DOCUMENT RESUME

ED 106 831 CS 001 889

AUTHOR	Smith, Marshall S., Ed.
TITLE	Assessment of Reading Comprehension, Panel 5; Conference on Studies in Reading (Washington, D.C., August 1974).
INSTITUTION	National Inst. of Education (DHEW), Washington D.C.
PUB DATE	Jun 75
NOTE	32p.
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Cognitive Processes; Conference Reports; Decoding (Reading); Functional Illiteracy; Readability; Reading; Reading Ability; *Reading Comprehension; *Reading Processes; *Reading Research; *Reading Skills; Research Needs; Research Problems; Task Analysis
IDENTIFIERS	*National Conference on Studies in Reading

ABSTRACT

The aim of the panel on reading comprehension, sponsored by the National Institute of Education Conference on Studies in Reading, was to determine the influence of intellectual factors, general experiences with the world, knowledge specific to the particular reading task, and general decoding skills to successful reading performance. In addition, an attempt was to be made to find ways to characterize the difficulty and other demand characteristics of written material and the task the reader is asked to perform. Finally, an attempt was to be made to integrate the results of the work just described with other theoretical and practical work in reading. The contents of the document are as follows: "Analysis of the Reading Demands in American Life," which lists several projects for the analysis of functional reading tasks; "Identification of Psychological Factors Involved in Reading Comprehension," which discusses the development of information-processing models for selected reading tasks; "Evaluating Individual Performance and Instructional Outcomes," which has as its aim the development of methods for evaluating understanding and models for individual testing; and "Comprehension and the Design and Evaluation of Written Material for Effective Human Use," which has as its aim the development of techniques to determine rational procedures for designing written material for various prespecified human uses. (WR)

Assessment of Reading Comprehension
No. 5

CONFERENCE ON STUDIES ON READING

"It was unlawful, as well as unsafe, to teach a slave to read.
'It will forever unfit him to be a slave. He will at once become unmanageable and of no value to his master.'
These words sank deep into my heart. From that moment, I understood the pathway from slavery to freedom. Though conscious of the difficulty of learning without a teacher, I set out with high hope and fixed purpose, at whatever cost of trouble, to learn how to read."

Frederick Douglas

NATIONAL INSTITUTE OF EDUCATION
Washington, D.C.
June, 1975

U.S. Department of Health, Education and Welfare

NIE CONFERENCE ON STUDIES IN READING

PANEL 5

ASSESSMENT OF READING COMPREHENSION

PROBLEM STATEMENT

Specify an array of research activities aimed at improving the capacity to assess reading comprehension.

PARTICIPANTS

PANEL CHAIRPERSON:

Dr. Ernst Rothkopf
Bell Laboratories

NIE LIAISON:

Mr. Monte Penney

PANEL MEMBERS:

Dr. Richard Anderson
University of Illinois

Dr. Alfredo Castaneda
Stanford University

Dr. Walter Kintsch
University of Colorado

Dr. Thomas Sticht
Human Relations Organization

FIELD CONSULTANTS:

Mrs. Virginia K. Brooks
Washington, D.C. 20024

Dr. John R. Bormuth
Center for Advanced Study in the Behavioral Sciences

Dr. Edward Coleman
University of Texas

Dr. Fred B. Davis
Botafogo, GB Brazil

Dr. Linnea Ehri
University of California

Dr. Lawrence T. Frase
Bell Laboratories

Dr. Wells Hively
Spalding Youth Center

Dr. Marcia Johnson
State University of New York

Dr. Lou Mosberg
University of Delaware

Dr. Michael Posner
University of Oregon

Dr. David E. Rumelhart
University of California, San Diego

Dr. Amiel Sharon
Educational Testing Service

Dr. Jerry Short
University of Virginia

Dr. Edward Smith
Stanford University

PANEL 5

ASSESSMENT OF READING COMPREHENSION

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

PREFACE

The National Institute of Education (NIE) came into being during 1972. Its authorizing legislation requires the NIE to:

- Help solve or alleviate the problems of, and achieve the objectives of, American Education.
- Advance the practice of education as an art, science, and profession.
- Strengthen the scientific and technological foundations of education.
- Build an effective education research and development system.

In order to aid in meeting these general objectives, the National Council on Education Research (NIE's policymaking body) approved the creation of five priority areas in December, 1973. One of the priority areas was the Essential Skills Program.* The purpose of that program was:

> To investigate through research and development, ways to aid all children to obtain skills essential for functioning adequately in school and society.

The initial focus of the Essential Skills Program was in the area of reading. During 1974, the Essential Skills Program carried out an intensive effort designed to formulate plans for funding research and development activities in reading. A variety of meetings were held with groups of teachers, school administrators, and scientists to designate directions for the program. The most ambitious of the meetings was held in Washington, D.C., in August, 1974, and directly involved over 175 individuals -- 50 as Conference participants and 125 as consultants to the Conference. This report is the product of one of the 10 panels of the August Conference.

The impetus for the Conference stemmed from a number of concerns about the state of Federal funding of research and development in education. Four concerns stood out in particular for reading.

1. Research in the field of reading was fragmented and noncumulative.

*During the past few months, the Essential Skills Program has been renamed the Learning Division of the Basic Skills Group. Both the Basic Skills Group and the Learning Division continue to follow the guidelines set out by the National Council in December, 1973 (above).

2. The Federal Government was not making constructive use of the state of knowledge in the field in their decisions to fund new research and development.

3. There was a lack of positive and firm coordination between the Federal Government and the professional research and practicioner organizations around the country.

4. A large number of scientists in a variety of disciplines carry out research with relevance to reading. We considered it important to attract these scientists to work in the applied areas of educational research.

The Conference itself was a step in meeting these concerns. During the past year, the NIE has been developing plans for funding research and development in reading for the next two years. Suggestions from the Conference have played an important role in this process. But planning is an ongoing process and we hope by publishing and widely disseminating the reports from the Conference to stimulate discussion of the reports, of research and development in the field of reading, and, indirectly, of the plans of the Institute.

