or modeling) of techniques	3
Production of learning materials	4
Development of assessment materials	5
Experimenting with methods yourself	6
Other (Specify) ________________	7

14. **After your last in-service course related to reading instruction, was any followup or support provided to help you implement the methods?**

No	1
Yes	2

15. **At the end of this school year how many years will you have been teaching?**

Full-time _______ Years

Part-time _______ Years

16. **How many years have you been teaching this grade level (i.e., fourth grade)?**

Full-time _______ Years

Part-time _______ Years

17. **About how often do you read each of the following?** *(Circle one number per line only.)*

	Never or almost never	About once a year	About once a term	About once a month	About once a week or more
a. Articles on teaching	1	2	3	4	5
b. Articles on reading	1	2	3	4	5
c. Books on history or politics	1	2	3	4	5
d. Books on the arts	1	2	3	4	5
e. Books on science	1	2	3	4	5
f. Novels or short stories	1	2	3	4	5
g. Poems	1	2	3	4	5
h. Plays	1	2	3	4	5
i. Books for children	1	2	3	4	5

B. Questions 18 to 29 pertain to the class being tested.

18. **At the end of this school year, how many years will you have been teaching this group of students?**

_______ Years

19. **Is the class tested a multi-grade class?** *(Circle one number only.)*

No	1
Yes	2

20. **In this class, what is the total number of students and the total number of fourth grade students?**

_______ Total students in this class

_______ Fourth grade students in this class

21. **How many 4th grade students in this class need remedial help in reading?** *(If none, please enter "0.")*

_______ Students

22. **How many 4th grade students in this class receive remedial help in reading?** *(If none, please enter "0.")*

_______ Students

23. **How often do you meet individually with parents of the students in this class?** *(Circle one number only.)*

Never	1
About once a year	2
About once a term	3
About once a month	4
About once a week or more	5

24. **How often do you meet individually with parents of students who have difficulties in reading?** *(Circle only one.)*

Never	1
About once a year	2
About once a term	3
About once a month	4
About once a week or more	5

25. **What is the <u>total instructional time</u> (in hours and minutes), excluding breaks, for this class in a typical week?** *(For all subject areas.)*

_______ Hours **and** _______ Minutes per week

26. **How much time <u>per school week</u> is typically devoted to the teaching and learning of English (including reading, writing, speaking, literature, listening, and other language skills) for this class?**

_______ Hours **and** _______ Minutes per week

27. **How much of this time <u>per school week</u> is typically devoted to the teaching and practice of reading for this class?**

_______ Hours **and** _______ Minutes per week

28. **How many fourth grade students in <u>this</u> class do <u>not</u> speak English as their first language?** *(If none, please enter '0.')*

_______ Students

29. **How much time per school week is typically devoted to the teaching and learning of ESOL (including reading, writing, speaking, literature, listening, and other language skills) for this class?**

_______ Hours **and** _______ Minutes per week

C. Questions 30 to 53 have to do with your teaching activities.

30. How often are your students typically involved in the following reading activities? *(Circle one number per line only.)*

	Reading Activities	**Frequency**			
		Almost never	About 1 or 2 times a month	About 1 or 2 times a week	Almost every day
a.	Learning letter-sound relationships and/or phonics	1	2	3	4
b.	Word-attack skills (e.g., prediction)	1	2	3	4
c.	Silent reading in class	1	2	3	4
d.	Answering reading comprehension exercises in writing	1	2	3	4
e.	Independent silent reading in a library	1	2	3	4
f.	Listening to students reading aloud to a whole class	1	2	3	4
g.	Listening to students reading aloud to small groups or pairs	1	2	3	4
h.	Listening to teachers reading stories aloud	1	2	3	4
i.	Discussion of books read by students	1	2	3	4
j.	Learning new vocabulary systematically (e.g., from lists)	1	2	3	4
k.	Learning new vocabulary from texts	1	2	3	4
l.	Learning library skills	1	2	3	4
m.	Reading plays or dramas	1	2	3	4
n.	Playing reading games (e.g., forming sentences from jumbled words)	1	2	3	4
o.	Dramatizing stories	1	2	3	4
p.	Drawing in response to reading	1	2	3	4
q.	Orally summarizing their reading	1	2	3	4
r.	Relating experiences to reading	1	2	3	4
s.	Reading other students' writing	1	2	3	4
t.	Making predictions during reading	1	2	3	4
u.	Diagramming story content	1	2	3	4
v.	Looking for the theme or message	1	2	3	4
w.	Making generalizations and inferences	1	2	3	4
x.	Studying the style or structure of a text	1	2	3	4
y.	Comparing pictures and stories	1	2	3	4
z.	Student leading discussion about passage	1	2	3	4
aa.	Reading in other subject areas	1	2	3	4
bb.	Writing in response to reading	1	2	3	4

31. **Does every student have a book available for his or her use during the normal reading instruction period?**

No .... 1 (Go to Q.33)
Yes 2

32. **How many different textbooks are available for each student?**

________ Textbooks (Go to Q.34)

33. **How many students must share a textbook due to lack of resources?**

________ Students

34. **Please rank <u>five</u> of the following <u>aims</u> of reading instruction in order of the importance you attach to each of them.** *(Place "1" next to the most important and so on to "5" for the least important. Choose only five aims, and use each rank only once.)*

	Aims	**Importance**
a.	Developing skill in reading aloud	________
b.	Developing a lasting interest in reading	________
c.	Improving students' reading comprehension	________
d.	Developing students' research and study skills..............................	________
e.	Extending students' vocabulary	________
f.	Developing students' critical thinking...	________
g.	Expanding students' world views	________
h.	Deepening students' emotional development................................	________
i.	Improving word-attack skills	________
j.	Increasing speed of reading	________
k.	Expanding students' reading choice	________
l.	Making reading enjoyable..	________

35. **How often do you use the following instructional strategies when teaching reading?** *(Circle one number per line only.)*

	Instructional Strategies	**Frequency** Almost never	 About 1 or 2 times a month	 About 1 or 2 times a week	 Almost Every day
a.	Introduce the background of a passage before reading it	1	2	3	4
b.	Ask students to describe their strategies for understanding	1	2	3	4
c.	Encourage parents to be involved with the reading program	1	2	3	4
d.	Maintain a graded sequence of text difficulty	1	2	3	4
e.	Ask questions to assess text comprehension	1	2	3	4
f.	Ask questions to deepen understanding	1	2	3	4
g.	Show students how to understand a text	1	2	3	4
h.	Compare stories, poems, fables and tales	1	2	3	4
i.	Read aloud to students	1	2	3	4
j.	Encourage parents to read to children	1	2	3	4
k.	Encourage students to read more	1	2	3	4
l.	Encourage students to use the library more	1	2	3	4
m.	Use materials you have prepared yourself	1	2	3	4
n.	Teach about different text genres (e.g., stories, poems, fables and tales)	1	2	3	4
o.	Provide instruction by means of computer (computer-aided instruction)	1	2	3	4
p.	Peer tutoring	1	2	3	4
q.	Cooperative learning	1	2	3	4

36. **How often do you teach or encourage students to improve their comprehension by using these strategies?** *(Circle one number on each line.)*

		Frequency			
		Almost never	About 1 or 2 times a month	About 1 or 2 times a week	Almost every day
a.	Picturing in their minds what is happening as they read	1	2	3	4
b.	Trying to predict what will come next	1	2	3	4
c.	Looking back over what they have read	1	2	3	4
d.	Writing down notes or ideas about what they have read	1	2	3	4
e.	Comparing what they have read with experiences they have had	1	2	3	4
f.	Thinking about similar things they have read	1	2	3	4
g.	Talking to somebody else about what they have read	1	2	3	4
h.	Writing something of their own on what they have read	1	2	3	4

37. **Are the students in this class all at the same reading level?**

No 1
Yes 2

38. **Do you divide the students in this class into groups for reading instruction?**

