DOCUMENT RESUME

ED 297 963 SE 049 461

AUTHOR Ryan, Joseph M.
TITLE Improving Mathematics Curriculum and Instruction: A South Carolina Model for Using Statewide Test Data.
INSTITUTION Southeastern Educational Improvement Lab., Research Triangle Park, NC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 88
CONTRACT 400-86-0007
NOTE 54p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; Elementary Education; *Elementary School Mathematics; *Evaluation Utilization; Mathematics Education; *Mathematics Tests; *Test Interpretation; Test Results; Test Theory; *Test Use

ABSTRACT

This monograph is designed for educators who must make decisions about curriculum, instruction, and the training of teachers. Many important educational decisions are made based primarily, if not exclusively, on test scores. This reliance on test scores has evolved for a variety of reasons. This monograph starts with test scores as a given in American education today, not as an endorsement of the emphasis given to test scores, but as a recognition of what has become common practice. The purpose of this monograph is to describe a set of procedures for reanalyzing test data to extract as much useful information as possible for making important educational decisions. The monograph and procedures described in it focus on the reanalysis of currently available test data, not on collecting new test data and initial analysis. The reasons for focusing on the reanalyses of currently available test data are described in the text. In this monograph, the reanalysis of mathematics test data is described as a model for other educators interested in taking full advantage of test data that have already been collected. (PK)

Improving Mathematics Curriculum and Instruction:

A South Carolina Model for Using Statewide Test Data

by Joseph M. Ryan
University of South Carolina

SEIL Mathematics Improvement Program

SOUTHEASTERN EDUCATIONAL IMPROVEMENT LABORATORY

RESEARCH REPORT

SOUTHEASTERN EDUCATIONAL IMPROVEMENT LABORATORY

P.O. Box 12746 ▲ 200 Park Offices ▲ Suite 204
Research Triangle Park ▲ NC 27709
919-549-8216
800-237-4829 (*outside NC*)

The Southeastern Educational Improvement Laboratory (SEIL) is a regional educational laboratory serving the education communities in the six southeastern states: Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina. Working with and through existing educational organizations in the region, SEIL offers information and technical assistance to improve writing and mathematics instruction, educational leadership, dropout prevention, rural education, instructional technology, school reform efforts, and the teaching profession.

The Southeastern Educational Improvement Laboratory is "client-owned." That is, its agenda is set by those it serves: educators and other individuals and groups supporting education in the six service states. A 24-member board of directors includes representatives of teachers, school administrators, state and local boards of education, parents, business and industry, professional organizations, higher education, state legislators, as well as the six chief state school officers. Regional and state advisory networks contribute to the decision-making and service-delivery process.

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement of the United States Department of Education, under Contract Number 400-86-0007. The content of this publication does not necessarily reflect the views of the Office of Educational Research and Improvement, the Department, or any other agency of the U.S. Government.

Improving Mathematics Curriculum and Instruction:

A South Carolina Model For Using Statewide Test Data

by Joseph M. Ryan
Department of Educational Psychology
College of Education
University of South Carolina

Southeastern Educational Improvement Laboratory
Mathematics Improvement Program

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This monograph was first suggested by Dr. Frederick E. Smith at the Southeastern Educational Improvement Laboratory, whose interest, encouragement, support, and advice have been most helpful and are most gratefully appreciated.

The editing provided by Betty Howie at the Lab has been most valuable in making the text more clear and succinct.

Jeanne Miyasaka of the South Carolina Department of Education has contributed significantly by critiquing and editing several versions of the monograph. The Conceptual Model for the South Carolina project owes much to her insights.

David Potter and Peter Samulski, also of the S.C. Department of Education, were kind enough to read and respond to early drafts of the document.

The author would like to note his special appreciation for the diligent and careful efforts of Jackie Richardson in typing this document.

Many people worked on the South Carolina Project on which this monograph is based. Special appreciation is extended to Dr. Paul Sandifer and the staffs in the Office of General Education and the Office of Research at the S.C. Department of Education.

The preparation of this monograph was supported with funds from the Southeastern Educational Improvement Laboratory. This sponsorship is gratefully acknowledged and appreciated.

How to use this monograph

This monograph is designed for educators who must make important decisions about curriculum, instruction, and the training of teachers. These decisions and the educational impact they have on students are greatly influenced by the information available to decision makers, and there are many useful, important, and valid sources of information on the basis of which such decisions can be made. These include direct and indirect observations of teachers and students, interviews with teachers and students, teachers' lesson plans, samples of students' homework and seatwork, and scores from a wide variety of tests. Important decisions that will have a major impact on students and their teachers should be based on information from a wide range of sources such as these.

Many important educational decisions, however, are made based primarily, if not exclusively, on test scores. This reliance on test scores has evolved for a variety of reasons, the study of which is beyond the scope of this monograph. This monograph starts with test scores as a given in American education today, not as an endorsement of the emphasis given to test scores, but as a recognition of what has become common practice.

The purpose of this monograph is to describe a set of procedures for reanalyzing test data to extract as much useful information as possible for making important educational decisions. The monograph and procedures described in it focus on the reanalysis of currently available test data, not on collecting new test data and initial analysis. The reasons for focusing on the reanalyses of currently available test data are described in the text.

In this monograph, the reanalysis of mathematics test data from the South Carolina Basic Skills Assessment Program is described as a model for other educators interested in taking full advantage of test data that have already been collected. The South Carolina project illustrates a wide range of reanalysis techniques, procedures for interpreting the results of the reanalyses, procedures for disseminating information about the project, and some developments in curriculum, instruction, and teacher training that have emerged from the project.

As a general model, the procedures employed in the South Carolina Project probably need to be modified when applied to a particular testing program at the school, school district, or statewide level. In adapting the procedures from the South Carolina project to a new setting, a group of educators should be identified who have responsibility for studying this monograph. Mathematics educators should be heavily represented on any such group, along with researchers who would bring technical expertise to the project.

After examining the details of the South Carolina project, the Project Study Group would draft a Project Plan in which they select or modify elements of the South Carolina Project and add new components which are appropriate to their particular situation. It is critical to emphasize that the Project Plan developed in this fashion cannot be limited to listing the statistical procedures that will be employed. An effective Project Plan must also explicitly address procedures for interpreting the results of the reanalysis and procedures for disseminating information about the projects findings, interpretations, and recommendations.

As a final note, it is important to mention that the procedures employed in the South Carolina Project can be adapted to other subject areas besides mathematics. The original project focused on students' mathematics achievement, although students' reading achievement was integrated into the study. A subsequent study in South Carolina focused exclusively on reading, using the mathematics project as a general model.

LIST OF FIGURES

LIST OF TABLES

Improving Mathematics Curriculum and Instruction:

A South Carolina Model For Using Statewide Test Data

by Joseph M. Ryan
Department of Educational Psychology
College of Education
University of South Carolina

Southeastern Educational Improvement Laboratory
Mathematics Improvement Program

INTRODUCTION

The time-tested educational formula of "diagnosis and prescription" is being applied across the state of South Carolina in a project that began in 1983. The project sprang from a concern for providing educators with instructionally useful diagnostic information without using additional instructional time and more testing. When the question of quality diagnostic information first arose, student performance on the state's Basic Skills Assessment Program (BSAP) reading and mathematics tests was leveling off after initial gains. In the early grades, students were still showing year-to-year gains in achievement, but the rate of gain was clearly slowing down. In the middle school, the gains had never been very encouraging at any grade. Some students still were not approaching grade-level standards.

The early achievement gains were, to some degree, related to the development and publication of statewide testing objectives toward which teachers and other educators could specifically focus instruction and curriculum. These objectives were published by the South Carolina Department of Education in a guide entitled *Teaching and Testing Our Basic Skills*. The benefits of this clarification of objectives were short-lived. Additional progress in student achievement seemed to require more detailed identification of students who continued to have problems, as well as the nature of their specific problems.

The additional diagnoses required for continued progress in student achievement could have been obtained from additional testing, but were also available from previously collected statewide test data. This project, which was designed to respond to the need for additional information, focused on the second approach, the reanalysis of currently available statewide test data, for a variety of reasons. In general, statewide test data seemed badly "underutilized," in that they were collected at considerable expense, but were used only once, primarily to classify students. In addition, currently available test data could be examined in complete detail before additional testing was employed; thus, no additional instructional time would be lost to testing. Reanalysis of available data is far less expensive than new testing, since new testing involves test design, development or purchase, distribution, administration, collection, and processing. Reanalysis of existing test data also provides diagnostic information in the framework of the instructional objectives defined in *Teaching and Testing Our Basic Skills*, a framework already familiar to teachers and other educators.