To some extent the format for the Conference was influenced by three other similar efforts of the Federal Government. In the area of health research, the conferences leading to the National Cancer Plan and the National Heart and Lung Institute Plan served as partial models. Within NIE, the Teaching Division had held a major planning effort in the area of teaching research during the early summer of 1974. The intent in each of these efforts was to develop a coherent set of documents that would be responsive to the needs of the American public and to knowledge in the field.

We felt it necessary to structure the Conference in two important ways. First, after extensive consultation with scientists and practicioners in the field we arrived at the conclusion that major efforts in the past had often ignored or down-played the critical importance of the stage of reading called "reading comprehension." Although we realized the impossibility of actually separating out "reading comprehension" from the earlier stages of learning to read -- which requires the learner to be able to translate written letters and words into speech -- our advice suggested that the comprehension or "reading for meaning" stage required far more attention than it had received in the past. Consequently, seven of the ten panels focused on problems in this area. Second, to direct the focus of the panels to planning future research we requested the panelists to organize their ideas into general approaches within the problem area, within the approaches to suggest programs for research, and, finally, when possible to specify particular research or development projects.

The seven panels addressing problems in comprehension spanned a wide range of concerns. The first three panels focused on basic research issues. Their panel reports are titled: Semantics, Concepts, and Culture, The Structure and Use of Language, and Attention and Motivation. The fourth panel was asked to consider the problem of Modeling the Reading Process. The fifth panel directed its attention to the issue of measuring how well people read and its report is titled Assessment of Reading Comprehension. The sixth and seventh reports directed themselves respectively at the practical problems of the Application of Existing Reading Comprehension Research and Reading Comprehension and the High School Graduate. The final three panels directed their attention to three pressing concerns in early reading: Learning and Motivation in Early Reading; Reading Strategies for Different Cultural and Linguistic Groups; and Essential Skills and Skill Hierarchies in Reading.

Although the reports have undergone some revision and editing since the Conference, the major part of the work was done in concentrated sessions in the space of a few days. The resulting documents are not polished or exhaustive. They are meant to be working documents to stimulate debate, suggestions, and comments. Such comments or requests for other reports should be directed to:

Director, Learning Division
National Institute of Education
Washington, D.C. 20208

The work of organizing the Conference was carried out by members of the Essential Skills staff at the NIE -- each of the panels had an NIE staff person as a permanent liaison. Special acknowledgements are due to Susan Duffy and Donald Fisher for their assistance in preparing the reports for publication and to Arthur Young & Company for coordination and arrangements before, during, and after the Conference. Finally, the work of NIE cannot proceed without the kind of skill, involvement, and hard work given by the panel chairpeople, panelists, and consultants for this Conference. The ideas and emphases in the reports are the products of their cumulative expertise.

Marshall S. Smith
Conference Chairperson

LIST OF PANEL REPORTS AND CHAIRPERSONS

1. Semantics, Concepts, and Culture, Dr. George Miller, Rockefeller University

2. The Structure and Use of Language, Dr. Thomas Trabasso, Princeton University

3. Attention and Motivation, Dr. Sheldon White, Harvard University

4. Modeling the Reading Process, Dr. Richard Venezky, Wisconsin University

5. Assessment of Reading Comprehension, Dr. Ernst Rothkopf, Bell Laboratories

6. Application of Existing Reading Comprehension Research, Dr. Lauren Resnick, University of Pittsburgh

7. Reading Comprehension and the High School Graduate, Dr. Mina Shaugnessy, City University of New York

8. Learning and Motivation in Early Reading, Dr. Richard Hodges, University of Chicago

9. Reading Strategies for Different Cultural and Linguistic Groups, Dr. Manuel Ramirez, University of California, Santa Cruz

10. Essential Skills and Skill Hierarchies in Reading, Dr. Irene Athey, University of Rochester

PANEL 5

ASSESSMENT OF READING COMPREHENSION

INTRODUCTION

Descriptive Statement

The term "reading comprehension" has come to mean many things. Some people use the phrase as if it refers to a skill or trait, i.e., an attribute of an individual, all or in part acquired through schooling and other experience. Others, ourselves included, use the term to describe performance involving written material, meeting some minimum quality standards. Views about what performances should be considered in this connection and about appropriate standards for measurement differ widely. These differences in views and in practice are not so much due to sharp conflicts in mutually exclusive or antagonistic theories or beliefs as they are to the diversity and loose organization that has characterized reading research.

Any systematic approach to the assessment of reading comprehension must select from the wide range of possibilities and explicitly delineate and describe the nature of the performances which will be considered to imply reading comprehension. This is, in part, necessary to limit any systematically guided research undertaking to manageable size. These explicit descriptions in essence become working definitions of reading comprehension. These working definitions should capture socially important aspects of the use of written materials. They also should provide palpable and realistic domains of phenomena which the theorist and the experimenter will attempt to understand and explain.

This panel tends to favor looking at the many practical uses of written material as sources of those performances we wish to identify with comprehension. However, we do not wish to restrict ourselves exclusively to tasks derived from an analysis of social and practical demands. We do not see reading comprehension as a single process that needs to be explained. Rather, it entails many different performances involving many different kinds of written material. Each may need a different model and may be affected by different factors, although we hope that some general conceptual model will prove possible.

The view of comprehension we propose does not look to the written material for the primary index of the appropriateness of comprehension. We do not see reading comprehension as extracting the "true" content of the message, although we do not completely reject the use of the content of the written message. We believe serious logical and practical difficulties may be avoided by using some referents outside the written document to determine whether adequate comprehension has been achieved. These outside referents, for example, demand purely verbal performances such as answers to questions derived from instructional goals, or may involve some other activity such as assembling a bicycle.

The main problem areas we see in reading comprehension assessment include the systematic classification and analysis of practical reading performances in American life, the theoretical characterization of the psychological competencies needed to perform various reading tasks, and the development of techniques for measuring personal reading competence, the effectiveness of instruction, and the effectiveness of written material.