No.......... 1 **(Go to Q.41)**
Yes 2

39. **How often do you use each of these types of groupings?** *(Circle one number per line only.)*

	Frequency			
	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
Age groups	1	2	3	4
Ability groups	1	2	3	4
Interest groups	1	2	3	4
Other (Please specify)	1	2	3	4

____

633

40. **How many groups do you typically form?**

_______ Groups

41. **This year how frequently did you teach your class to read each of the following kinds of text?** *(Circle one number per line only.)*

	Frequency				
	Almost never	3 or 4 times a year	About once a month	At least once a week	Nearly every day
a. **Narration:** Texts that tell a story or give the order in which things happen	1	2	3	4	5
b. **Exposition:** Texts that describe things or people or explain how things work, why things happen, or persuasive arguments	1	2	3	4	5
c. **Documents:** Tables, charts, diagrams, lists, maps	1	2	3	4	5

42. **What percentage of classroom time is devoted to teaching each of the following kinds of text?**

	Percent
a. Narration	_______
b. Exposition (including arguments)	_______
c. Documents	_______
	100%

43. **Below you will find a number of statements about issues in reading instruction. Please state your degree of agreement/disagreement with each statement by circling the appropriate number.** *(Circle one number on each line.)*

		Strongly disagree	Disagree	Uncertain	Agree	Strongly agree
a.	When my students read to me, I expect them to read every word accurately	1	2	3	4	5
b.	Teachers should keep careful records of every student's reading progress	1	2	3	4	5
c.	Students should not be encouraged to read a word they don't know	1	2	3	4	5
d.	All students should enjoy reading	1	2	3	4	5
e.	Most of what a student reads should be assessed	1	2	3	4	5
f.	Every day students should be read to by the teacher from a story book	1	2	3	4	5
g.	Reading aloud by students to a class is a waste of time	1	2	3	4	5
h.	Most students improve their reading best by extensive reading on their own	1	2	3	4	5
i.	Students should always understand why they are reading	1	2	3	4	5
j.	Teachers should always group students according to their reading ability	1	2	3	4	5
k.	9-year-olds should not have access to books they will read in the next year at school	1	2	3	4	5
l.	Class sets of graded reading material should be used as the basis for the reading program	1	2	3	4	5
m.	Students who can't understand what they read haven't been taught proper comprehension skills	1	2	3	4	5
n.	Every mistake a student makes in reading aloud should be corrected at once	1	2	3	4	5
o.	All students' comprehension assignments should be marked carefully to provide them with feedback	1	2	3	4	5
p.	Students should not start a new book until they have finished the last	1	2	3	4	5
q.	Parents should be actively encouraged to help their students with reading	1	2	3	4	5
r.	Students should learn most of their new words from lessons designed to enhance their vocabulary	1	2	3	4	5
s.	Reading learning materials should be carefully sequenced in terms of language structures and vocabulary	1	2	3	4	5
t.	Students should take a book home to read every day	1	2	3	4	5

43. (Continued)

		Strongly disagree	Disagree	Uncertain	Agree	Strongly agree
u.	Students should be encouraged to read texts they have written	1	2	3	4	5
v.	Students should always understand what they are reading	1	2	3	4	5
w.	Students should always choose their own books to read	1	2	3	4	5
x.	A word recognition test is sufficient for assessing students' reading levels	1	2	3	4	5
y.	Teachers should carefully follow the sequence of the textbook	1	2	3	4	5
z.	Students should undertake research projects to improve their reading	1	2	3	4	5

44. **Do you regularly (i.e., at least once a week) do the following activities to encourage your students to read outside school?** *(Circle one number per line only.)*

		No	Yes
a.	Suggest books (to students) to read	1	2
b.	Suggest newspaper articles to students to read	1	2
c.	Read stories to students	1	2
d.	Hold discussions about books	1	2
e.	Encourage students to borrow library books	1	2
f.	Other	1	2

45. How often do you use the following methods/materials to discover your students' needs in reading? *(Circle one number per line only.)*

		Never	About once a year	About once a term	About once a month	About once a week or more
a.	Listening to students' reading.....	1	2	3	4	5
b.	Teacher-made vocabulary tests	1	2	3	4	5
c.	Exercises in workbooks and textbooks	1	2	3	4	5
d.	Standardized commercial reading tests	1	2	3	4	5
e.	Knowledge of students' reading interests..........................	1	2	3	4	5
f.	Comments from other teachers	1	2	3	4	5
g.	Informal observation....................	1	2	3	4	5
h.	Interviews....................................	1	2	3	4	5
i.	Tests in workbooks and textbooks	1	2	3	4	5
j.	Students' writing	1	2	3	4	5

46. How often do you assess these aspects of reading with all or most of your class? *(Circle one number per line only.)*

		Never	About once a year	About once a term	About once a month	About once a week or more
a.	Word recognition.........................	1	2	3	4	5
b.	Vocabulary..................................	1	2	3	4	5
c.	Text comprehension....................	1	2	3	4	5
d.	Literary appreciation....................	1	2	3	4	5
e.	Use of background knowledge...	1	2	3	4	5
f.	Sentence understanding.............	1	2	3	4	5
g.	Phonic skills................................	1	2	3	4	5
h.	Reading study skills.....................	1	2	3	4	5
i.	Amount of reading.......................	1	2	3	4	5
j.	Decoding	1	2	3	4	5

47. **How often do you use these assessment methods?** *(Circle one number per line only.)*

		Never	About once a year	About once a term	About once a month	About once a week or more
a.	Multiple-choice questions on material read	1	2	3	4	5
b.	Listening to students reading aloud	1	2	3	4	5
c.	Records of student interests	1	2	3	4	5
d.	Oral discussions	1	2	3	4	5
e.	Oral questions on material read	1	2	3	4	5
f.	Written open-ended questions on material read	1	2	3	4	5
g.	Student-teacher interviews	1	2	3	4	5

48. **Do you assign homework in <u>reading</u> to this class?**

No	1	(Go to Q.51)
Yes	2	

49. **How often do you assign reading homework to this class?** *(Circle one number only.)*

Less than once a week	1
1 or 2 times a week	2
3 or 4 times a week	3
More than 4 times a week	4

50. **How much time do you expect an average student to spend on <u>assigned homework</u> in reading each week?**

_______ Hours and _______ Minutes per week

51. **How often do you assign homework that requires reading in other subject areas ?** *(Circle only one.)*

Never	1
Less than once a week	2
1 or 2 times a week	3
3 or 4 times a week	4
More than 4 times a week	5

52. **How much time do you expect an average student to spend on assigned homework that requires reading in other subject areas each week?**

________ Hours **and** ________ Minutes per week

53. **How often are the following teaching practices reflected in your class?** *(Circle only one number per line.)*

			Frequency		
	Never	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
a. Students are assigned specific topics to study	1	2	3	4	5
b. Students are told how what they know relates to a topic	1	2	3	4	5
c. Students are informed as to the purposes of lessons	1	2	3	4	5
d. Students receive feedback from the teacher on their ideas	1	2	3	4	5
e. Students are directed to proceed based upon set guidelines	1	2	3	4	5
f. Students deal with issues and topics related to their own experiences	1	2	3	4	5
g. Students establish their own purposes and goals	1	2	3	4	5
h. Students have a choice in what they will do	1	2	3	4	5
i. Students decide how they will approach their texts	1	2	3	4	5
j. Students share their ideas with each other	1	2	3	4	5
k. Students are directed to answer a set of the teacher's questions	1	2	3	4	5
l. Students are given feedback by the teacher on the themes or main ideas of the selections they read	1	2	3	4	5
m. Students are given the opportunity to discuss various possible themes for the selection	1	2	3	4	5
n. Spontaneous student responses are discouraged	1	2	3	4	5
o. Students are encouraged to compare their written texts with other students' written texts	1	2	3	4	5

53. (Continued)