The purpose of this publication is to describe the South Carolina project as a model for other educators interested in taking full advantage of test data that have already been collected. It is important to recognize that the South Carolina model illustrates a broad range of reanalysis procedures, but does not define the specific procedures that would be most appropriate in a given situation. As a general model, the procedures used in the South Carolina project would need to be modified when applied to a particular set of test data from a school, school district, or statewide testing program. Some suggestions about effective ways to use the information in this monograph are provided in the Foreword (page iii).

The South Carolina project has involved much more than the reanalysis of test data, which was only the first step in the process. The interpretation of the reanalysis and the dissemination of project information to mathematics educators around the state were actually the more critical components of the project. These components were important because the purpose of the project was to bring about changes in curriculum and instruction that would have a positive impact on student achievement. With this purpose in mind, merely reanalyzing test data would have not been adequate. It was essential that mathematics educators receive and understand the information from the project so that they could initiate the necessary changes in curriculum and instruction.

This monograph describing the South Carolina project is presented in six sections:

1) Introduction
2) A Conceptual Model of the Project Showing the Relationships Among Mathematics Achievement, Information About Achievement, Factors That Affect Achievement, and Project Outcomes
3) Procedures for Reanalyzing Statewide Test Data
4) Procedures for Interpreting Reanalysis Results in Terms of Factors That Affect Achievement
5) Project Outcomes
6) Summary and Discussion

A CONCEPTUAL MODEL OF THE PROJECT

Showing the relationships among mathematics achievement, information about achievement, factors that affect achievement, and project outcomes

The ultimate value of any reanalysis depends on the extent to which the factors affecting student achievement are modified based on that reanalysis. Technical complexity and detail may be important features of a worthwhile reanalysis. However, for educational purposes, the bottom line is the impact that the reanalysis information has on the factors that enhance students' future learning.

The South Carolina project focused very careful attention on the interplay among student mathematics achievement, information about achievement, the interpretation of information in terms of the factors that affect student achievement, and project outcomes. A conceptual model was developed to show the relationships among the various components of the project. This model was very important for several reasons. First, the conceptual model provided an integrated structure through which all aspects of the project could be coordinated. Second, it showed the interplay among the various components of the project. Third, it provided clear and specific direction to the reanalyses by continually focusing on information that could be used to modify the factors that affect achievement. Reanalyses that were theoretically interesting from a substantive or technical perspective were noted, but omitted because of the focus provided by the model. Fourth, the conceptual model provided a coherent and informative structure for disseminating project information through a variety of media.

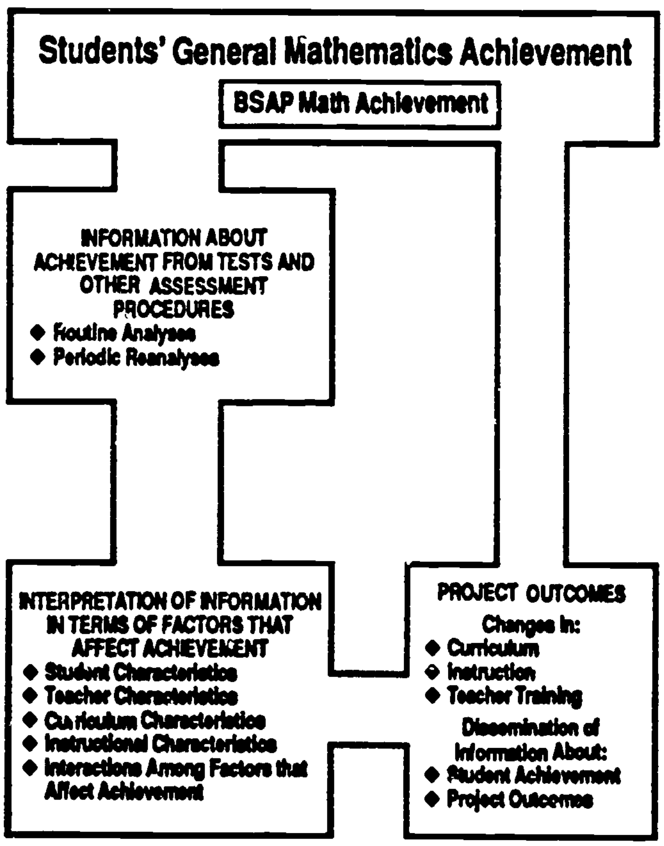

Figure 1. A conceptual model of the project showing the relationships among mathematics achievement, information about achievement, factors that affect achievement, and project outcome

The conceptualmodel with its four major components is depicted in Figure 1 (left). Student achievement is represented in the upper rectangle, information about student achievement is shown in the rectangle below achievement, factors that affect achievement are represented in the interpretation of information box, and, finally, project outcome activities are represented by the rectangle on the right side of the figure.

Student achievement

Student achievement, as conceptualized by the model, has two components: achievement in general and achievement as it relates specifically to the objectives assessed by the BSAP tests. From a curriculum perspective, it is critical to note that most basic skills programs and virtually all minimum competency programs assess only a subset of the skills that constitute an appropriate comprehensive mathematics curriculum. The part-to-whole relationship of basic skills and minimum competency programs to a comprehensive mathematics curriculum must be emphasized. Such constant reminders help minimize the possibility that a mathematics curriculum will shrink to include only those skills assessed by basic skills/minimum competency programs.

Information about achievement

For the project described in this document, information about achievement is focused on student achievement on the BSAP tests. Information about student achievement is provided annually from routine analyses, which are part of each testing cycle, and periodic reanalyses of the type being described here. The routine analyses and periodic reanalyses may interact as they did in the South Carolina project. One of the reanalyses procedures was so useful that it was incorporated into the routine analyses procedures performed each year. This particular procedure and the others used in the reanalyses will be described in detail in the next section, "Procedures for Reanalyzing Statewide Test Data."

It is critical to note that, as a standard practice, information about student achievement should not be limited to test data when conducting a typical student assessment. Other sources of information, such as direct observation, teacher and student interviews, and samples of students' work, would normally be included in conducting a comprehensive assessment of student achievement. The focus of this monograph examines a special case in which test data alone are being examined in assessing student achievement. This specific focus should not be interpreted as an endorsement of assessment procedures that exclude other sources of information as a general procedure.

Interpretation in terms of factors that affect achievement

As seen in Figure 1, information about achievement must be interpreted before it can be educationally useful. Specifically, achievement must be interpreted by an individual or individuals in terms of factors that affect future learning and achievement. With this focus, the essential question becomes, "What does the analysis of student achievement data mean in terms of changes to curriculum, instruction, teacher characteristics, and student characteristics that can facilitate future achievement?"

A special feature of the project is the procedure used to interpret the data analyses and thus to answer this question. A project study group composed of mathematics educators from across the state was convened to interpret the results of the reanalyses. They were especially qualified for this task since the focus of the interpretation was on recommending changes in curriculum, instruction, and teacher preservice and in-service training aimed at improving student achievement in those skills identified as problem areas. The role of teachers and other educators in interpreting research analyses (or in this case reanalyses) differs from the traditional view that sees them as consumers of completed research in which the researchers interpret the analyses. The involvement of teachers and other mathematics educators as the interpreters of the reanalyses is a major component of the project and will be described in detail in the section entitled "Procedures for Interpreting Reanalysis Results in Terms of Factors That Affect Achievement."

Factors affecting achievement

The Conceptual Model shown in Figure 1 (page 3) identifies only these factors affecting achievement: student, teacher, curriculum, and instructional characteristics, as well as interactions among these four factors. This model is a simplification of a far more complex reality in which many factors have an impact on student achievement. These other factors, such as parents' educational level, are not included since they cannot be modified or can be modified only under very special circumstances.

The major emphases of the project were on the characteristics of curriculum, instruction, and teachers' preservice and in-service training that could be changed to have a positive impact on subsequent student achievement. The project study group interpreted the reanalyses with the intention of making recommendations for outcome or implementation activities in curriculum, instruction, and teacher training. A complete list of the recommendations for changes in these three areas is given in the section on interpreting the results of the reanalyses.

Student characteristics, shown as one of the factors affecting student achievement, were viewed as entry characteristics that had to be considered in making recommendations for changes in curriculum, instruction, and teacher training.

Project outcomes

A final component of the project is concerned with project outcomes designed to implement both the recommendations of the project study group and dissemination activities. In general, project outcomes focused on curriculum, instruction, and teacher training. In particular, the State Department of Education has initiated a variety of specific activities in response to the recommendations of the project study group. A second type of project outcome deals with dissemination activities. Dissemination mechanisms used in the South Carolina project included oral presentations, written materials, television broadcasts, and videotaped presentations. These will be described in detail in the section entitled "Project Outcomes."

Summary

The reader will probably recognize that the conceptual model showing the relationships among mathematics achievement, information about achievement, factors affecting achievement, and project outcomes is a "macro" version of the well-understood diagnostic-prescriptive model for instruction. The analysis, reanalysis, and interpretation of mathematics data represent diagnosis, while recommendations for modifications in the factors affecting achievement—such as curriculum, instruction, and teacher training—constitute the prescription. The project outcomes are designed to have an impact on subsequent achievement, which will then become the subject of future reanalyses.