To a certain degree, each of these problem areas overlaps with other reading panels. This is because we conceive of general assessment issues in comprehension as defining a domain of generality for basic theoretical work, and as a source for practical test development and instructional research.

Summary of Background and Current Knowledge

Scientific research on readers started at the end of the last century. Despite considerable work, the accumulation of systematic knowledge in this area has been slow. The scientific analysis of reading has been difficult because reading appears to depend on many interrelated human capabilities, such as thinking, learning, and perception. Each of these, in turn, is imperfectly understood. Other sources of difficulty have been the lack of explicit conceptualization of reading activities, the lack of exact measurement techniques, and the lack of productive theoretical models.

During the past few decades, several developments bearing on future reading research have taken place. These include the development of more powerful methods for characterizing readability, and of criterion-referenced achievement tests. Failures in machine translation and difficulties in applying the concepts of the transformational grammarians to language behavior have taught us the importance of semantics and the folly of treating "understanding" as a primitive term requiring no further definition.

Theoretical Framework

The point of departure of our approach to assessment issues in reading comprehension is the study of the literacy demands of school, personal life, occupations, and society. The second stage is the analysis of the psychological demands of various uses of written documents. We think this is best done through a detailed task or process analysis which examines reading activity in great detail at a surface level. Attempts should then be made to translate the results of these analyses into more fundamental terms, to make guesses about underlying processes, and to build the kind of verbal and definitional bridge necessary to tie these analyses to the concepts of the psychological laboratory. We hope this conceptual bridging will tie the practical aspect of the work on comprehension assessment to the basic scientific research on language memory, and learning proposed by the panel.

In general, our aim is to determine the influence of intellectual factors, general experience with the world, knowledge specific to the particular reading task, and general decoding skills on successful reading performance. We will also make an effort to find ways to characterize the difficulty and other demand characteristics of written material and the task the reader is asked

to perform. Finally, we will attempt to integrate the results of the work described above with other theoretical and practical work in reading. This will involve the explicit formulation of (1) basic experimental and theoretical questions, (2) procedures for measuring individual performance and for evaluating instructional results, and (3) guidelines for the design of effective written materials.

APPROACH 5.1

ANALYSIS OF THE READING DEMANDS
IN AMERICAN LIFE

Rationale and Background

What must individuals know to use effectively the written material they will encounter in life? In order to find out, it is necessary to determine in considerable detail the various uses of written material and the practical and personal demands society and cultures make on various individual members in the use of these materials. The development of improved task analysis techniques has made possible the analysis of the literacy requirements in some occupations. The explicit characterization of functional literacy demands is a key phase in reading research, because it defines reading and because it is necessary for analysis of the psychological demands on readers. No view of comprehension can stand without sound characterization of functional demands on readers.

Ideally, analyzing the functional demands on readers also provides standards against which theorists of language, memory, and learning can test their intellectual mettle. It is perhaps too grand a goal to expect a comprehensive theory of human memory. However, we might soberly hope for a complete psychological account of simple tasks such as searching through a book index.

Besides these general issues, we need to understand better the nature of the reading tasks involved in various life situations, because more realistic methods for assessing reading comprehension are needed. The point is that reading assessment procedures should have greater "ecological validity," or ability to provide information about how well students are acquiring functional reading skills. These tests might also serve as screening devices capable of reflecting actual reading demands of jobs. The reading tasks, consisting of actual materials and questions addressed to the materials by job incumbents, would constitute a job-sample test in the reading domain and should help resolve questions of validity and bias of selection instruments. Candidates for employment who failed such tests could be given constructive feedback about the specific skill areas they needed to improve, rather than vague advice to pursue general remedial literacy training. Employers might be more willing to offer reading instruction if it were to be limited to reading skills essential to the job.

Of course, although employability is a major concern to individuals and society, other areas of life make demands on literacy skills: managing a family (shopping, caring for the children, consumer purchasing, etc.); attending school, participating in community affairs. We have little knowledge of the types of reading tasks involved in such activities and, hence, little understanding of how to teach and evaluate the skills needed to perform these various tasks.

Therefore, the general purpose of this approach is to suggest research on the nature of reading tasks in various school, occupational, and personal situations so that instructional systems, reading comprehension measures, and scientific studies of the reading process may reflect more faithfully the nature of reading as it actually occurs in American life.

Program 5.1.1: Analysis of Functional Reading Tasks.

This program focuses on the identification of reading tasks which contribute to the solving of some life problem.

Research Considerations

Limited availability of R&D funds forces a narrowing of interest from the domain of all possible reading tasks to a subset of those more pragmatic, functional ones. Therefore, this program focuses on reading tasks performed for educational or other practical reasons rather than for recreational or entertainment purposes--tasks involving the informational aspects and functional consequences rather than the emotional consequences of the written message. This focus reflects the judgment that performance of such functional reading tasks is of immediate concern to the individual seeking a satisfactory life in our society, and to the society which bears the economic burden of preparing future generations to contribute to its general welfare.

Many reading comprehension tests and reading instructional systems exist today, but most reflect merely the "typical" K-12 curriculum. Much evidence suggests that this curriculum and the tests constructed for it do not adequately represent the world outside the school and, hence, are incapable of properly preparing students for full participation in society. The proposed program of research will provide the information needed to construct instructional systems and tests of reading comprehension which are fully representative of the uses of reading materials both within the school system and in the greater society. Additionally, information about the reading tasks in various courses of instruction will permit statements about the relative reading demands of various domains of life activity. Such information would be useful for occupational and educational counseling.

Because of the unmanageably large number of settings for reading tasks, there is a need to divide the settings into representative domains (e.g., occupational reading tasks) and then into job clusters prior to sampling. Fortunately, some research has already been done to prepare the way for more comprehensive projects to identify functional reading tasks.