		Frequency				
		Never	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
p.	Students are encouraged to compare their written texts with the reading selection	1	2	3	4	5
q.	Students are given guided practice with skills	1	2	3	4	5
r.	Students are invited to consider how skills apply to what they have written	1	2	3	4	5
s.	Students are encouraged to work independently on classwork	1	2	3	4	5
t.	Spontaneous student responses are encouraged	1	2	3	4	5
u.	Students are encouraged to use the reading selection as a source for ideas when writing their texts	1	2	3	4	5
v.	Students are told what they have learned and have yet to learn	1	2	3	4	5
w.	Students are given the opportunity to consider what they think they have learned, as well as their perception of their strengths and weaknesses..........	1	2	3	4	5
x.	Students are given the opportunity to assess their own progress	1	2	3	4	5
y.	Students are given the opportunity to provide input on how they will be assessed..................	1	2	3	4	5
z.	Specific skills are taught at certain times	1	2	3	4	5
aa.	Students are given teacher feedback on how they compare with other students.............	1	2	3	4	5
bb.	Students are expected to follow the activities outlined in the lesson the teacher has planned	1	2	3	4	5
cc.	Student needs necessitate changes to the lesson	1	2	3	4	5
dd.	Students are given the opportunity to work on a variety of different projects.........................	1	2	3	4	5

D. Questions 54 to 58 have to do with your classroom library.

54. Do you have a classroom library (i.e., a small book or magazine corner in your classroom)?

No .... 1 **(Go to Q.59)**
Yes 2

55. About how many books with different titles are in it?

________ Books with different titles

56. About how many different titles of magazines/newspapers do you have in it?

________ Different titles of magazines/newspapers

57. Can your students borrow books from the classroom library to take home? *(Circle one number only.)*

No 1
Yes 2

58. When do students use the classroom library? *(Circle only one.)*

Once a week as prescribed by the teacher 1
Several times a week as prescribed by the teacher 2
Whenever the students have free time 3
Whenever they like 4

E. Questions 59 to 62 have to do with your school library.

59. Is there a school library in your school? *(Circle one number only.)*

No	1	**(Go to Q.63)**
Yes	2	

60. How often does this class visit the school library? *(Circle one number only.)*

Less than once a month	1
1 or 2 times a month	2
3 or 4 times a month	3
5 or more times a month	4

61. Can your students borrow books from the school library to take home? *(Circle one number only.)*

No	1
Yes	2

62. When do students use the school library? *(Circle only one.)*

Once a week as prescribed by the teacher	1
Several times a week as prescribed by the teacher	2
Whenever the students have free time	3
Whenever they like	4

F. Questions 63 to 67 have to do with school organization.

63. Is your work as a teacher evaluated by the school principal (or deputy school principal)?

No	1
Yes	2

64. **Does the school principal (or deputy principal) ...** *(Circle one number per line only.)*

	No	Yes
a. discuss explicit achievement standards for the subject that you teach?	1	2
b. ask for evaluation results or progress of your students in reading?	1	2
c. make suggestions about the choice of instructional methods in reading?	1	2
d. encourage contacts among teachers?	1	2
e. initiate activities directed at the professional development of teachers?	1	2
f. make suggestions about the content that must be covered in reading?	1	2

65. **Do you have staff meetings at your school?**

No .... 1 (Stop here)
Yes 2

66. **How often do you have staff meetings at your school?** *(Circle one only.)*

Less than once a year 1
Once a year 2
Once a term 3
Monthly 4
Weekly 5

67. **How often do the following items occur as subjects of discussion during staff meetings?** *(Circle one number per line only.)*

	All staff meetings	Most staff meetings	Some staff meetings	Not in any staff meetings
a. Curriculum content	1	2	3	4
b. The way the subject matter is presented	1	2	3	4
c. Professional development of teachers	1	2	3	4
d. Issues of student welfare and guidance	1	2	3	4
e. Organizational issues (e.g., school climate, coordination of work among teachers, the way decision-making procedures are conducted)	1	2	3	4
f. Other topics (e.g., purely administrative tasks, leisure and social activities)	1	2	3	4

Thank you very much for your cooperation!

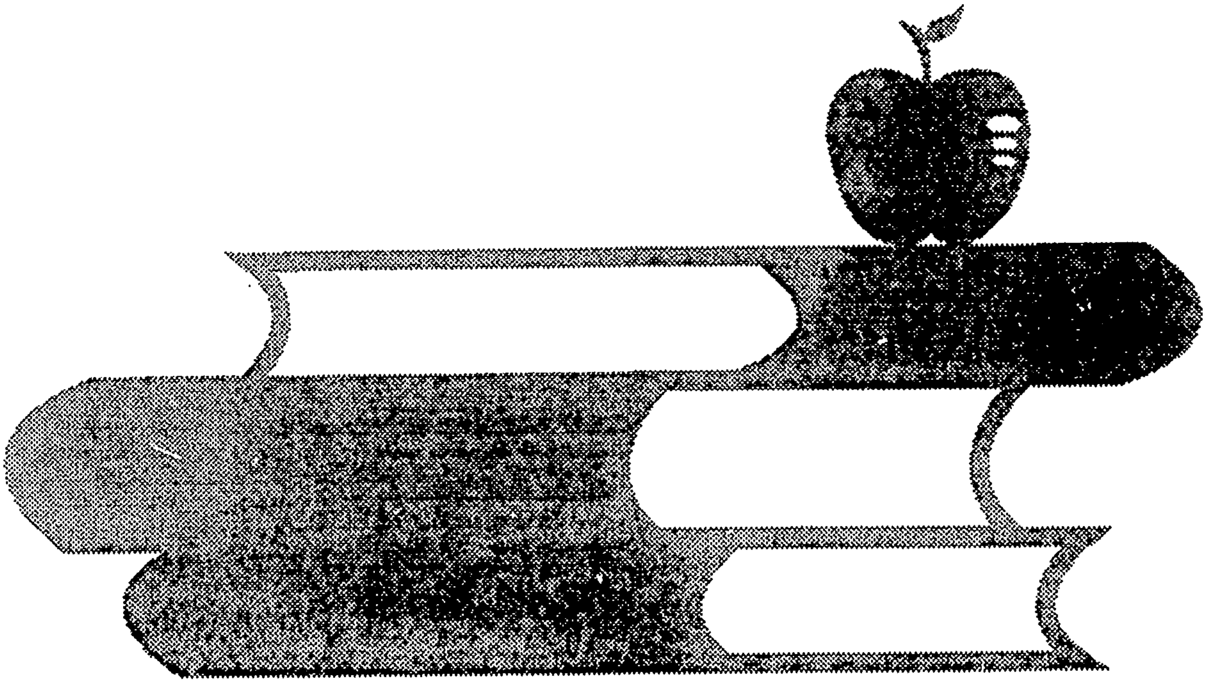
OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92

ID: ____________________

READING LITERACY

TEACHER QUESTIONNAIRE

Ninth Grade



CODE: ________

Directions:

The following questionnaire is part of an international study of Reading Literacy and attempts to identify differences in reading instruction. It is recognized that teachers around the world are likely to respond quite differently from one another.

Some questions are more relevant to particular countries. However, all teachers are asked to respond to every question so that international comparisons can be made.

Please answer all questions in such a way as to reflect most clearly your teaching practices. Most questions require you to circle your selected response. Others require you to write in a number. Where it is appropriate to enter "0" in the answer, please do so. Do not leave blanks.

We thank you for your effort.

Public reporting burden for this collection of information is estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the U.S. Department of Education, Information Management and Compliance Division, Washington, D.C. 20202-4651; and to the Office of Management and Budget, Paperwork Reduction Project 1850-0645, Washington, D.C. 20503.