PROCEDURES FOR REANALYZING STATEWIDE TEST DATA

Introduction

The reanalysis of any test data should be performed with three questions in mind: (1) What is the purpose of the reanalysis? (2) Who is the intended audience for the information from the reanalysis? (3) What is the structure of the tests to be reanalyzed?

Purpose

In the South Carolina study, the purposes of the reanalysis were to describe general trends in students' mathematics achievement, to identify specific areas of mathematics in which students continued to have difficulty, and to identify changes in curriculum, instruction, and teachers' preservice and in-service training that might lead to improved achievement in the problem areas identified by the study. In terms of the conceptual model of the project described in the preceding section, the South Carolina study examined BSAP mathematics achievement to extract information about areas in which students had deficiencies, interpreted the information about identified deficiencies, and disseminated the information to mathematics educators. Recommendations were then made by a representative group of educators for changes in curriculum, instruction, and teachers' preservice and in-service training that would have a positive impact on achievement in those areas identified by the reanalyses as problematic.

The study had an explicitly practical and applied orientation. Such an approach places theoretical questions about learning and instruction, as well as technical issues in measurement and statistics, in a particular light: given the purposes of the reanalysis, theoretical and technical questions clearly become secondary to the more practical and applied issues the study was designed to address. A variety of interesting and potentially important theoretical issues came to light during the study. These were noted in some detail and set aside for future research having a different focus.

Intended audience

Mathematics educators were the intended audience for the reanalysis of the South Carolina BSAP mathematics data base. The nature of this intended audience, like the purposes of the study, had major implications for the reanalysis that was performed, the way the report of the reanalysis was written, and the procedures used for interpretation and dissemination of the results. The intended audience was viewed as a professional group committed to and experienced in teaching mathematics, but usually without any particular interest in theoretical or technical issues. The nature of the intended audience led to reanalysis procedures and a report that were descriptive in nature, with little or no discussion of theoretical explanations for students' achievement patterns or of technical details of the reanalysis procedures. Procedures used for dissemination were chosen specifically to suit the intended audience and will be discussed in detail in a later section.

Structure of South Carolina BSAP mathematics tests

The South Carolina BSAP tests in grades 1, 2, 3, 6, and 8 are composed of 30 items; 6 items assess each of 5 objectives. The BSAP objectives include: Concepts (C), Operations (O), Geometry (G), Measurement (M), and Problem Solving (P). All objectives except Problem Solving are divided into subskills assessed at various grade levels. The chart on the next page (Figure 2) lists the subskills for each objective with solid circles indicating the grades at which each is assessed. Subskills are referred to by the first letter of their objective and by the number indicating the subskill (e.g., C4 is the Concept Objective, Subskill 4, "Determining Place Value").

The BSAP tests are designed to measure student achievement at the end of the school year relative both to end-of-year grade-level appropriate skills and to a grade-level standard of achievement. The BSAP tests are thus grade-level-appropriate basic skills tests, not minimum competency tests. The results of the BSAP tests have generally been used for two purposes. First, student performance is reported in terms of whether or not grade-level standards have been reached. Second, if the performance for an objective is deficient, the student's performance for that objective is noted. The deficiency level for each objective is based on the standard set at the total test level.

The BSAP tests are equated within each grade over the years, but not across the grades. Thus, comparisons within a grade over time are appropriate, but between-grade comparisons are not appropriate.

The structure of the BSAP test, including the objectives and subskills, suggests four possible levels for reanalysis:

1. Total test score.
2. Objectives.
3. Subskills.
4. Items.

It is very important to describe historical trends in student performance relative to the total test score, even though this information might not be directly useful in identifying areas in which students are deficient. There is value in documenting the broader context of student achievement as a backdrop for examining student deficiencies in specific detail. A study that excludes the larger context and examines only specific student deficiencies can leave the erroneous impression that, overall, students are doing poorly when, in fact, they are doing very well overall and poorly only in a small number of very specific areas of mathematics. This issue is similar to a messenger who prefaces a message with the comment, "I have some good news and some bad news." It is important that studies of student deficiencies describe student achievement in the broader context, generally the good news, before describing information about student weaknesses, or the bad news.

Reanalyses performed at the total test score level

Total test scores are used to document students' general performance or overall growth relative to grade level standards over a fixed period of time. In the South Carolina case study, student achievement in mathematics from 1981 to 1984 was examined. The procedures employed could be used over any in-

OBJECTIVE: CONCEPTS

Subskills	1	2	3	4	5	6	7	8	9	10	11	12
1. Counting	●	●	●	○								
2. Identifying Equivalencies	●	●	●	○	○	●	○	●	○	○	●	○
3. Establishing Relationships	●	●	●	○	○	●	○	●	○	○	●	○
4. Determining Place Value	●	●	●	○	○	●	○	●	○	○	●	○
5. Interpreting Tables and Graphs	○	●	●	○	○	●	○	●	○	○	●	○
6. Recognizing Square Roots									○	○	●	○
7. Using Formulas									○	○	●	○

OBJECTIVE: MEASUREMENT

Subskills	1	2	3	4	5	6	7	8	9	10	11	12
1. Identifying Units of Measurement	●	●	●	○	○	●	○	●	○	○	●	○
2. Estimation		○	●	○	○	●	○	●	○	○	●	○
3. Using Measuring Devices	●	●	●	○	○	●	○	●	○	○	●	○
4. Conversions and Operations		○	●	○	○	●	○	●	○	○	●	○
5. Scale Drawings				○	○	●	○	●	○	○	●	○

OBJECTIVE: OPERATIONS

Subskills	1	2	3	4	5	6	7	8	9	10	11	12
1. Addition	●	●	●	○	○	●	○	●	○	○	●	○
2. Subtraction	●	●	●	○	○	●	○	●	○	○	●	○
3. Multiplication		●	●	○	○	●	○	●	○	○	●	○
4. Division			●	○	○	●	○	●	○	○	●	○

OBJECTIVE: PROBLEM SOLVING

Objective	1	2	3	4	5	6	7	8	9	10	11	12
The student can solve problems involving the use of mathematics	●	●	●	○	○	●	○	●	○	○	●	○

OBJECTIVE: GEOMETRY

Subskills	1	2	3	4	5	6	7	8	9	10	11	12
1. Identification	●	●	●	○	○	●	○	●	○	○	●	○
2. Comparison	●	●	●	○	○	○	○	●	○	○	●	○
3. Application				○	○	●	○	●	○	○	●	○

Figure 2. BSAP objectives and subskill structure within objectives

terval of time. In the South Carolina study, students' overall performance in mathematics and reading was examined to describe their general overall performance as completely as possible. Students were classified into four categories, based on their performance in mathematics and reading. These categories were:

+R+M = Students who reached both the reading and mathematics standard.

-R+M = Students who did not reach the standard in reading but reached the standard in mathematics.

+R-M = Students who reached the standard in reading but not in mathematics.

-R-M = Students who did not reach the standard in both reading and mathematics.

Besides providing a broad description of students' overall achievement, these four categories identify groups of students for whom different types of remedial efforts might be appropriate. For example, there are two types of students who do not reach the mathematics standard, the +R-M and -R-M students. Remedial efforts for the +R-M students could be presented in a fairly standard grade- level format because the students appear to be able to read grade- level material. However, remedial efforts for the -R-M students could not use regular grade-level materials because these students apparently would have difficulty reading them.

Student achievement at the total test-score level can be usefully described in three ways:

1) Changes in the percentage of students in the four categories over time, displayed both in tables and figures.
2) Variation in the rate of change in the percentage of students in the various categories from one year to the next, displayed in both tables and figures.
3) Changes in the percentage of students doing either extremely well or very poorly on the test over time, displayed both in tables and figures.

Changes in the percentage of students in the four categories over time

The most obvious and standard format for describing students' overall achievement is to display changes in the percentage of students in the four categories over time. This information is displayed in Table 1 (page 11) for the four categories from 1981 to 1984. Of particular interest is the increase in the percentage of +R+M students and the decrease in the percentage of -R-M students from 1981 to 1984, as shown in Table 2 (page 11). The data in Table 2 show an average increase of 12.6 percent in the +R+M category and an average decrease of 10.6 percent in the -R-M category. The total shift in the student population in the +R+M and -R-M categories combined over the five grades is 23.3 percent from 1981 to 1984. It is important to point out that discussions of student deficiencies should be placed in the context of this overall positive shift of almost 25 percent.