Methods for task analysis (particularly regarding job analysis) have been rather thoroughly worked out by applied psychologists in industry and in the Departments of Labor and Defense. These procedures permit a determination of the types of reading materials in various settings; uses to which the materials are put; the importance of various reading tasks; the frequency of reading tasks; differences between reading tasks for entry and advancement through job levels; and differences between reading tasks for learning a job versus doing a job. Modification and refinement of existing task analysis procedures will be needed to extend these procedures to nonoccupational settings.

If similar task analysis procedures can be applied to the other communication tasks of writing, speaking, and comprehending spoken language (auding), it may be possible to obtain estimates of the relative importance of reading with respect to other forms of communication. This information is important because reading is only one of the communications skills essential to pleasant existence in our society. Limited research suggests that writing, speaking, and auding are closely intertwined with the use of written materials in many practical situations. The ecological value of reading can be more fully understood within the context of these other communication tasks. Thus, research on nonreading communication tasks is of relevance to the analysis of the role of reading tasks in various life settings. Though the development of writing, speaking, and auding task tests are recommended, they are assigned low priority. Such efforts may be desired at a later date depending on the results of the analyses of functional reading tasks proposed in this program.

This program is divided into four projects: (1) the identification of reading and other communication tasks in American life; (2) the development of inventories of reading tasks scaled for difficulty; (3) the use of such inventories for defining the reading demands of various occupational clusters, community service roles, and other demands of activity identified as important in American life; and (4) task and/or process analysis of selected reading tasks.

Project 5.1.1.1: Identification of Functional Reading Tasks. Past projects concerned with identifying functional reading tasks (e.g., the ETS survey of reading activities and the Texas Adult Performance Level Study) have not provided enough information to develop maximally useful instructional curriculums and assessment devices capable of faithfully representing the functional reading demands of the Nation. The proposed project will remedy this deficiency by providing a detailed analysis of what and why people read in selected critical domains of American life.

Project 5.1.1.2: Scaling of Reading Tasks for Difficulty. The purpose of this project is to develop and administer reading task tests to specified populations to determine difficulty indexes for the tasks identified by Project 5.1.1.1.

This research may require large-scale testing of national samples of the population. The result should be an inventory or catalog of reading tasks scaled for difficulty. Such an inventory can provide a resource bank to (1) facilitate test construction and (2) permit evaluation of the literacy demands of the domains composed of various reading tasks in the catalog. The latter possibility is the purpose of the next project.

Project 5.1.1.3: Determination of Reading Difficulty Levels of Selected Domains of American Life Activities. In Project 5.1.1.1, a set of critical domains of American life will be defined and reading tasks for these domains will be identified. In Project 5.1.1.2, these tasks will be scaled for difficulty to produce a catalog of reading tasks. Such a catalog would present

the frequency of occurrence of the various tasks and their scaled difficulty value. This project will take the next step toward making the catalog useful. For any domain represented in the catalog, the weighted mean of the difficulty values for the tasks would estimate the difficulty level of the domain. In addition, the known difficulty values may be useful in helping to estimate the overall difficulty of other domains which have tasks in common with the domains represented in the catalog. This information could be used in advising and counseling regarding education and literacy training needs for active participation in the activities of the domain of interest. It is, of course, recognized that the reading difficulty of a particular practical domain is not fixed but may be changed by substantial investment in revisions of the written materials used. The economic and other systemic implications of various strategies for reducing or simplifying reading demands should be studied.

Project 5.1.1.4: Task and/or Process Analyses of Various Uses of Written Material. The most important single element of this program is task and/or process analysis of the more important uses of written material. After tasks have been identified, scaled, and cataloged, the next step is to describe the conditions of each particular reading task in sufficient detail to permit and stimulate its scientific analysis.

Techniques adequate for this type of analysis have been developed in training and human factors work. The results of these analyses typically describe the condition initiating a particular action, analyze the stimulus display, and describe the responses required at each stage, the feedback conditions, and the signs that performance is adequate. Some inferences about psychological processes may also be made.

The application of these techniques to reading activity should result in explicit descriptions of human performance in important social settings. This should provide a challenge to the psychological theorist and experimenter to test their theories and knowledge by providing reasonably complete accounts of such reading tasks as searching an index or finding the price of an item in a catalog. A rigorous quantitative model of a practical reading task would be just as interesting as a model of associative memory and perhaps socially useful as well. Reading task analysis and attempts to account for the human competencies which the reading tasks demand may provide reality tests for an experimental science, as well as produce other salutary consequences.

APPROACH 5.2

IDENTIFICATION OF PSYCHOLOGICAL FACTORS INVOLVED IN READING COMPREHENSION

Rationale and Background

The purpose of Approach 5.1 is to provide an inventory of culturally important reading tools and detailed analyses of the empirically observable performances they require. The next issue of importance is to identify the psychological factors that determine performance in these various reading tasks.

Can productive conceptualizations be found that allow the decomposition of performance into basic information processing components? How much is reading performance determined by general factors and how much by context-specific elements? We do not believe that it is possible to develop a general comprehension model applicable to all reading tasks, though any eventual model may involve the same components in different ways.

We have emphasized before that comprehension is not a single unitary process. Thus, each reading task may need its own performance model. However, it is hoped that there will be important commonalities among these models. If one is able to specify the component processes in various reading tasks, it is plausible that the number of basic psychological performance components for the various major reading tasks is small. Furthermore, several of these components may have been investigated in the psychological laboratory already, although not necessarily in the context of reading research.

We therefore propose that components of reading performance be isolated via cognitive analysis of the reading tasks determined in Approach 5.1 (Program 5.2.1), and that psychological study of these basic components of information processing be undertaken. The three aspects of this inquiry concern the role of the reader's knowledge (Program 5.2.2), the characteristics of the document being read (Program 5.2.3), and the information-processing analysis of the comprehension process itself (Program 5.2.4). The outcome of this work should be a series of explicit models of the reading process (for different tasks) which could be used as a foundation for the construction of instruments used in assessing reading performance (Approach 5.3) as well as in the improvement of reading materials (Approach 5.4). Although a theoretical approach to these problems is possible and often necessary, success will ultimately depend upon our understanding of what to assess and improve.