A. Questions 1 to 17 have to do with you and your education.

1. **What is your sex?** *(Circle one number only.)*

Male	1
Female	2

2. **What is your date of birth?**

|__|__| |__|__| |__|__|
Month Day Year

3. **What language was spoken in your home when you were a child?** *(Circle one number only.)*

English	1
Other (Specify) ______________	2
English and another language	3

4. **What is your ethnicity/race?** *(Circle only one.)*

Asian or Pacific Islander	1
American Indian or Alaskan Native	2
Hispanic	3
White (non-Hispanic)	4
Black (non-Hispanic)	5

5. **What is the highest level of education you have completed?** *(Circle only one.)*

High school	1
Some college	2
College graduate (Bachelor's degree)	3
Some post-baccalaureate	4
Master's degree	5
Education specialist degree	6
Doctorate	7

6. **Prior to becoming a teacher, did you attend an accredited teacher education program?**

No	1	**(Go to Q.9)**
Yes	2	

7. **How many teacher education courses did you complete?**

________ Courses

8. **What percentage of your total teacher education was devoted to learning about the teaching of English/Language Arts/Reading?** *(Circle only one.)*

0% 1
1-24% 2
25-49% 3
50-74% 4
75-100% 5

9. **What type of teacher certification do you hold?** *(Circle only one.)*

a. Regular or standard 1
b. Probationary 2
c. Temporary, provisional, or emergency ... 3

10. **Do you hold special certification as an English/Language Arts/Reading teacher** *(including all types of certification)*?

No 1
Yes 2

11. **How many courses did you complete related to the teaching of English/Language Arts/Reading since your initial teacher certification?** *(Circle only one.)*

None 1
One 2
Two 3
Three 4
Four or more 5

12. **How many times have you been to in-service teacher training courses in English/Language Arts/Reading in the last three years?** *(Circle only one.)*

Not at all .... 1 (Go to Q.15)
Once 2
Twice 3
Three times 4
Four or more times 5

13. What was the predominant mode of instruction used by the instructors in the last in-service teacher training course in English/Language Arts/Reading you attended? *(Circle only one.)*

Listening to presentations	1
Group discussions	2
Demonstrations (or modeling) of techniques	3
Production of learning materials	4
Development of assessment materials	5
Experimenting with methods yourself	6
Other (Specify) ______________	7

14. After your last in-service course related to English/Language Arts/Reading instruction, was any followup or support provided to help you implement the methods?

No	1
Yes	2

15. At the end of this school year, how many years will you have been teaching?

Full-time _______ Years

Part-time _______ Years

16. How many years have you been teaching English/Language Arts/Reading?

Full-time _______ Years

Part-time _______ Years

17. About how often do you read each of the following? *(Circle one number per line only.)*

		Never or almost never	About once a year	About once a term	About once a month	About once a week or more
a.	Articles on teaching	1	2	3	4	5
b.	Articles on reading comprehension	1	2	3	4	5
c.	Books on history or politics	1	2	3	4	5
d.	Books on the arts	1	2	3	4	5
e.	Books on science	1	2	3	4	5
f.	Novels or short stories	1	2	3	4	5
g.	Poems	1	2	3	4	5
h.	Plays	1	2	3	4	5
i.	Articles on literature	1	2	3	4	5

B. Questions 18 to 22 pertain to the class being tested.

18. How many students are enrolled in this class?

_______ Students

19. How many students in this class do not speak English as their first language? *(If none, please enter "0.")*

_______ Students

20. How many students in this class need remedial help in reading? *(If none, please enter "0.")*

_______ Students

21. How many students in this class receive remedial help in reading? *(If none, please enter "0.")*

_______ Students

22. What is the total instructional time (in hours and minutes), excluding breaks, for this class in a typical week? *(For all subject areas.)*

_______ Hours **and** _______ Minutes per week

C. Questions 23 to 32 have to do with your teaching activities.

23. **How much time per school week is devoted to the teaching and learning of English/Language Arts/Reading for this class?**

_______ Hours and _______ Minutes per week

24. **What assessment methods do you use most often in this English/Language Arts/Reading class?** *(Rank order by assigning a "1" to the most frequent, "2" to the next, and "7" to the least frequent.)*

		Frequency
a.	Teacher quizzes	_______
b.	Multiple-choice questions	_______
c.	Records of student interests	_______
d.	Oral discussions	_______
e.	Oral discussions on material read	_______
f.	Written open-ended questions on material read	_______
g.	Essays in response to literature	_______

25. **How frequently do you teach students in your class how to read each of the following kinds of writing?** *(Circle one number per line only.)*

		Frequency				
		Almost never	3 or 4 times a year	About once a month	At least once a week	Nearly every day
a.	**Narration:** Texts that tell a story or give the order in which things happen	1	2	3	4	5
b.	**Exposition:** Texts that describe things or people or explain how things work, why things happen, or persuasive arguments	1	2	3	4	5
c.	**Documents:** Tables, charts, diagrams, lists, maps	1	2	3	4	5

26. **What percentage of classroom time is devoted to teaching each of the following kinds of text?**

		Percent
a.	Narration	______
b.	Exposition (including arguments)	______
c.	Documents	______
		100%

27. **What do you regularly (i.e., at least once a week) do to encourage your students to read outside school?** *(Choose the two you do most frequently.)*

a.	Hold discussions about books	2
b.	Suggest titles/authors..............................	2
c.	Encourage them to borrow books from the school library	2
d.	Give special reading assignments............	2
e.	Other *(Please specify)*	2

____

28. **How often are your students typically involved in the following activities?** *(Circle one number per line only.)*

	Activities	**Frequency**			
		Almost never	About 1 or 2 times a month	About 1 or 2 times a week	Almost every day
a.	Silent reading in class	1	2	3	4
b.	Answering text comprehension questions in writing..	1	2	3	4
c.	Independent silent reading in a library	1	2	3	4
d.	Listening to students reading aloud	1	2	3	4
e.	Discussion of books	1	2	3	4
f.	Learning new vocabulary systematically (e.g., from lists)..	1	2	3	4
g.	Learning new vocabulary from texts	1	2	3	4
h.	Learning library skills.......................................	1	2	3	4
i.	Reading plays or dramas	1	2	3	4
j.	Summarizing their reading...............................	1	2	3	4
k.	Relating experiences to reading	1	2	3	4
l.	Reading other students' writing	1	2	3	4
m.	Studying the style or structure of a text...........	1	2	3	4
n.	Reading in other subject areas	1	2	3	4
o.	Writing in response to reading.........................	1	2	3	4
p.	Participating in a discussion of texts led by students	1	2	3	4
q.	Learning to use illustrations (graphs, diagrams, tables) to understand text...............	1	2	3	4

29. **How often do you assign <u>homework</u> that requires reading or writing to this class?** *(Circle only one.)*

Never	1
Less than once a week	2
1 or 2 times a week	3
3 or 4 times a week	4
More than 4 times a week	5

30. **Please rank <u>five</u> of the following <u>aims</u> of reading instruction in order of the importance you attach to each of them.** *(Place "1" next to the most important and so on to "5" for the least important. Choose only five aims, and use each rank only once.)*

	Aims	Importance
a.	Developing a lasting interest in reading	________
b.	Improving students' reading comprehension strategies	________
c.	Developing students' research and study skills	________
d.	Extending students' vocabulary	________
e.	Developing students' critical thinking	________
f.	Expanding students' world views	________
g.	Increasing speed of reading	________
h.	Expanding students' variety of reading choice	________
i.	Teaching students how to apply study strategies to other subjects	________
j.	Increasing students' appreciation of literature	________
k.	Teaching students how to interpret diagrams and graphs	________

31. **How often do you teach or encourage students to improve their comprehension by using these strategies?** *(Circle one number on each line.)*

		Never	Once in a while	Quite often	Most of the time
a.	Picturing in their minds what is happening as they read	1	2	3	4
b.	Trying to predict what will come next	1	2	3	4
c.	Looking back over what they have read	1	2	3	4
d.	Writing down notes or ideas about what they have read	1	2	3	4
e.	Comparing what they have read with experiences they have had	1	2	3	4
f.	Thinking about similar things they have read	1	2	3	4
g.	Talking to somebody else about what they have read	1	2	3	4
h.	Writing something of their own on what they have read	1	2	3	4