Grade 1	All Students	+R+M	-R+M	+R-M	-R-M
1984	49,572	36,304 (73.2%)	3,937 (8.0%)	3,364 (6.8%)	5,967 (12.0%)
1983	46,440	31,207 (67.2%)	4,015 (8.7%)	3,782 (8.7%)	7,436 (16.0%)
1982	46,561	28,076 (60.4%)	3,379 (7.3%)	5,409 (11.6%)	9,652 (20.8%)
1981	47,412	28,245 (59.6%)	4,190 (8.8%)	4,908 (10.4%)	10,069 (21.2%)
Grade 2	**All Students**	**+R+M**	**-R+M**	**+R-M**	**-R-M**
1984	42,590	29,698 (69.7%)	5,415 (12.7%)	2,469 (5.8%)	5,008 (11.8%)
1983	43,239	27,312 (63.2%)	5,552 (12.8%)	3,124 (7.2%)	7,251 (16.8%)
1982	44,226	24,641 (55.7%)	3,865 (8.7%)	5,660 (12.8%)	10,060 (22.8%)
1981	45,491	24,728 (54.4%)	6,865 (15.1%)	3,278 (7.2%)	10,620 (23.3%)
Grade 3	**All Students**	**+R+M**	**-R+M**	**+R-M**	**-R-M**
1984	42,840	29,402 (68.6%)	4,447 (10.4%)	3,554 (8.3%)	5,437 (12.7%)
1983	43,957	28,246 (64.3%)	4,465 (10.2%)	4,939 (11.2%)	6,307 (14.4%)
1982	45,207	25,555 (56.5%)	5,062 (11.2%)	5,624 (12.4%)	8,966 (19.8%)
1981	46,952	24,743 (52.7%)	4,065 (8.7%)	6,952 (14.8%)	11,192 (23.8%)
Grade 6	**All Students**	**+R+M**	**-R+M**	**+R-M**	**-R-M**
1984	46,515	23,015 (49.5%)	3,433 (7.4%)	7,176 (15.4%)	12,891 (27.7%)
1983	49,987	24,093 (48.2%)	4,117 (8.2%)	6,458 (13.0%)	15,319 (30.7%)
1982	48,127	21,827 (45.4%)	2,526 (5.3%)	8,190 (17.0%)	15,584 (32.4%)
1981	46,546	18,764 (40.3%)	3,278 (7.0%)	6,832 (14.7%)	17,672 (38.0%)
Grade 8	**All Students**	**+R+M**	**-R+M**	**+R-M**	**-R-M**
1984	47,135	21,530 (45.7%)	3,867 (8.2%)	6,807 (14.4%)	14,931 (31.7%)
1983	45,969	16,858 (36.7%)	2,468 (5.4%)	9,032 (19.7%)	17,601 (38.2%)
1982	45,504	16,033 (35.2%)	2,452 (5.4%)	7,744 (17.0%)	19,275 (42.4%)
1981	46,667	17,024 (36.5%)	2,901 (6.2%)	6,795 (14.6%)	19,947 (42.7%)

Code

+R+M: Students Passing Both Tests
-R+M: Students Failing Reading, Passing Mathematics
+R-M: Students Passing Reading, Failing Mathematics
-R-M: Students Failing Both Tests

Table 1. Number (and percentage) of students meeting or failing to meet BSAP standards in reading and mathematics

Grade	Increased Percentage of Students in the +R+M Group	Decreased Percentage of Students in the -R-M Group	Total Shift in Student Population
1	+13.6	-9.2	22.8
2	+15.3	-11.5	26.8
3	+15.9	-11.1	27.0
6	+9.2	-10.3	19.5
8	+9.2	-11.0	20.2
Average	+12.6	-10.6	23.3

Table 2. Increased and decreased percentage of students in the +R+M and -R-M groups from 1981 to 1984

21

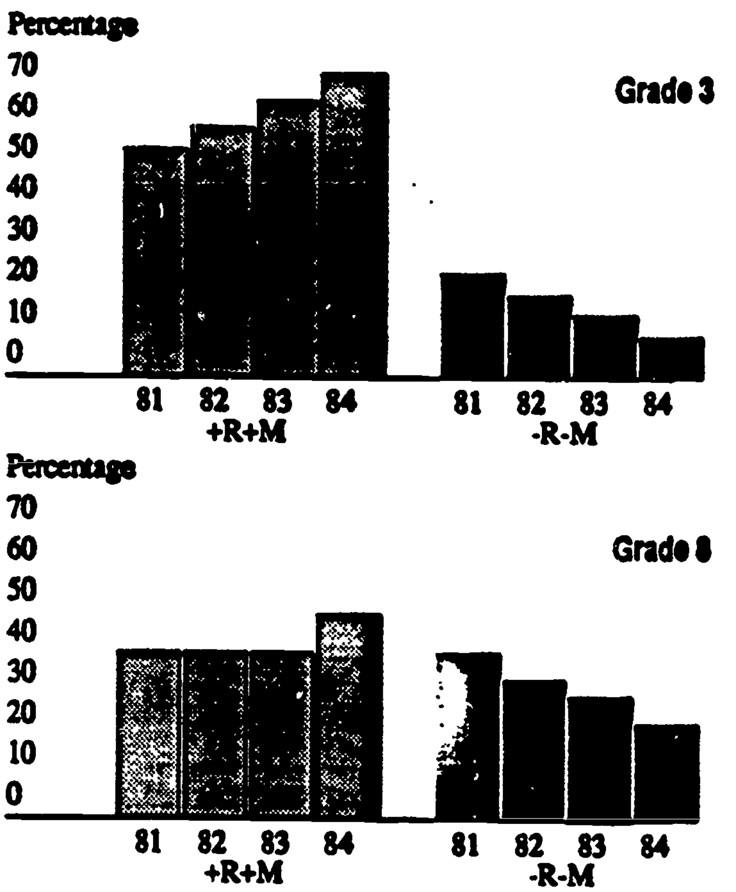

Figure 3. Percentage of +R+M and -R-M students in grades 3 and 8, from 1981 to 1984

Changes in the percentage of students in the four categories can also be shown graphically, using a bar chart or histogram. For purposes of illustration, a histogram is presented for +R+M student and the -R-M students in grades 3 and 8 respectively in Figure 3 (left). These types of figures are extremely helpful in describing changes over time to nontechnical audiences who may not be comfortable reading data tables.

Rate of change from one year to the next

It is critical to recognize and describe the year-to-year rates of change in the percentage of students in the various student categories over time. Using the South Carolina example, the changes in the percentage of +R+M students from 1981 to 1982, from 1982 to 1983, and from 1983 to 1984 for grades 1, 2, 3, 6, and 8 are shown in Table 3 (below). The information in Table 3 clearly shows that the change in the percentage of students in the +R+M category from 1981 to 1984 was not equally distributed across the years. In grades 1, 2, and 3, there was a modest increase from 1981 to 1982, a major shift from 1982 to 1983, and another modest increase from 1983 to 1984. In these early grades, there is still evidence of an increasing percentage of students in the +R+M category from 1983-1984, but the rate of increase is decreasing. This type of situation is analogous to the difference between speed and acceleration. The data show that there is consistent improvement from year to year (e.g., there is a positive rate of speed), but the rate of improvement (e.g., the acceleration) is decreasing. The decreasing rate of improvement is an important trend to document, because it has major policy implications. It is important to recog nize that the trends can vary at different grades. For example, the trends

	Grade				
Years	1	2	3	6	8
1983-1984	6.0	6.5	4.3	9.0	1.3
1982-1983	6.8	7.5	7.8	1.5	2.8
1981-1982	0.8	1.3	3.8	-1.3	5.1

Table 3. Year-to-year increases in the percentage of +R+M students in grades 1,2, 3, 6, and 8, from 1981 to 1984

over time are quite different for grades 6 and 8.

The year-to-year change in the percentage of students in the +R+M category can also be displayed graphically as seen in Figure 4 (below) for grades 1, 2, 3, 6, and 8. The year-to-year shifts in grades 1, 2, and 3 are clearly very similar, whereas the shifts in grades 6 and 8 are quite different.