Program 5.2.1: Information-processing Models for Selected Reading Tasks.

The aim of this program is to interpret the results of socially important reading tasks in terms of theoretical formulations now evolving in cognitive psychology (information processing models, stage analyses, and experimental methodologies to test such models). Attempts to construct information-processing models of reading comprehension should be informed by analyses of practical reading tasks. The following two projects suggest ways of incorporating the information from these analyses into viable cognitive models. However, the eventual development of powerful models will also depend on other knowledge. The remaining programs and projects in this approach are an attempt to define the scope of the required knowledge.

Project 5.2.1.1: Component Analysis. Component analysis consists of a review by experts in cognitive psychology of the step-by-step task analyses supplied in Approach 5.1. Task analysis components need to be identified with information-processing components identified by cognitive psychology. The possibility that the same solution could be achieved through different processes must be considered.

Project 5.2.1.2: Factors Specific to Reading Comprehension. By comparing information processing models of reading tasks with models of how information is acquired by other means (listening, pictures, graphs), elements specific to reading performance might be isolated.

Program 5.2.2: Use of Previous Knowledge in Obtaining New Information From Reading Material.

Research in this area can be classified according to the size of the units of knowledge being studied. Both linguistic units (words-sentences-texts) and the corresponding conceptual units (word concepts-propositions-text bases) are of interest.

Project 5.2.2.1: Word Recognition. We need studies concerned with the processing of the different types of information involved in word recognition, e.g., graphemic-phonemic-semantic. Special attention should be paid to the apparently close interdependence of the grapheme decoding process and semantic analysis (i.e., comprehension) because of direct implications for beginning reading instruction.

Project 5.2.2.2: Word Retrieval and Vocabulary Development. The intent of this project is to gain some understanding of the structure of the internal lexicon and its development, and how the reader retrieves lexical information in the process of comprehension.

Project 5.2.2.3: Use of Knowledge at the Propositional Level. In reading, lexical information is not retrieved from isolated words, but usually from multiword units processed more or less simultaneously. We need to know how this is done. These studies will investigate how such units are represented in "semantic memory" and the nature of the retrieval process. Existing research has been almost exclusively concerned with (1) determining the truth or falsity of sentences on the basis of a person's knowledge (or "semantic memory"), or (2) using information about simple, unstructured storage systems to explain more complex, highly structured ones. While the former studies are perhaps too simplistic, the latter ones are too speculative and almost certainly wrong in detail; nevertheless, this area is lively and energetic and some progress is being made. The main problem is that too much of this kind of work is restricted to one type of sentence: set inclusion statements. The problem of knowledge use is much more general than that. Different types of propositions (other kinds of relatively simple semantic relationships, such as "part of," "property of," etc., but also more complex relationships such as "caused by," or "purpose of") must be studied and larger units of knowledge (e.g., themes of a story and larger bodies of knowledge such as the history of a certain period) should be investigated.

Project 5.2.2.4: The Role of Inferences in Comprehension. Knowledge structures must be understood as dynamic and flexible rather than fixed and complete. Knowledge structures are the outcome of the application of inference rules to knowledge elements. The psychological, linguistic, and logical study of the nature of these inference rules seems very important to an improved understanding of reading. We are not talking here about the rules of standard logic, but about what has been called the "natural logic," i.e., these rules include specifically such nonlogical instances as the conversation postulates involved in interpreting the statement, "It is cold in here," as a request to turn down the air-conditioning. Operations performed with nonpropositional memory representations (analog representations, imagery) need to be investigated, both theoretically and experimentally.

The study of such inference rules is directly relevant to the assessment of reading comprehension. It is probably true that many comprehension failures are not failures to understand the words per se, but failures to recognize the intended interpretation, i.e., connecting what is being read with relevant bits of knowledge and drawing certain inferences from it. To what extent this is the case, and how such inference failures are to be assessed (or corrected) is a research question. But unless we have a much better understanding of "natural logic," progress in this area is impossible.

Program 5.2.3: Message Characteristics.

There are two aspects of this problem: Specification of the form of a text (its "surface structure"), and specification of content (the terms "text base," or "underlying conceptual base" are often used). Sentence grammars, as well as the text grammars now being considered by linguists, seem sufficient for the first purpose. Considerable research emphasis needs to be devoted to the second. The development of techniques for understanding the content of a message independently of how it is expressed, perhaps in a paraphrase or even in a graph or a picture, would strongly accelerate

research progress in the reading area. Specifically, such an ability is important for many assessment problems described in Approaches 5.3 and 5.4. The explicit specification of content, although very important, is an extremely difficult problem, and the likelihood of rapid progress is small.

Project 5.2.3.1: Research on the Representation of Knowledge. Psychologists, linguists, and computer scientists need to make a concentrated effort to develop and standardize current systems for the representations of meaning. Different purposes should be considered, e.g., how to represent abstractly the structure of a narrative or description versus the structure of instructions to perform a certain task.

There is also a need to test experimentally the psychological validity of the representations developed in Project 5.2.3.1. For instance, if such a system uses propositions as a unit, then it is necessary to inquire whether propositions actually function as performance units, i.e., whether number and type of propositions in a text predict reading time, or whether or not propositions are recalled as units.

Project 5.2.3.2: Standardization of Textual Materials for Purposes of Research. Such a project would be immediately useful to researchers in this area. Sentences and paragraphs as well as words could be scaled along various dimensions (e.g., reading time, abstractness, ease of recall under standard conditions, etc.) and analyzed according to some of the proposed theoretical systems. It will be important to scale various types of material to take into account the tasks identified in Approach 5.1. This scaled material could then be used in research, and, even more importantly, other newly developed material could be characterized with respect to these available normed texts. In this way, some standard of comparison for such materials could be obtained.

Program 5.2.4: Comprehension Processes.