32. **How often are the following teaching practices reflected in your class?** *(Circle only one number per line.)*

		Frequency				
		Never	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
a.	Students are assigned specific topics to study	1	2	3	4	5
b.	Students are told how what they know relates to a topic........	1	2	3	4	5
c.	Students are informed as to the purposes of lessons.............	1	2	3	4	5
d.	Students receive feedback from the teacher on their ideas..	1	2	3	4	5
e.	Students are directed to proceed based upon set guidelines....................................	1	2	3	4	5
f.	Students deal with issues and topics related to their own experiences	1	2	3	4	5
g.	Students establish their own purposes and goals....................	1	2	3	4	5
h.	Students have a choice in what they will do.........................	1	2	3	4	5
i.	Students decide how they will approach their texts	1	2	3	4	5
j.	Students share their ideas with each other....................................	1	2	3	4	5
k.	Students are directed to answer a set of the teacher's questions	1	2	3	4	5
l.	Students are given feedback by the teacher on the themes or main ideas of the selections they read	1	2	3	4	5
m.	Students are given the opportunity to discuss various possible themes for the selection......................................	1	2	3	4	5
n.	Spontaneous student responses are discouraged..........................	1	2	3	4	5
o.	Students are encouraged to compare their written texts with other students' written texts...	1	2	3	4	5

32. (Continued)

		Frequency				
		Never	Less than once a week	1 or 2 times a week	3 or 4 times a week	More than 4 times a week
p.	Students are encouraged to compare their written texts with the reading selection	1	2	3	4	5
q.	Students are given guided practice with skills	1	2	3	4	5
r.	Students are invited to consider how skills apply to what they have written	1	2	3	4	5
s.	Students are encouraged to work independently on classwork	1	2	3	4	5
t.	Spontaneous student responses are encouraged	1	2	3	4	5
u.	Students are encouraged to use the reading selection as a source for ideas when writing their texts	1	2	3	4	5
v.	Students are told what they have learned and have yet to learn	1	2	3	4	5
w.	Students are given the opportunity to consider what they think they have learned, as well as their perception of their strengths and weaknesses	1	2	3	4	5
x.	Students are given the opportunity to assess their own progress	1	2	3	4	5
y.	Students are given the opportunity to provide input on how they will be assessed	1	2	3	4	5
z.	Specific skills are taught at certain times	1	2	3	4	5
aa.	Students are given teacher feedback on how they compare with other students	1	2	3	4	5
bb.	Students are expected to follow the activities outlined in the lesson the teacher has planned	1	2	3	4	5
cc.	Student needs necessitate changes to the lesson	1	2	3	4	5
dd.	Students are given the opportunity to work on a variety of different projects	1	2	3	4	5

D. Questions 33 to 39 have to do with your school library and class resources.

33. During the instructional period associated with English/Language Arts/Reading, does every student have a book available for his or her use?

No	1	**(Go to Q.35)**
Yes	2	

34. How many different textbooks are available for each student to use during the instructional period associated with reading?

________ Textbooks **(Go to Q.36)**

35. How many students must share a textbook due to lack of resources?

________ Students

36. Is there a school library in your school? *(Circle one number only.)*

No	1	**(Go to Q.40)**
Yes	2	

37. How often does this class visit the school library? *(Circle one number only.)*

Hardly ever	1
Once a month	2
Once a week	3
More than once a week	4

38. Can your students borrow books from the school library to take home? *(Circle one number only.)*

No	1
Yes	2

39. When do students use the school library? *(Circle only one.)*

Once a week as prescribed by the teacher	1
Several times a week as prescribed by the teacher	2
Whenever the students have free time	3
Whenever they like	4

E. Questions 40 to 44 have to do with school organization.

40. **Is your work as a teacher evaluated by the school principal (or deputy school principal)?**

No	1
Yes	2

41. **Does the school principal (or deputy principal)** ... *(Circle one number per line only.)*

	No	Yes
a. discuss explicit achievement standards for the subject that you teach?	1	2
b. ask for evaluation results or progress of your students in reading?	1	2
c. make suggestions about the choice of instructional methods in reading?	1	2
d. encourage contacts among teachers?	1	2
e. initiate activities directed at the professional development of teachers?	1	2
f. make suggestions about the content that must be covered in reading?	1	2

42. Do you have **staff meetings** at your school?

No	1	**(Stop here)**
Yes	2	

43. How often do you have **staff meetings** at your school? *(Circle one only.)*

Less than once a year	1
Once a year	2
Once a term	3
Monthly	4
Weekly	5

44. **If you have staff meetings, please indicate how often the following items occur as subjects of discussion during staff meetings.**

	Not in any staff meetings	Some staff meetings	Most staff meetings	All staff meetings	If "all staff meetings," what percentage of time is typically devoted to this item? %
a. Curriculum content	1	2	3	4	______
b. The way the subject matter is presented..	1	2	3	4	______
c. Professional development of teachers.....	1	2	3	4	______
d. Issues of student welfare and guidance..	1	2	3	4	______
e. Organizational issues (e.g., school climate, coordination of work among teachers, the way decision-making procedures are conducted).....................	1	2	3	4	______
f. Other topics (e.g., purely administrative tasks, leisure and social activities).................................	1	2	3	4	______

Thank you very much for your cooperation!

657

Attachment A-4

School Questionnaires

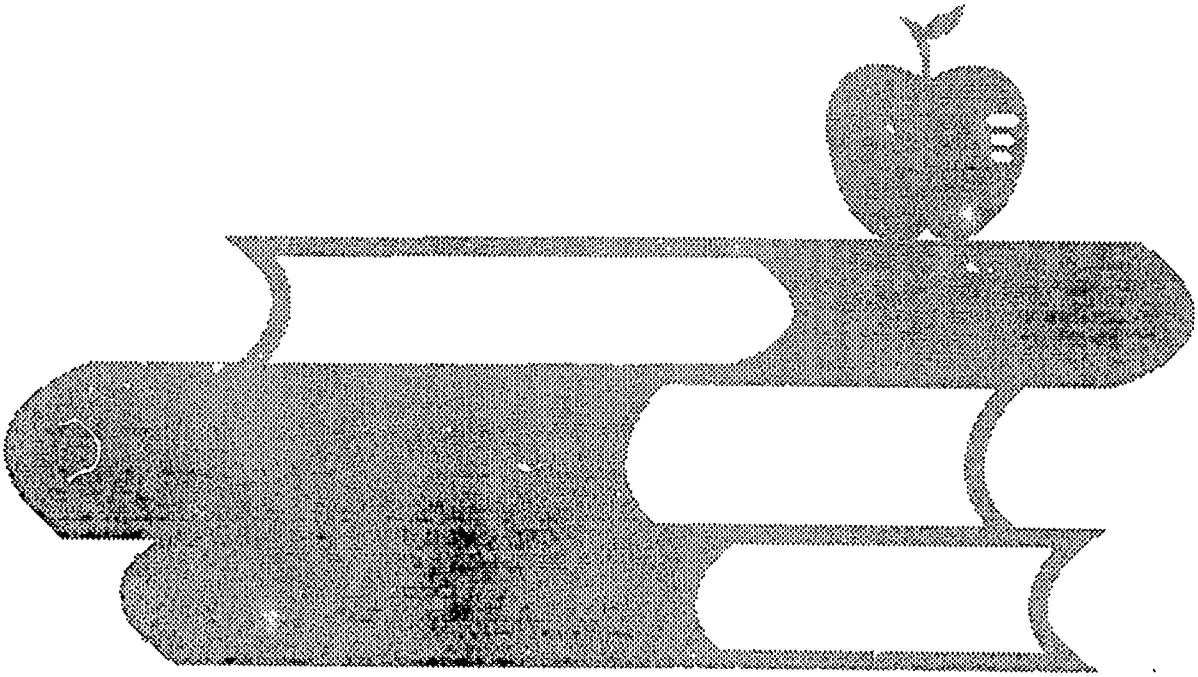
OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92

ID: ____________________

READING LITERACY

SCHOOL QUESTIONNAIRE

Fourth Grade



Directions:

The following questionnaire is part of an international study of Reading Literacy. The questions asked attempt to gather information which captures the wide range of experience and practice which is likely to exist across countries.