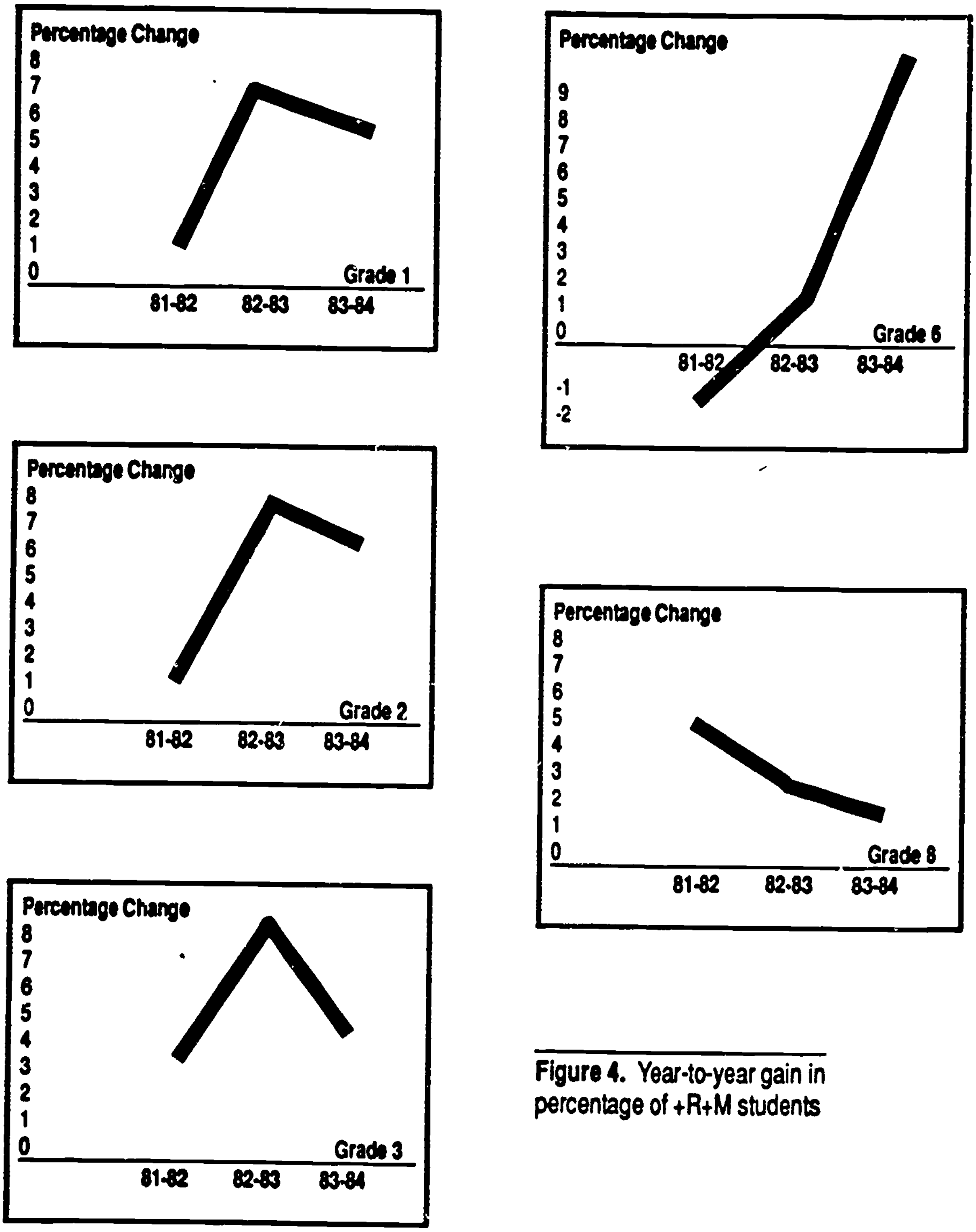

Figure 4. Year-to-year gain in percentage of +R+M students

Change in the percentage of students doing extremely well or very poorly

A final way to view student performance on the total test compares the percentage of students doing extremely well or very poorly in 1981 in contrast to 1984. The terms "extremely well" and "very poorly" are somewhat vague and need to be clarified. The basic question involves determining whether the students who passed both the reading and mathematics standard (+R+M) in 1984 passed the mathematics standard by as much as did their 1981 counterparts. Likewise, is is informative to determine whether students who failed to reach both the reading and mathematics standards (-R-M) in 1984 are any closer to reaching the mathematics standard than their 1981 counterparts.

The importance of examining the performance of students at the extreme ends of the achievement scale derives from a concern that a basic skills program may focus instructional efforts at the middle range of students. An increase in the percentage of students reaching grade-level standard could be achieved very effectively by ignoring both the most capable students who are going to reach standard with little help and the very low-ability students who may not reach the standard even with considerable assistance. An instructional effort with these emphases might increase the proportion of students reaching grade-level standard, but it would do so at the expense of the most capable students and the students most in need of assistance. Such an approach is unacceptable, and statewide test data should be examined to determine if trends reflecting this approach to instruction are apparent.

The standard error of measure (S.E.M.) is used to make this determination. The S.E.M. is a number of points on a scale that is used to construct a confidence interval around a test score. An interval of plus and minus one S.E.M. around a score gives a 68 percent confidence interval; e.g., there is a probability of .68 that a person's true score falls within the interval. Plus and minus two S.E.M.s gives a 95 percent confidence interval; e.g., there is a probability of .95 that a person's true score falls within the interval.

The use of S.E.M.s to examine student performance on the extreme ends of the achievement scale can be illustrated by considering six hypothetical students. The performance of these hypothetical students is shown in Figure 5 on the next page. Student A passes the standard by greater than two S.E.M.s, while Student B passes the standard by greater than one S.E.M., but less than two S.E.M.s. Both of these students have passed the standard, but Student A has passed it at a higher level than Student B. Student C also passes the standard, but does so by less than one S.E.M., performing less well than the other two students.

On the other side of the standard, consider Students X, Y, and Z, all of whom fail to make the standard. Student X is less than one S.E.M. below standard; Student Y is greater than one, but less than two S.E.M.s below standard; and Student Z is greater than two S.E.M.s below standard. All three students have failed to make the standard. However, Student X did not miss the standard by as much as Student Y, who was closer to making the standard than Student Z.

The approach illustrated with these six hypothetical students is used to determine if the +R+M students in 1984 are passing the mathematics standards by as much as the 1981 +R+M students and also to determine if the -R-M students in 1984 are any closer to making the mathematics standard than the 1981 -R-M students. The distribution in percentages of students' scores on the BSAP mathematics tests according to their distance above or below the standard for each test is given in Tables 4 through 8, beginning on page 32. Each page displays the information for one grade, with the score distribution for the 1984, 1983, 1982, and 1981 tests presented from the top to the bottom of the page. The table for each grade presents the percentage of students in each of five groups: all students combined and +R+M, +R-M, -R+M, and -R-M students by year from 1981 to 1984. The distances of the students' scores from the standard are expressed in standard errors of measurement (S.E.M.) arranged in the following order from the top to the bottom of the table for each year:

a) Greater than two S.E.M.s above the standard (abbreviated: GT.2 above).
b) Greater than one and less than two S.E.M.s above the standard (abbreviated: GT.1 and LT.2).
c) Scoring at the standard or above it by less than one S.E.M. (abbreviated: at or LT.1).
x) Below the standard by less than one S.E.M. (abbreviated: below by LT.1).
y) Greater than one, but less than two S.E.M.s below the standard (abbreviated: GT.1, but LT.2).
z) Greater than two S.E.M.s below the standard (abbreviated: GT.2).

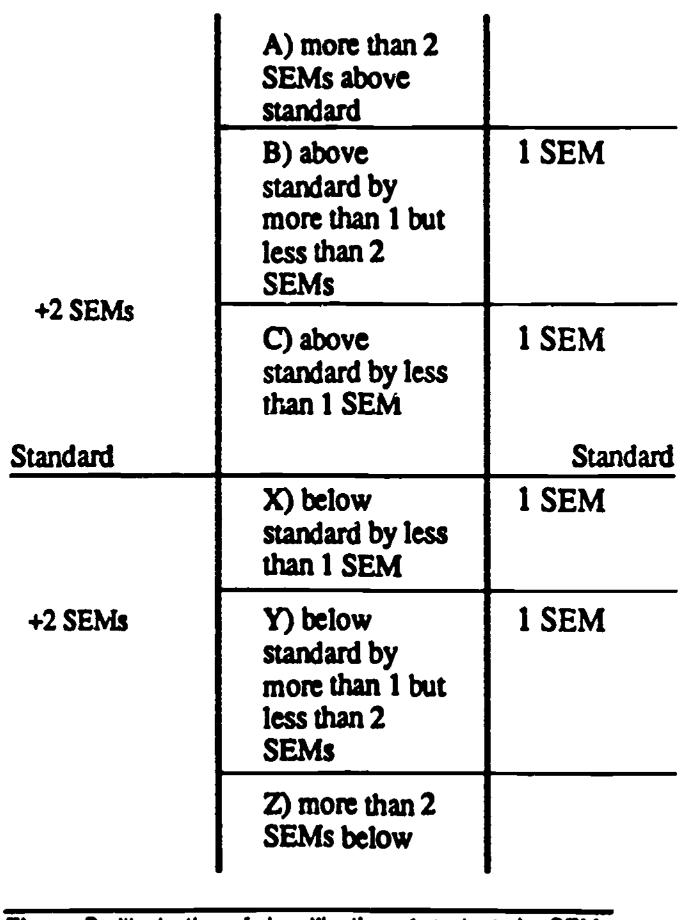

Figure 5. Illustration of classification of students by SEMs

The percentage of all students who passed or failed the standard is listed in the middle row for each year. The total number and percentages of students in each group are listed in the last two rows for each year.