Studies in this area are to identify the psychological components of comprehension processes, particularly reading tasks. Two issues appear to be of the most immediate significance: the use of context in comprehension and the effects of task constraints on how much comprehension occurs.

Project 5.2.4.1: Context Factors and Their Influence Upon Performance. We should undertake research to investigate the role of both the linguistic context (previous sentences in a text, old and new information in a sentence) and the nonlinguistic context (general communication situation, pictures, etc.) in determining comprehension. A common view today is that comprehenders use a large number of the cues available to them, the message itself as well as the context, to arrive at an interpretation of the reading material specific to their particular reading purposes (task constraints). Clearly, this research cannot proceed independent of the research on the role of inferences in comprehension described in Programs 5.2.2 and 5.2.3. We should also direct

research toward diminishing the discrepancy existing today between formal models (Section 5.2.3.1) which generally neglect context effects, and those more informal approaches aimed at providing demonstrations of the importance of context in comprehension.

Project 5.2.4.2: Comparing the Efficiency of Information Processing in Reading and Listening Tasks. The comparison of reading and listening performance provides important cues about whether performance deficits are specifically related to the translation and representation of written materials into internal representations. If deficits appear both for reading and listening, intellectual or experiential determinants must be suspected. We propose that the methodology of this type of analysis be investigated. The role of task characteristics as well as that of individual differences must be considered in this respect.

Project 5.2.4.3: Levels of Comprehension. Even with the same readers and the same reading materials, the psychological processes involved in comprehension may be entirely different, ranging from reading without comprehension to making full use of the readers' knowledge about what they read, and the inference rules available to them.

Research needs to proceed along several related lines. We should study how task constraints affect the level at which the reader processes some material (e.g., incidental versus intentional learning, organizing activities, adjunct questions). Secondly, there is, today, much interest in a levels-of-memory approach. It seems that deeper levels of comprehension result in more permanent memory traces. We need to explore these issues in order to construct scientifically based assessment instruments. An important goal of this research would be to determine those environmental conditions that affect the level to which an individual will process a text. This has practical significance because it may suggest instructional practices.

Project 5.2.4.4: The Interrelationship Between Reading Speed and Comprehension. Is there a trade-off between these factors? How is it to be modified? What role does automaticity play? Comprehension tests must be sensitive to both aspects of reading performance. Techniques to monitor comprehension during the actual reading process should be explored.

APPROACh 5.3

EVALUATING INDIVIDUAL PERFORMANCE AND INSTRUCTIONAL OUTCOMES

Rationale and Background

The purpose of this effort is to develop rational methods for evaluating individual understanding and use of written materials, and to develop models for individual testing and formative and summative decisions about instruction.

The presumption is that current assessment practices have several deficiencies. Evaluation is not based on an adequate analysis of the general and specific literacy demands of society, nor on validated conceptions of the processes entailed in reading. The items in reading comprehension tests are not selected or constructed in an orderly, rule-governed fashion. The intuitions and judgment of item writers are paramount. Universally employed empirical item-analysis procedures cannot be counted on to maintain validity. Finally, reading comprehension tests could be made more functional than they are at present. Diagnosing individual deficits in reading comprehension, selecting text material which individuals could read with maximum instructional benefit, and evaluating higher level reading instruction are all important functional uses for which comprehension tests should be designed.

The greatest problem with traditional and still current assessment procedures is the poor fit between purposes and methods of assessment, and between measurement instruments and information requirements. Tests are being used for purposes for which they were never intended, and test results are consequently misinterpreted.

Traditional psychometric theory is based on the measurement of individual differences; it can guide the construction of test results to determine how much one individual knows relative to other individuals. This type of information is needed where relative attainment is the criterion for selection and promotion. Traditional psychometric methods address restricted purposes. Traditional test theory informs the test maker which of a previously generated set of items to include in a test in order to insure fine interindividual discriminations, but it fails to say how the items should be produced in the first place.

A more common and important purpose for assessment is to determine the absolute level of competence -- exactly what the individual knows. Current test theory provides no method to assess actual level of comprehension. No method is developed yet that generates test items representative of a particular domain of knowledge as required for a valid and reliable comprehension test. Consequently, current test items reflect the item writer's own judgments and intuitions; any claim that the test validly assesses true comprehension is unwarranted.

Program 5.3.1: Assessing Various Kinds of Comprehension.

The goal of this program is to formulate explicit procedures for generating test items or tasks upon which successful performance will imply comprehension. A key word here is "explicitness." Ideally, the boundaries of the class of documents, the rules for generating items or tasks, and the criteria for evaluating performance would be so clearly defined that different evaluators could construct equivalent assessment devices.

Since there are a number of kinds of reading material and many purposes for reading, it follows that a variety of assessment procedures will be required. Research done under Approach 5.1 will indicate which projects in this area should be given priority.

Research on assessment procedures should be guided by the principle that many factors interact to determine reading performance, including the characteristics of the document, the skills and knowledge of the reader, the purpose or task demand, and the linguistic and nonlinguistic milieu in which the performance takes place. The tradition of individual assessment has perhaps emphasized too much the measurement of context-insensitive, "generalized capacities." One result has been the classification of some people as deficient, when, in fact, they are quite capable of performing successfully in specific settings.

While the main thrust should be to provide prototypes for the assessment of different kinds of reading comprehension, the research in this area should also pay attention to efficient and economical methods of obtaining information. Sequential-contingent testing may prove to be an interesting technique. If the content and skill domain can be structured in some nonarbitrary manner, response-contingent testing should also be explored.

Project 5.3.1.1: Assessment of Comprehension of Nonnarrative Documents. It seems possible to forecast that research done under Approach 5.1 will reveal that many socially important reading tasks do not involve narrative or expository text. Rather, the documents are catalogs, lists, reference manuals, and so on. Representative reading tasks involving such written materials might include using an index to locate information in a manual and following step-by-step directions. We should undertake research and development into exemplary procedures for measuring competence at such tasks. The measures should stress speed and efficiency as well as accuracy.