Please answer all questions in such a way as to reflect most accurately the situation in your school. It is important that all questions are answered. Most questions require you to circle your selected response. Others require you to write in a number. Where it is appropriate to enter "0" in the answer, please do so. Do not leave it blank. All information will be treated in the strictest confidence.

Public reporting burden for this collection of information is estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the U.S. Department of Education, Information Management and Compliance Division, Washington, D.C. 20202-4651; and to the Office of Management and Budget, Paperwork Reduction Project 1850-0645, Washington, D.C. 20503.

A. Questions 1 to 8 have to do with your school.

1. **What is the total enrollment of full-time students in your school?** *(If there are no boys or no girls, please enter "0.")*

_______ Number of boys
_______ Number of girls

2. **What is the total enrollment of full-time fourth grade students in your school?** *(If there are no boys or no girls, please enter "0.")*

_______ Number of boys
_______ Number of girls

3. **<u>How many</u> of the full-time fourth grade students in your school are:**

a. Asian or Pacific Islander, ... _______ Students
b. American Indian or Alaskan Native, ... _______ Students
c. Hispanic, ... _______ Students
d. White (non-Hispanic), or .... _______ Students
e. Black (non-Hispanic)? .... _______ Students

4. **Is your school ...** *(Circle only one.)*

A public school, or	1
A private school?	2

5. **Which of the following best describes the community in which this school is located?** *(Circle only one.)*

A rural or farming community	1
A small city or town of fewer than 50,000 people that is not a suburb of a large city	2
A medium-sized city (50,000 to 100,000 people)	3
A suburb of a medium-sized city	4
A large city (100,000 to 500,000 people)	5
A suburb of a large city	6
A very large city (over 500,000 people)	7
A suburb of a very large city	8
A military base or station	9
An Indian reservation	10

6. **Please indicate the availability of the following resources in relation to your school?** *(Circle one number on each line.)*

	Not readily available	Available in neighboring town or city (less than 2 hours of normal one way travel time)	Available locally (within 30 minutes of normal one way travel time)
Public library	1	2	3
Bookstore/book department in a store	1	2	3
Secondary level school	1	2	3
A higher education institution	1	2	3
Museum	1	2	3

7. **What is the degree of parent cooperation with the school in terms of support for the school's educational principles or goals (compared with other schools you know)?** *(Circle one only.)*

Much below average	1
Below average	2
Average	3
Above average	4
Much above average	5

8. **Which of the following resources and activities are there in your school?** *(Circle one number on each line.)*

	No	Yes
School library	1	2
Reading room for students	1	2
Student/school newspaper or magazine	1	2
Teacher (Professional) library	1	2

B. Questions 9 to 11 are about the school library. If you have indicated in Question 8 that your school does not have a library, please go straight to Question 12.

9. **Approximately how many books with <u>different titles</u> does your school library contain?** *(Exclude magazines and periodicals.)*

_______ Books with different titles

10. **Approximately how many books with <u>different titles</u> were <u>added</u> to your school library in the last year?** *(Exclude magazines and periodicals.)*

________ Books with different titles

11. **Can fourth grade students in your school borrow books from the school library to take home?** *(Circle one only.)*

No... 1
Yes 2

C. Questions 12 and 13 pertain to the number of teachers in your school.

12. **<u>How many</u> of the full-time (or full-time equivalent) teachers in your school are:**

a. Asian or Pacific Islander,.................................. ________ Teachers
b. American Indian or Alaskan Native,.................. ________ Teachers
c. Hispanic,.. ________ Teachers
d. White (non-Hispanic), or.................................. ________ Teachers
e. Black (non-Hispanic)?...................................... ________ Teachers

13. **For each of the following categories, how many full-time (or full-time equivalent) employees are there in your school?** *(If there are no male or no female teachers, please enter "0.")*

	Number of males	Number of females
a. Regular classroom teachers (exclude special education)	________	________
b. Special education teachers	________	________
c. Guidance counselors	________	________
d. Librarians and other professional media staff	________	________
e. Reading specialists	________	________
f. Other professional staff (other curriculum specialists, administrative and business staff, social workers)	________	________
g. Teacher aides (paraprofessionals who assist teachers)	________	________

D. Questions 14 to 28 pertain to instruction in your school.

14. **What is the total instructional time (in hours and minutes), excluding breaks, in a typical week in your school for all subject areas?**

_______ Hours and _______ Minutes per week

15. **How many days per year is your school scheduled to be open?**

_______ Days per year

16. **How many days of instruction were lost in the last school year due to accidents, snow days, floods, strikes, festivals, staff days, etc.?** *(If there were no days lost, please enter "0.")*

_______ Days lost last year

17. **On an average day, approximately what percentage of students are absent from school?**

_______ Percent

18. **What type of standardized tests of reading achievement does your district administer to students?** *(Circle one only.)*

Our district does not administer standardized tests of reading achievement ..	1
Norm-referenced	2
Criterion-referenced	3
Both norm- and criterion-referenced	4

19. **Do you use the results of students' standardized tests to evaluate:** *(Circle one per line.)*

	No	Yes
Student progress?	1	2
Curriculum?	1	2
Teachers?	1	2
Textbooks/materials?	1	2
Special programs?	1	2

20. **Rate your satisfaction with the following sources of evidence of your students' progress.** *(Circle one number on each line.)*

	Highly satisfied				Highly dissatisfied
Norm-referenced test scores...............	1	2	3	4	5
Criterion-referenced test scores..........	1	2	3	4	5
Student work sample...........................	1	2	3	4	5
Teachers' judgment............................	1	2	3	4	5
Grade report.......................................	1	2	3	4	5

21. **Please describe the special reading programs your school has outside the normal classroom activities (e.g., enrichment programs or special programs for the disadvantaged)?**

22. **Does your school sponsor any informal initiatives to encourage reading (e.g., book clubs)?**

No..	1
Yes	2

23. **Does your school have a program for the improvement of reading instruction (teaching and learning of reading)?**

No..	1
Yes	2

If "Yes," please describe

665

24. **Do you have any serious problems in <u>providing for the teaching/learning of reading</u> in your school?**

No	1	(Go to Q.26)
Yes	2	

25. **Rank the following seven problems in <u>providing for the teaching and learning of reading</u> in your school in order of severity.** *(Please assign "1" to the most serious and "7" to the least serious problem).*

		Rank of severity
a.	Insufficient specialized staff	________
b.	Insufficient time	________
c.	Students' lack of interest	________
d.	Insufficient funds and classroom material	________
e.	Insufficient parental support	________
f.	Insufficient number of teacher aides	________
g.	Other (Specify)________________	________

26. **Please rank the following activities in order of importance in your work as a school principal.** *("1" is the most important activity, "8" is the least important activity, "NA" = not applicable. Do not assign equal rankings.)*

		Rank of importance
a.	Representing the school at official meetings	______
b.	Evaluation of staff	______
c.	Contacts with local community (e.g., parents, community organizations, local industry)	______
d.	Discussing educational objectives with the teaching staff	______
e.	Administrative tasks concerning the functioning of the school (e.g., regulations, disciplinary duties, school budget, timetable)	______
f.	Using records of students' progress	______
g.	Taking care of issues of student welfare and guidance	______
h.	Activities aimed at the professional development of teachers	______

27. **How often do you systematically evaluate your teachers?** *(Circle one only.)*