To illustrate the use of these tables, consider grade 1 for the year 1981 at the bottom of Table 4 (p. 32). The total number of first grade students (All Ss column) taking the test is 47,412 as listed on the next to last line on the page. There were 28,245 students or 59.6 percent of all first graders in the +R+M group. Of these, 41.6 percent passed the mathematics standard by greater than two (GT.2 above) S.E.M.s, 38.1 percent of the students passed by greater than one and less than two (GT.1 and LT.2) S.E.M.s, and 20.4 percent were at or above the standard by less than one (LT.1) S.E.M.

The corresponding group of students for 1984 is found at the top of Table 4. In 1984, there were 36,304 students or 73.2 percent of all first graders in the +R+M group. This is an increase of 13.6 percent over 1981. Of these 36,304 in the 1984

1984 SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	38.2	50.7	12.8		
b) GT.1 & LT.2	28.6	34.9	38.4		
c) at or LT.1	14.4	14.4	48.8		
Pass/Fail	81.2/18.8				
x) below by LT.1	4.4			38.1	15.4
y) GT.1 but LT.2	6.0			38.4	27.8
z) GT.2 below	8.4			23.5	56.8
Total	49,572	36,304	3,937	3,364	5,967
Percent of total	100.0	73.2	8.0	6.8	12.0

1983 SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	30.8	44.6	9.2		
b) GT.1 & LT.2	27.8	36.8	35.4		
c) at or LT.1	17.3	18.6	55.3		
Pass/Fail	75.9/24.2				
x) below by LT.1	5.6			39.0	15.3
y) GT.1 but LT.2	7.6			39.1	27.6
z) GT.2 below	11.0			21.9	57.1
Total	46,440	31,207	4,015	3,782	7,436
Percent of total	100.0	67.2	8.7	8.1	16.0

1982 SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	16.0	25.9	4.6		
b) GT.1 & LT.2	31.6	48.2	34.5		
c) at or LT.1	20.1	25.9	61.1		
Pass/Fail	67.7/32.4				
x) below by LT.1	6.9			36.2	13.1
y) GT.1 but LT.2	9.6			38.6	24.5
z) GT.2 below	15.9			25.2	62.4
Total	46,561	28,076	3,379	5,409	9,652
Percent of total	100.0	60.4	7.3	11.6	20.8

1981 SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	25.5	41.6	9.0		
b) GT.1 & LT.2	25.8	38.1	35.9		
c) at or LT.1	17.0	20.4	55.1		
Pass/Fail	68.3/31.6				
x) below by LT.1	6.0			33.3	12.2
y) GT.1 but LT.2	9.2			38.0	24.6
z) GT.2 below	16.4			28.7	63.2
Total	47,412	28,245	4,190	4,908	10,069
Percent of total	100.0	59.6	8.8	10.4	21.2

Table 4. Distribution of students by group membership on BSPA mathematics. Grade 1

1984

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	40.8	55.0	19.6		
b) GT.1 & LT.2	26.9	31.5	39.4		
c) at or LT.1	14.7	13.5	41.0		
Pass/Fail	82.4/17.6				
x) below by LT.1	4.6			39.9	19.1
y) GT.1 but LT.2	6.2		39.0	33.3	
z) GT.2 below	6.8			21.1	47.6
Total	42,590	29,698	5,415	2,469	5,008
Percent of total	100.0	69.7	12.7	5.8	11.8

1983

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	33.6	50.1	15.6		
b) GT.1 & LT.2	25.9	33.5	36.7		
c) at or LT.1	16.5	16.4	47.7		
Pass/Fail	76.0/24.0				
x) below by LT.1	5.6			37.9	17.3
y) GT.1 but LT.2	7.8			38.6	29.9
z) GT.2 below	10.6			23.5	52.8
Total	43,239	27,312	5,552	3,124	7,251
Percent of total	100.0	63.2	12.8	7.2	16.8

1982

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	23.7	40.6	12.2		
b) GT.1 & LT.2	23.2	36.2	34.4		
c) at or LT.1	17.6	23.2	53.4		
Pass/Fail	64.5/35.6				
x) below by LT.1	7.0			32.8	12.1
y) GT.1 but LT.2	10.4			37.6	24.5
z) GT.2 below	18.2			29.7	63.4
Total	44,226	24,641	3,865	5,660	10,060
Percent of total	100.0	55.7	8.7	12.8	22.8

1981

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	27.4	46.5	13.9		
b) GT.1 & LT.2	25.2	35.7	38.2		
c) at or LT.1	16.9	17.8	47.9		
Pass/Fail	69.5/30.5				
x) below by LT.1	6.4			37.2	15.8
y) GT.1 but LT.2	9.1			39.4	27.0
z) GT.2 below	15.0			23.4	57.2
Total	45,491	24,728	6,865	3,278	10,620
Percent of total	100.0	54.4	15.1	7.2	23.3

Table 5. Distribution of students by group membership on BSPA mathematics. Grade 2

1984

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	39.1	54.3	17.8		
b) GT.1 & LT.2	19.5	24.4	26.9		
c) at or LT.1	20.4	21.3	55.3		
Pass/Fail	79.0/21.0				
x) below by LT.1	8.6			55.6	31.1
y) GT.1 but LT.2	5.6		26.9	26.4	
z) GT.2 below	6.8			17.5	42.5
Total	42,840	29,402	4,447	3,554	5,437
Percent of total	100.0	68.6	10.4	8.3	12.7

1983

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	36.5	53.7	19.7		
b) GT.1 & LT.2	17.3	23.0	24.7		
c) at or LT.1	20.7	23.3	55.6		
Pass/Fail	74.5/25.5				
x) below by LT.1	9.9			50.7	28.9
y) GT.1 but LT.2	6.7			26.7	25.7
z) GT.2 below	9.1			22.6	45.4
Total	43,957	28,246	4,465	4,939	6,307
Percent of total	100.0	64.3	10.2	11.2	14.4

1982

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	26.6	44.4	13.5		
b) GT.1 & LT.2	17.1	25.5	23.6		
c) at or LT.1	24.0	30.1	62.9		
Pass/Fail	67.7/32.3				
x) below by LT.1	12.8			54.3	30.5
y) GT.1 but LT.2	9.2			29.1	28.1
z) GT.2 below	10.3			16.6	41.4
Total	42,207	25,555	5,062	5,624	8,966
Percent of total	100.0	56.5	11.2	12.4	1[illegible]8

1981

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	24.2	43.9	11.8		
b) GT.1 & LT.2	15.8	26.1	24.2		
c) at or LT.1	21.4	30.0	64.0		
Pass/Fail	61.4/38.7				
x) below by LT.1	12.1			47.8	20.9
y) GT.1 but LT.2	10.0			28.3	24.5
z) GT.2 below	16.6			23.9	54.6
Total	45,491	24,728	6,865	3,278	10,620
Percent of total	100.0	52.7	8.7	14.8	23.8

Table 6. Distribution of students by group membership on BSPA mathematics. Grade 3

1984

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	29.5	57.1	17.0		
b) GT.1 & LT.2	10.5	18.0	21.5		
c) at or LT.1	16.9	24.9	61.5		
Pass/Fail	56.9/43.1				
x) below by LT.1	11.1			41.2	16.9
y) GT.1 but LT.2	10.4		29.6	21.3	
z) GT.2 below	21.6			29.2	61.8
Total	46,515	23,015	3,433	7,176	12,891
Percent of total	100.0	49.5	7.4	15.4	27.7

1983

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	22.9	45.8	9.4		
b) GT.1 & LT.2	16.5	29.5	27.8		
c) at or LT.1	17.1	24.7	62.8		
Pass/Fail	56.5/43.6				
x) below by LT.1	11.4			45.4	18.0
y) GT.1 but LT.2	15.6			38.6	34.7
z) GT.2 below	16.6			16.0	47.3
Total	49,987	24,093	4,117	6,458	15,319
Percent of total	100.0	48.2	8.2	13.0	30.7

1982

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	26.6	56.7	16.3		
b) GT.1 & LT.2	8.9	17.4	19.2		
c) at or LT.1	15.2	25.9	64.5		
Pass/Fail	50.7/49.4				
x) below by LT.1	10.8			39.3	12.6
y) GT.1 but LT.2	10.7			29.0	17.9
z) GT.2 below	27.9			31.8	69.5
Total	48,127	21,827	2,526	8,190	15,584
Percent of total	100.0	45.4	5.3	17.0	32.4