Project 5.3.1.2: Assessment of Comprehension of Text. All too often, in school and in instructional research, people are expected to learn text rather than learn from text. Research into techniques for assessing comprehension measures other than the capacity to reproduce or recognize the surface form of passages should receive high priority. A representative but not exhaustive list of the tasks from which subsurface comprehension of text may be inferred is, (1) recognizing or producing paraphrases, (2) applying concepts and principles, (3) drawing inferences, and (4) organizing and synthesizing information.

We have made some progress in explaining orderly procedures for assessment of text comprehension, but much more remains to be done. Also, analysis is required to determine which sort of measure is appropriate depending upon the reader, the purpose, and the nature of the text. Different measures cannot be regarded as interchangeable.

Program 5.3.2: Domain-referenced Tests of Reading Comprehension.

The aim of this program is to develop tests which indicate directly the kind and difficulty of materials a person is capable of reading with specified comprehension. The established practice is to represent individuals' reading competencies on a grade-equivalent scale. This indicates how they compare to others, but permits only oblique answers to such questions as whether people can understand a story on the front page of a newspaper or complete an insurance claim form. The sorts of tests envisioned would scale categories of words, texts, and other documents for difficulty. The measure of competence would be the probability that the examinee could perform an indicated task within the specified category of verbal material.

Project 5.3.2.1: Domain-referenced Vocabulary Assessment. The simple vocabulary test in which examinees select the alternative most nearly synonymous with the target word is a powerful predictor of learning from written materials. Measures of vocabulary difficulty account for the lion's share of variance in the readability of discourse. Yet, information about the size of an individual's vocabulary, or a specified population's vocabulary, currently is not gathered and calibrated in ways useful to making decisions about the design or selection of instructional materials, reference manuals, and job aids. Procedures ought to be developed for estimating from a sample the probability that a person knows the meaning of a word in a specified corpus, for instance, a word appearing fewer than 50 times per million in the Kucera and Francis norms or words likely to be encountered in automotive maintenance manuals.

Project 5.3.2.2: Domain-referenced Tests of Text Comprehension. At the present time, measures of reading "ability" and characterizations of reading materials do not coordinate well. The most directly useful measure of reading competence would be an estimate of the probability that a person could perform a specified task with respect to a certain class (i.e., domain) of documents. With this information in hand, a teacher or supervisor could make informed judgments more easily.

A concrete example illustrates this point. Catalogs constitute a reasonably homogeneous class of documents. They are ubiquitous in occupations that entail the repair of complex equipment. The service manager of a garage who knew that a job applicant could successfully answer queries about the order number, dimensions, and cost of a king-size mattress from consulting a mail-order catalog would be able to infer that the applicant might make a suitable apprentice in the parts department. Of course, more elaborate measures of vocabulary load and structural complexity could and should be developed to characterize the range of documents with which success could be predicted.

Program 5.3.3: Diagnosing Individual Deficits in Comprehension and Their Remedies.

The purpose of this program is to refine domain-referenced tests to the extent that they are capable of yielding specific information about individual learning problems. Simply knowing the individual's level of comprehension is not as useful as knowing at what point comprehension fails. Precise information on where individuals encounter difficulty can fill two important needs in education: The information can (1) indicate what individuals should do to fill in their own comprehension gaps, and (2) suggest how instruction should be revised to be more effective in the future.

Project 5.3.3.1: Assessment Based on Task Hierarchies. The aim is to develop tests based on task-learning hierarchies so that missed items indicate at what point in the hierarchy comprehension fails.

Project 5.3.3.2: Assessment Based on Response Analysis. Operationally defined means of generating distractors for multiple-choice tests and means for content analysis of freely composed answers are to be constructed so that response analysis can reveal points of miscomprehension.

Program 5.3.4: Determining the Content and Effectiveness of Higher Level Reading Instruction.

The purpose of this effort is to furnish the basis for decisions about the content and scope of reading instruction beyond the level at which basic decoding skills have been mastered. One issue is whether there ought to be developmental reading exercises aimed at general literacy, apart from the study of specific disciplines and the preparation for specific occupations. If there is to be separate reading instruction, should it stress the acquisition of information and ideas from a variety of linguistic structures, or the development of self-monitored strategies for maintaining full processing of text, or both?

Project 5.3.4.1: Instruction in the Extraction of Information from Various Linguistic Structures. Linguists and investigators in related fields have been actively engaged in specifying syntactic and semantic structures, not only of individual sentences, but of lengthy passages as well. Examples of such structures are those governed by "because," "however," "problem-solution," and their paraphrases. Research has indicated that there are substantial problems involved in the extraction of information from such structures. A plausible approach would be to develop a program of reading comprehension instruction that would teach people to extract information and ideas from passages which systematically present the variety of important linguistic structures.

Project 5.3.4.2: Instruction in Heuristics for Self-management of Reading Activities. Evidence from sources as diverse as psycholinguistic studies (Program 5.2.4) and analyses of the reasons for poor performance in college implicates failures to process text at a sufficiently deep level as a cause of inadequate comprehension. There are several ways this problem can be approached. One rationale for adjunct questions (Approach 5.4) is that they maintain processing activities. Another approach would be to teach people self-management tactics. While a research and development effort is required before specific recommendations can be made, it might turn out, for instance, that people would be best advised not to take notes while they read but rather attempt to write a precis after studying a short-text segment.

APPROACH 5.4

COMPREHENSION AND THE DESIGN AND EVALUATION OF
WRITTEN MATERIAL FOR EFFECTIVE HUMAN USE

Rationale and Background

The maturing of scientific conceptions of comprehension is likely to have substantial impact on how written materials are produced and used. The emergence of suitable analytical machinery and reliable measurement techniques in this area may make it possible to provide useful information for writers and editors. The results of experimentation and theorizing in this area can be expected to influence the course of formative evaluation and to improve the means by which schools and other users of written materials can evaluate the products offered to them. Procedures for the use of written materials for various purposes can be provided to users. Development of such instructions may also be influenced by the development of rational procedures for measuring comprehension.