Never	1	(Go to Q.29)
Less than once a year	2	
About once a year	3	
More than once a year	4	

28. **Which of the following procedures do you use to gather information for your evaluation?** *(Circle one per line.)*

	No	Yes
Interviews	1	2
Written or oral self reports by teachers	1	2
Observational data on teachers' classroom work	1	2
Student ratings of teachers' performance	1	2
Other forms of systematic evaluation	1	2

E. Questions 29 to 40 have to do with you and your education.

29. **What is your sex?**

Male	1
Female	2

30. **What is your date of birth?**

|__|__| |__|__| |__|__|
Month Day Year

31. **What is your ethnicity/race?** *(Circle one only.)*

Asian or Pacific Islander	1
American Indian or Alaskan Native	2
Hispanic	3
White (non-Hispanic)	4
Black (non-Hispanic)	5

32. **Prior to becoming a principal, did you attend an accredited educational administration program?**

No	1
Yes	2

33. **Approximately how many courses have you completed related to educational administration since your initial certification?**

________ Courses

34. **How many times have you been to in-service educational administration courses in the last three years?** *(Circle only one.)*

Not at all	1
Once	2
Twice	3
Three times	4
Four or more times	5

35. **At the end of this school year, how many years will you have been a school principal?** *(Please include any years as assistant principal.)*

In your total career .... ________ Years

In your present school...................................... ________ Years

36. **Prior to becoming a principal, how many years did you teach altogether?** *(If none, please enter "0.")*

________ Years (or full-time years equivalent)

37. **How many courses related to the teaching of English/Language Arts/Reading did you complete as part of your professional training?**

______________________________ Courses

38. **How many courses related to the teaching of English/Language Arts/Reading have you completed since your initial certification?** *(Circle one only.)*

None	1
One	2
Two	3
Three	4
Four or more	5

39. **How many times have you been to in-service teacher training courses in English/Language Arts/Reading in the last three years?** (*Circle only one.*)

Not at all	1
Once	2
Twice	3
Three times	4
Four or more times	5

40. About how often do you read each of the following? *(Circle one number per line only.)*

		Never or almost never	About once a year	About once a term	About once a month	About once a week or more
a.	Articles on teaching	1	2	3	4	5
b.	Articles on reading comprehension	1	2	3	4	5
c.	Books on history or politics	1	2	3	4	5
d.	Books on the arts	1	2	3	4	5
e.	Books on science	1	2	3	4	5
f.	Novels or short stories	1	2	3	4	5
g.	Poems	1	2	3	4	5
h.	Plays	1	2	3	4	5
i.	Articles on literature	1	2	3	4	5

Thank you very much for your cooperation!

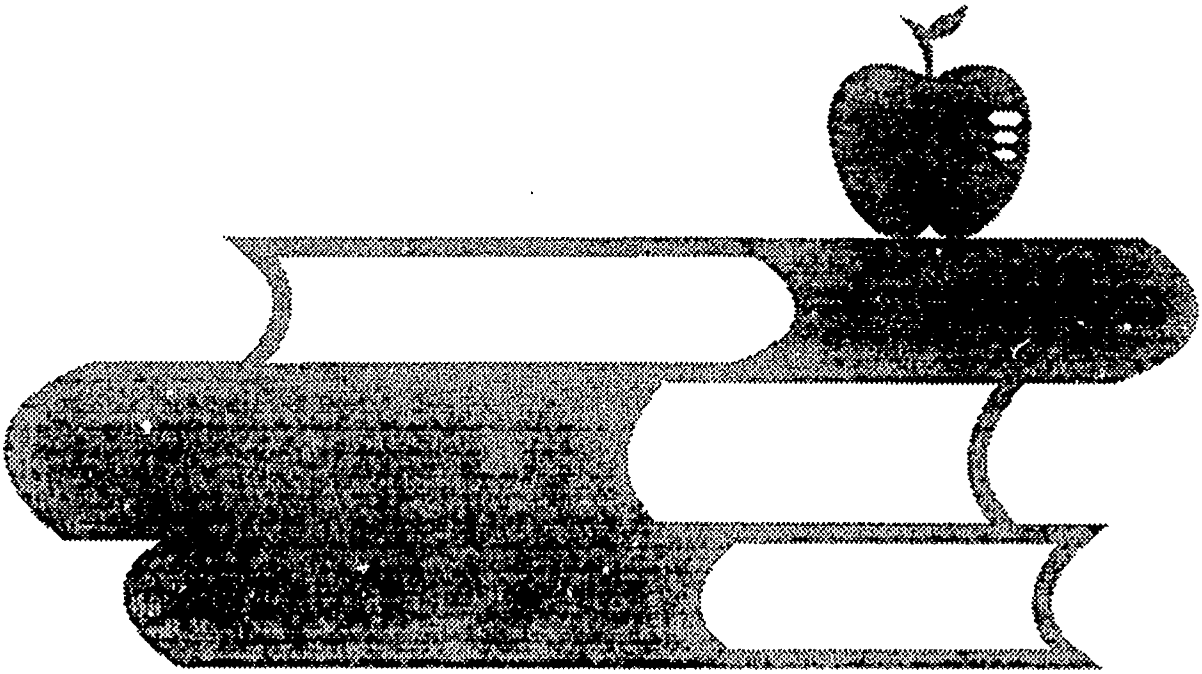
OMB Clearance #: 1850-0645
OMB Expiration Date: 12/92

ID: ______________________

READING LITERACY

SCHOOL QUESTIONNAIRE

Ninth Grade



Directions:

The following questionnaire is part of an international study of Reading Literacy. The questions asked attempt to gather information which captures the wide range of experience and practice which is likely to exist across countries.

Please answer all questions in such a way as to reflect most accurately the situation in your school. It is important that all questions are answered. Most questions require you to circle your selected response. Others require you to write in a number. Where it is appropriate to enter "0" in the answer, please do so. Do not leave it blank. All information will be treated in the strictest confidence.

Public reporting burden for this collection of information is estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the U.S. Department of Education, Information Management and Compliance Division, Washington, D.C. 20202-4651; and to the Office of Management and Budget, Paperwork Reduction Project 1850-0645, Washington, D.C. 20503.

A. Questions 1 to 8 have to do with your school.

1. **What is the total enrollment of full-time students in your school?** *(If there are no boys or no girls, please enter "0.")*

________ Number of boys

________ Number of girls

2. **What is the total enrollment of full-time ninth grade students in your school?** *(If there are no boys or no girls, please enter "0.")*

________ Number of boys

________ Number of girls

3. **How many of the full-time ninth grade students in your school are:**

a. Asian or Pacific Islander,.................................. ________ Students

b. American Indian or Alaskan Native,.................. ________ Students

c. Hispanic,... ________ Students

d. White (non-Hispanic), or................................. ________ Students

e. Black (non-Hispanic)?...................................... ________ Students

4. **Is your school ...** *(Circle only one.)*

A public school, or....................................	1
A private school?......................................	2

5. **Which of the following best describes the community in which this school is located?** *(Circle only one.)*

A rural or farming community	1
A small city or town of fewer than 50,000 people that is not a suburb of a large city	2
A medium-sized city (50,000 to 100,000 people)	3
A suburb of a medium-sized city	4
A large city (100,000 to 500,000 people)	5
A suburb of a large city	6
A very large city (over 500,000 people)	7
A suburb of a very large city	8
A military base or station	9
An Indian reservation	10

6. **Please indicate the availability of the following resources in relation to your school?** *(Circle one number on each line.)*

	Not readily available	Available in neighboring town or city (less than 2 hours of normal one way travel time)	Available locally (within 30 minutes of normal one way travel time)
Public library	1	2	3
Bookstore/book department in a store	1	2	3
Other secondary level school	1	2	3
A higher education institution	1	2	3
Museum	1	2	3

7. **What is the degree of parent cooperation with the school in terms of support for the school's educational principles or goals (compared with other schools you know)?** *(Circle one only.)*

Much below average	1
Below average	2
Average	3
Above average	4
Much above average	5

8. **Which of the following resources and activities are there in your school?** *(Circle one number on each line.)*

	No	Yes
School library	1	2
Reading room for students	1	2
Student/school newspaper or magazine	1	2
Teacher (Professional) library	1	2
Drama club	1	2
Debating club	1	2
Literature club	1	2
Writing club	1	2

B. Questions 9 to 11 are about the school library. If you have indicated in Question 8 that your school does not have a library, please go straight to Question 12.