1981

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	22.4	52.8	15.7		
b) GT.1 & LT.2	9.5	19.7	21.7		
c) at or LT.1	15.5	27.5	62.7		
Pass/Fail	47.4/52.7				
x) below by LT.1	10.7			39.7	12.7
y) GT.1 but LT.2	10.4			28.0	16.5
z) GT.2 below	31.6			32.3	70.8
Total	46,546	18,764	3,278	6,832	17,672
Percent of total	100.0	40.3	7.0	14.7	38.0

Table 7. Distribution of students by group membership on BSPA mathematics. Grade 6

1984

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	23.7	49.8	11.9		
b) GT.1 & LT.2	13.9	25.6	26.9		
c) at or LT.1	16.2	24.6	61.2		
Pass/Fail	53.9/46.1				
x) below by LT.1	11.7			42.5	17.4
y) GT.1 but LT.2	17.5		41.0	36.6	
z) GT.2 below	17.0			16.5	46.0
Total	47,135	21,530	3,867	6,807	14,931
Percent of total	100.0	45.7	8.2	14.4	31.7

1983

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	16.5	43.3	10.4		
b) GT.1 & LT.2	11.3	27.1	24.9		
c) at or LT.1	14.3	29.6	64.7		
Pass/Fail	42.1/57.9				
x) below by LT.1	11.5			35.3	12.0
y) GT.1 but LT.2	20.3			42.3	31.3
z) GT.2 below	26.1			22.4	56.7
Total	45,969	16,868	2,468	9,032	17,601
Percent of total	100.0	36.7	5.4	19.7	38.2

1982

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	18.2	49.6	13.4		
b) GT.1 & LT.2	10.0	24.5	25.9		
c) at or LT.1	12.4	25.9	60.9		
Pass/Fail	40.6/60.4				
x) below by LT.1	9.8			32.4	10.0
y) GT.1 but LT.2	18.3			40.3	27.0
z) GT.2 below	31.3			27.3	63.0
Total	45,504	16,033	2,452	7,744	19,275
Percent of total	100.0	35.2	5.4	17.0	42.4

1981

SEM above or below	All Students	+R+M	-R+M	+R-M	-R-M
a) GT.2 above	23.9	61.8	21.9		
b) GT.1 & LT.2	6.9	15.6	19.6		
c) at or LT.1	11.9	22.5	58.5		
Pass/Fail	42.7/57.2				
x) below by LT.1	8.9			32.8	9.8
y) GT.1 but LT.2	10.5			28.8	14.8
z) GT.2 below	37.8			38.4	75.3
Total	46,667	17,024	2,901	6,795	19,947
Percent of total	100.0	36.5	6.2	14.6	42.7

Table 8. Distribution of students by group membership on BSPA mathematics, Grade 8

+R+M group, 50.7 percent passed the mathematics standard by greater than two S.E.M.s. This is an increase of 9.1 percent over 1981. Thus, not only is there an increase of 13.6 percent in the percentage of students in the +R+M group from 1981 to 1984, but also within this group, there is an increase of 9.1 percent of +R+M students who exceed the standard by more than two S.E.M.s.

A similar type of analysis can be conducted for the -R-M groups for 1981 and 1984. Again regarding first grade students, Table 4 shows that in 1981 there were 10,069 or 21.2 percent of all first graders in the -R-M group. In 1984, there were 5,967 or 12.0 percent in the same group. Thus, there has been a decrease of 9.2 percent in the number of all students in the -R-M group. In 1981, 63.2 percent of the students in the -R-M group failed the standard by more than two S.E.M.s, whereas in 1984, only 56.8 percent in this group failed by more than two S.E.M.s. Thus, not only is there a decrease of 9.2 percent in the percentage of students in the -R-M group, but also within this group, there is a decrease of 6.4 percent in the number of students failing the standard by greater than two S.E.M.s.

A similar analysis can be performed for students in grades 2, 3, 6, and 8, based on the information in the appropriate tables. A summary of these analyses is given in Table 9 (below). For each grade, this table shows the change in the percentage of students in the +R+M group meeting the mathematics standard by greater than two

S.E.M.s and the change in percentage of students in the -R-M group failing the standard by greater than two S.E.M.s.

The information in Table 9 indicates that there is an increase in the percentage of +R+M students passing the mathematics standard by greater than two S.E.M.s in grades 1, 2, 3, and 6, but there is a decrease of 12.0 percent in grade 8. In the -R-M group, there is a decrease in the percentage failing the mathematics standard by greater than two S.E.M.s in all grades. The change in the -R-M group at grade 8 stands out, with the substantial reduction of -29.3 percent in the number failing the standard by greater than two S.E.M.s.

Grade	1981-84 Change in the Percentage of +R+M Students Meeting the Standard by Greater than Two SEMs	1981-84 Change in the Percentage of -R-M Students Failing the Standard by Greater than Two SEMs
1	+9	-6.4
2	+8.5	-9.6
3	+10.4	-12.1
6	+4.3	-9.0
8	-12.0	-29.3

Table 9. 1981-84 changes in the percentage of students passing or failing the mathematics standard by greater than two SEMs

The discussion of student performance at the extreme ends of the achievement scale demonstrates that not only did more students reach the reading and mathematics standards in 1984 compared to 1981, but that an increased percentage of 1984 +R+M students surpassed the mathematics standard by more than two S.E.M.s. At the other end of the scale, fewer students not only failed to reach both the reading and mathematics standards in 1984 compared to 1981, but a decreased percentage failed by two S.E.M.s. Thus, more students passed, and those who still failed moved closer to the grade-level standards. In short, the data indicate that students at the extremes of the achievement scale are not being ignored, since they continue to show progress in 1984 compared to their 1981 counterparts.

Reanalyses performed at the objective level

As mentioned previously, BSAP mathematics tests are composed of 30 items with 6 items measuring each of 5 objectives. Except for Problem Solving, the five objectives are further divided into subskills: Geometry has three subskills, Operations has four subskills, Measurement has five subskills, and Concepts has seven subskills. Not all subskills are measured at each grade. A listing of the subskills and the grades at which each is measured is contained in Figure 2 (page 9).

The original plan for the reanalyses included a description of student achievement on the five major objectives. The analyses were performed as planned, but turned out to be uniformative and, in some cases, misleading. These problems arose because the subskills within each of the five objectives differed widely in their difficulty. For example, at grade 1, there are 11 subskills reflected on the test. Among the 11 subskills are Concepts Subskills 1, 2, 3, and 4. Counting (C1) was the easiest subskill for the students, whereas Determining Place Value (C4) was the most difficult. Identifying Equivalencies (C2) and Establishing Relationships (C3) tended to be easy, like C1. Describing students' overall performance on these Concepts Objectives is problematic, because on one subskill students had their best performance (C1), while on another subskill students had their worst performance (C4). If student performance is averaged over all subskills, the result is misleading. The average would suggest that Concepts is much easier than C4 actually was and, at the same time, would suggest that Concepts is more difficult than C1 and, to a lesser degree, C2 and C3.

The resolution of this problem was to recognize that the objectives were not appropriate units of analysis for providing useful information about student achievement. They lack the general meaning that performance on the total test contains and fail to reflect the diversity of subskills that define the objective. Given the purposes of the study and the intended audience, the analyses of student performance at the objective level were determined to be of little use and were therefore abandoned.

Reanalyses performed at the subskill level

The relationship of subskill and item-level analysis

Reanalyses of student achievement at the subskill level were extremely useful in identifying specific areas of mathematics in which students appeared to have deficiencies. Reanalyses at the item level were a valuable follow-up and supplement to the subskill-level analyses. In describing student deficiencies, information from both the subskill- and item-level analyses was combined because the information was most useful when considered together. The procedures for performing the reanalyses at the subskill level are described in this section. The application of these procedures is illustrated using BSAP data for grades 1, 3, and 8. Item-level information is incorporated into these examples, even though the procedures for performing the item-level analyses are not described until the next section.

Procedures for subskill analysis

Subskill deficiencies were detected by combining information on individual items used to assess a subskill from all four years. In this procedure, the proportion of students answering each item correctly was determined. This proportion is referred to as the *item difficulty*. An item difficulty of .20 means that 20 percent of the students answered the item correctly on the test. Item A, with a difficulty of .20, is considered harder than Item B, with a difficulty of .70, because a smaller proportion of students was able to answer Item A correctly. The second step in the procedure involved combining the item difficulties for all items that assess each subskill from all years and then calculating the average difficulty of the items on each subskill. The average difficulty of the items on a subskill is the *subskill difficulty*. A subskill difficulty of .95 would indicate a very easy subskill since, on average, 95 percent of the students correctly answered the items assessing that subskill from 1981 to 1984. In contrast, a subskill difficulty of .40 would indicate a very difficult subskill since, on average, 40 percent of the students correctly answered the items assessing that subskill from 1981 to 1984. Subskill difficulties were calculated for students in each of the four groups (+R+M, -R+M, +R-M, and -R-M), for all students who did not meet the mathematics standard, and for all students combined.