Progress in our ability to assess comprehension is likely to influence the production of written materials in several ways. The programs advocated here tend to stress consideration of the purpose for which a written document is to be used. The spirit of this approach will tend to change the way writers approach their tasks and the nature of their products. The producer of written material is also likely to be helped by insights into the way the characteristics of the users determine whether a document is successfully used and what needs to be done to tailor a document to the experience, abilities, and interest of users. It can be expected also that progress toward describing the successful use of written materials in more explicit and exact terms will foster rational procurement and evaluation procedures for written materials by schools and other consumers.

The approaches presented above are all expected to provide guides for more effective writing, but writers will also need information about the characteristics of the populations for whom they are producing materials.

The main aim of the activities contemplated in Approach 5.4 is to develop rational procedures for designing written material for various prespecified human uses. We also hope to develop quantitative characterizations of the difficulty and the content of these materials.

Program 5.4 1: Techniques for Determining the Comprehension-relevant Characteristics of Specified Populations.

The primary aim of this program is to develop techniques to determine the characteristics of special populations that bear on their successful use of various kinds of written material. A special population would be of interest for studies of this kind if it represented a target readership for a writer's

project and was sufficiently large and homogeneous. The results of fundamental research on language and learning indicate that task-relevant knowledge and experience constitute the single most important population characteristic determining successful use of a written document. Because writers and readers are often drawn from different milieus, the writer's assumptions about these characteristics are often wrong.

The program seeks to explore useful techniques to characterize various aspects of population knowledge and to determine how the information for any given target population should be presented to the writer.

Project 5.4.1.1: Techniques for Making Concept and Lexical Inventories in Selected Populations. Writers are aided in their task if they can safely assume that key concepts are familiar to the population for which they are writing. Concepts include the ability to identify a number of environmental states relative to some descriptive terms such as roughness, color, or accuracy. Limited inventories of concepts are probably feasible for special populations such as certain occupational specialties. An effort should also be made to adopt the techniques developed in Project 5.3.2.1 into economical methods for characterizing the vocabulary skill of certain preselected populations. Specialized vocabularies would be of particular interest in connection with various kinds of technical writing. An effort should be made to explore the entire semantic range of each vocabulary item. It may be of value to the writer to know that a given population tends to understand a term in a concrete, specific sense but not in another, perhaps more abstract or figurative meaning.

Project 5.4.1.2: Conceptual and Lexical Inventories, and the Design of Written Material. This project attempts to determine what information about the conceptual and lexical inventories of target populations is of value to writers and editors for various purposes. This project contemplates a series of experimental studies to investigate the relationship between lexical and conceptual knowledge of populations, document content, and task demands.

Program 5.4.2: Determining Difficulty and Other Demand Characteristics of Written Material.

The purpose of this program is to develop economical quantitative indexes for predicting how well material will be comprehended by certain specified populations and how usable the material will be. The creation of these indexes is considerably beyond the capabilities of the current state-of-the-art readability formulas. The desired indexes would include lexical. semantic, and syntactic or stylistic determinants.

Project 5.4.2.1: Indicators of the Content of Instructional Documents for Well-specific Uses. Message content is not the primary benchmark for comprehension in the present approach. Nevertheless, the content of instructional documents strongly influences comprehension performance. The project seeks to explore two approaches: (1) The search for compact but comprehensive

indicators of topical content that will allow teachers to judge the usefulness of a document for a given purpose, and (2) the development of techniques for deriving content structures that characterize both the content and the organization of text.

Project 5.4.2.2: Indicators of the Cognitive Demand Characteristics of Nonnarrative Documents with Well-specified Uses. Almost nothing is known about how the cognitive demands of nonnarrative documents may be characterized. An active search for techniques by which structure and content of nonnarrative material may be quantitatively characterized may produce useful methods to characterize its difficulty. The nature of the task these documents should guide also requires careful scaling; this problem may be as difficult as determining the cognitive demands of the texts.

Project 5.4.2.3: Development of Improved or Augmented Indexes of Text Structure or Style. Classical approaches to readability measurement have not proved to be very powerful predictors of comprehension performances. There is substantial reason to believe that the variance in comprehension performance accounted for by structural (rather than content) text variables may be rather limited if there are no time constraints. Nevertheless, the availability of efficient computer procedures for processing text makes the development of new or augmented readability indexes economical even with small returns. New statistical estimates (proxes) of comprehension-relevant text features should therefore be explored. For example, the number of nominalizations can be estimated by searching for "ion" suffixes. This is not an exact method, because words such as nation or caution will be mistakenly counted, but it may be sufficient for predicting difficulty. Other refinements to readability formulas, such as weighing prepositions by their location in sentences, may be interesting to try. The principle for exploration should be to use approximate statistical methods in large bodies of text to evaluate variables proved relevant to comprehension in laboratory experiments involving small numbers of verbal elements. It is also worthwhile to investigate the feasibility of a quantitative stylistic usability index for nonnarrative written material such as procedural directions.

Program 5.4.3: Systems, Methods, and Aids in the Preparation of Special Purpose Written Material.

The preparation of written material is a difficult practical activity. The emergence of new scientific information about the use of written material may make the writer, producer, or editor's task more difficult, rather than easier at first. This builds resistance to innovation. Some imaginative work is needed to help writers or editors make use of psychological information.

Project 5.4.3.1: Computer Aids for Writing and Editing. Computer aids which provide prompt display of information about target populations may be of substantial help to the writer. When written material is reduced to computer readable form early in the production process, information about text may be displayed to the author or editor and compared automatically to computer standards for the target populations.

Project 5.4.3.2: Production Systems for Writers and Editors. This project involves investigation of how the writer's task should be structured in order not to exceed the writer's capacity to handle the various psychologically important components of the task. Alternate systems for the production of written material for well-specified purposes should be explored. It is not quite sufficient to supply writers and editors with basic psychological information on how humans understand without taking into consideration the human limitations of the writers.