9. **Approximately how many books with different titles does your school library contain?** *(Exclude magazines and periodicals.)*

________ Books with different titles

10. **Approximately how many books with different titles were added to your school library in the last year?** *(Exclude magazines and periodicals.)*

______ Books with different titles

11. **Can ninth grade students in your school borrow books from your school library to take home?** *(Circle one only.)*

No.. 1
Yes 2

C. Questions 12 and 13 pertain to the number of teachers in your school.

12. **How many of the full-time (or full-time equivalent) teachers in your school are:**

a. Asian or Pacific Islander,.................................. ______ Teachers
b. American Indian or Alaskan Native,.................. ______ Teachers
c. Hispanic,.. ______ Teachers
d. White (non-Hispanic), or................................. ______ Teachers
e. Black (non-Hispanic)?...................................... ______ Teachers

13. **For each of the following categories, how many full-time (or full-time equivalent) employees are there in your school?** *(If there are no male or no female teachers, please enter "0.")*

	Number of males	Number of females
a. Regular classroom teachers (exclude special education)	______	______
b. Special education teachers	______	______
c. Guidance counselors	______	______
d. Librarians and other professional media staff	______	______
e. Reading specialists	______	______
f. Other professional staff (other curriculum specialists, administrative and business staff, social workers)	______	______
g. Teacher aides (paraprofessionals who assist teachers)	______	______

D. Questions 14 to 28 pertain to instruction in your school.

14. What is the <u>total instructional time</u> (in hours and minutes), excluding breaks, in a <u>typical week</u> in your school for all subject areas?

_______ Hours and _______ Minutes per week

15. How many <u>days per year</u> is your school scheduled to be <u>open</u>?

_______ Days per year

16. How many <u>days of instruction</u> were lost in the last school year due to accidents, snow days, floods, strikes, festivals, staff days, etc.? *(If there were no days lost, please enter "0.")*

_______ Days lost **last** year

17. On an average day, approximately what <u>percentage</u> of students are <u>absent</u> from school?

_______ Percent

18. What type of standardized tests of reading achievement does your district administer to students? *(Circle one only.)*

Our district does not administer standardized tests of reading achievement ..	1
Norm-referenced	2
Criterion-referenced	3
Both norm- and criterion-referenced	4

19. Do you use the results of students' standardized tests to evaluate:

	No	Yes
Student progress?	1	2
Curriculum?	1	2
Teachers?	1	2
Textbooks/materials?	1	2
Special programs?	1	2

20. **Rate your satisfaction with the following sources of evidence of your students' progress.** *(Circle one number on each line.)*

	Highly satisfied				Highly dissatisfied
Norm-referenced test scores	1	2	3	4	5
Criterion-referenced test scores	1	2	3	4	5
Student work sample	1	2	3	4	5
Teachers' judgment	1	2	3	4	5
Grade report	1	2	3	4	5

21. **Please describe the special reading programs your school has outside the normal classroom activities (e.g., enrichment programs or special programs for the disadvantaged)?**

22. **Does your school sponsor any informal initiatives to encourage reading (e.g., book clubs)?**

No	1
Yes	2

23. **Does your school have a program for the improvement of <u>reading instruction</u> (teaching and learning of reading)?**

No	1
Yes	2

If "Yes," please describe

676

24. **Do you have any serious problems in providing for the teaching/learning of reading in your school?**

No .... 1 (Go to Q.26)
Yes 2

25. **Rank the following seven problems in providing for the teaching and learning of reading in your school** in order of severity. *(Please assign "1" to the most serious and "7" to the least serious problem).*

	Rank of severity
a. Insufficient specialized staff	________
b. Insufficient time	________
c. Students' lack of interest	________
d. Insufficient funds and classroom material	________
e. Insufficient parental support	________
f. Insufficient number of teacher aides	________
g. Other (Specify) ______________ ______________	________

26. **Please rank the following activities in order of importance in your work as a school principal.** *("1" is the most important activity, "8" is the least important activity, "NA" = not applicable. Do not assign equal rankings.)*

	Rank of importance
a. Representing the school at official meetings	______
b. Evaluation of staff	______
c. Contacts with local community (e.g., parents, community organizations, local industry)	______
d. Discussing educational objectives with the teaching staff	______
e. Administrative tasks concerning the functioning of the school (e.g., regulations, disciplinary duties, school budget, timetable)	______
f. Using records of students' progress	______
g. Taking care of issues of student welfare and guidance	______
h. Activities aimed at the professional development of teachers	______

27. **How often do you systematically evaluate your teachers?** *(Circle one only.)*

Never .... 1 (Go to Q.29)
Less than once a year 2
About once a year 3
More than once a year 4

28. **Which of the following procedures do you use to gather information for your evaluation?** *(Circle one per line.)*

	No	Yes
Interviews	1	2
Written or oral self reports by teachers	1	2
Observational data on teachers' classroom work	1	2
Student ratings of teachers' performance	1	2
Other forms of systematic evaluation	1	2

E. Questions 29 to 40 have to do with you and your education.

29. **What is your sex?**

Male	1
Female	2

30. **What is your date of birth?**

|__|__| |__|__| |__|__|
Month Day Year

31. **What is your ethnicity/race?** *(Circle one only.)*

Asian or Pacific Islander	1
American Indian or Alaskan Native	2
Hispanic	3
White (non-Hispanic)	4
Black (non-Hispanic)	5

32. **Prior to becoming a principal, did you attend an accredited educational administration program?**

No	1
Yes	2

33. **Approximately how many courses have you completed related to educational administration since your initial certification?**

_______ Courses

34. **How many times have you been to in-service educational administration courses in the last three years?** *(Circle only one.)*

Not at all	1
Once	2
Twice	3
Three times	4
Four or more times	5

35. **At the end of this school year, how many years will you have been a school principal?** *(Please include any years as assistant principal.)*

In your total career .... ________ Years

In your present school .... ________ Years

36. **Prior to becoming a principal, how many years did you teach altogether?** *(If none, please enter "0.")*

________Years (or full-time years equivalent)

37. **How many courses related to the teaching of English/Language Arts/Reading did you complete as part of your professional training?**

______________________ Courses

38. **How many courses related to the teaching of English/Language Arts/Reading have you completed since your initial certification?** *(Circle one only.)*

None	1
One	2
Two	3
Three	4
Four or more	5

39. **How many times have you been to in-service teacher training courses in English/Language Arts/Reading in the last three years?** (*Circle only one.*)

Not at all	1
Once	2
Twice	3
Three times	4
Four or more times	5

40. **About how often do you read each of the following?** *(Circle one number per line only.)*

	Never or almost never	About once a year	About once a term	About once a month	About once a week or more
a. Articles on teaching....................	1	2	3	4	5
b. Articles on reading comprehension...........................	1	2	3	4	5
c. Books on history or politics	1	2	3	4	5
d. Books on the arts	1	2	3	4	5
e. Books on science	1	2	3	4	5
f. Novels or short stories	1	2	3	4	5
g. Poems...	1	2	3	4	5
h. Plays..	1	2	3	4	5
i. Articles on literature.....................	1	2	3	4	5

Thank you very much for your cooperation!

680

☆ U S GOVERNMENT PRINTING OFFICE:1994-300-807/23102

United States
Department of Education
Washington, DC 20208–5650

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Fourth Class Special
Special Handling