The subskill difficulty was used in two ways to detect subskill deficiencies. First, for students who failed to reach the mathematics standard, the most difficult subskills were identified at each grade level. These subskills are those on which children who failed to reach the mathematics standard may need particular assistance.

The second procedure for detecting subskill deficiencies involved comparing the subskill difficulties of students who passed and students who failed to pass the mathematics standard. The difference between the subskill difficulties for students passing and failing the mathematics standard indicates the gap in learning between the mathematics standard need additional instruction.

To illustrate the use of the subskill discrimination, suppose that the subskill difficulty based on all students' meeting the mathematics standard (+R+M and -R+M) is .87. This means that 87 percent of the students passing the mathematics standard were able to answer correctly items assessing this subskill. Suppose, however, that the subskill difficulty based on all students who failed to meet the mathematics standard (+R-M and -R-M) is only .52. The differences between the subskill difficulties for these two groups is (.87 - .52) = .35. Thus the discrimination of .35 on this subskill means that 35 percent more students who met the standard answered these subskill items correctly.

The subskill analysis is presented for grades 1, 3, and 8 to illustrate the use of these procedures. The information for each of these grades is presented in narrative form with the supporting data in an accompanying table and graph. The tables and graphs will be explained in the discussion of Grade 1. Detailed information from the item-level analysis is integrated into the narrative. The procedures used for the item-level analysis will be described in the next section.

Grade 1 subskill deficiencies

The information for the grade 1 subskill analysis is shown in Table 10 (below) and Figure 6 (opposite). In Table 10, the objective, the subskill number within objectives, and subskill name are listed on the left. The subskills are ordered from top to bottom according to their discrimination indices. The most discriminating subskills are listed at the top; the least discriminating subskills are listed at the bottom. In grade 1, the Operations Subskills of Subtraction and Addition are the most discriminating, followed by the Concepts 4 Subskill, Determining Place Value. Concepts 1, Counting, is the subskill with the lowest discrimination index. In Table 10, the subskill discriminations are followed to the right by the subskill difficulty for each of the four groups. These columns are labeled "Proportion Answering Correctly by Group." Listed next is the subskill difficulty based on all students who did not make the standard (+R-M and -R-M). These difficulties have a ranking from 1 to 11: 1 indicates the most difficult subskill, and 11 indicates the easiest subskill. Based on this information, the Subskills Subtraction (O2), Addition (O1), and Determining Place Value (C4) are the three most difficult subskills for the students who do not meet the mathematics standard. Counting (C1) is the easiest subskill. The last column in Table 10 lists the subskill difficulty based on all students, both those who did and did not reach the standard.

The information from Table 10 is shown graphically in Figure 6 (opposite), in which the subskills are ordered from hard to easy, going from left to right. Location of each subskill from left to right is determined by the subskill difficulty based on the students who did not make the standard. Thus, Counting

Objective	Subskill	Subskill Name	Discrimination Index (Discrimination Rank*)	Proportion Answering Correctly by Group R+M+	R-M+	R+M-	R-M-	Proportion of Those Failing Mathematics Answering Correctly (Difficulty Rank**)	Overall Proportion Answering Correctly (Overall Difficulty Rank**)
Operations	2	Subtraction (1-62 digits, no regrouping)	.355(1)	.883	.806	.589	.482	.519 (1)	.779 (2)
Operations	1	Addition (1-62-digit addends, no regrouping)	.343(2)	.983	.866	.660	.541	.582 (3)	.834 (3)
Concepts	4	Determining Place Value	.254(3)	.798	.693	.570	.512	.532 (2)	.720 (1)
Problem Solving	-	No Subskills Listed	.239(4)	.974	.950	.796	.699	.732 (5)	.908 (6)
Measurement	1	Identifying Units of Measurement	.237(5)	.924	.853	.741	.646	.679 (4)	.852 (4
Concepts	2	Identifying Equivalencies	.209(6)	.985	.930	.896	.701	.770 (8)	.923 (9)
Measurement	3	Using Measuring Devices	.182(8)	.969	.943	.858	.745	.785 (9)	.918 (8)
Geometry	1	Identification	.165(9)	.938	.901	.794	.755	.769 (7)	.891 (5)
Concepts	3	Establishing Relationships	.116(10)	.977	.969	.898	.842	.861 (10)	.947 (10)
Concepts	1	Counting	.076(11)	.980	.977	.939	.885	.904 (11)	.959 (11)

*1 = most discriminating **1 = most difficult

Table 10. Subskill analysis for grade 1, subskills ranked from most to least likely to differentiate those passing from those failing mathematics

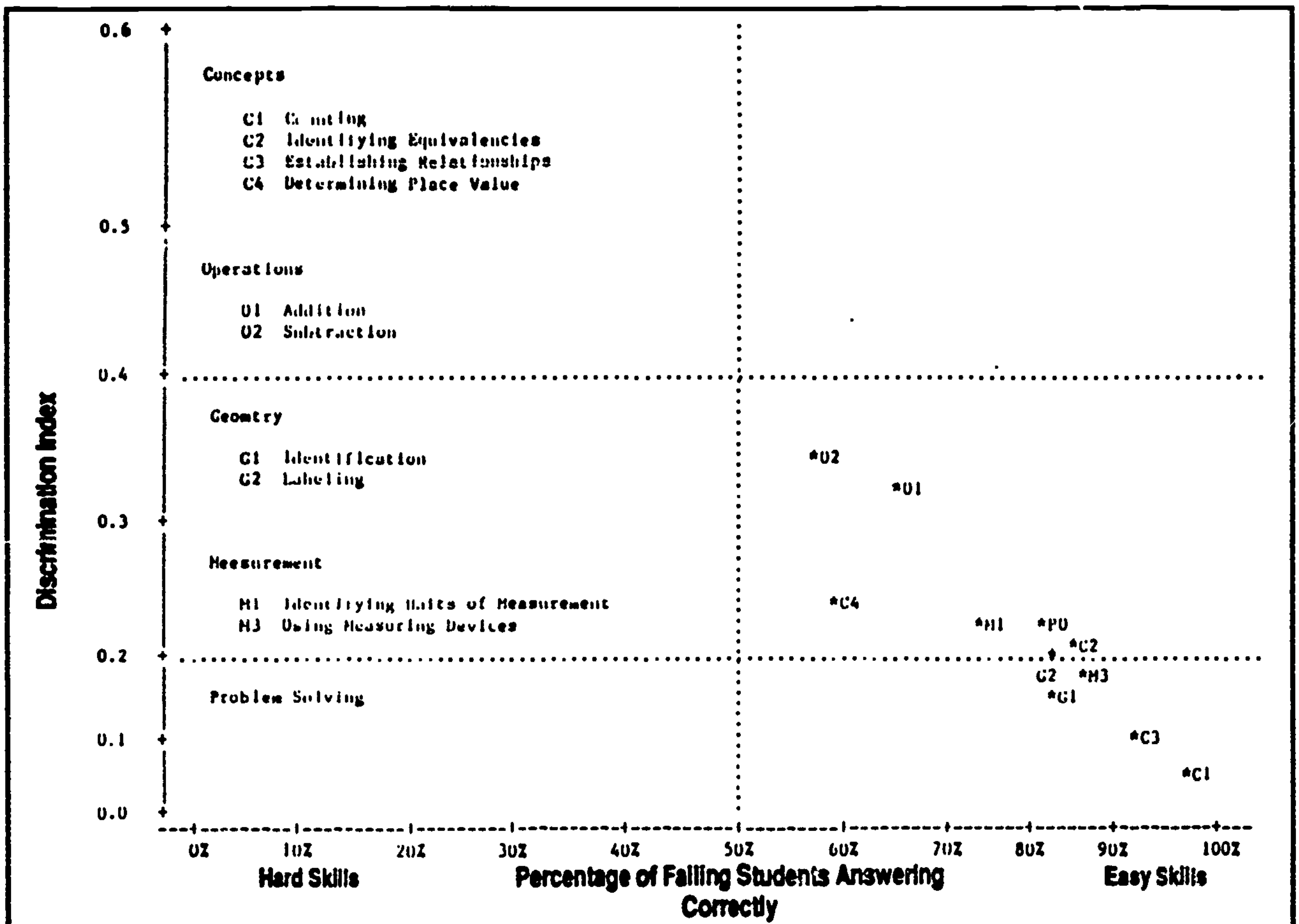

Figure 6. Plot of subskill discrimination by difficulty for grade 1

(C1) is just above the 90 percent point on the right, because the subskill has a difficulty of .904 for the students who do not make the standard (see Table 10). Subtraction (O2) is positioned over the 52 percent, because it has a subskill difficulty of .519 for the students who do not make the standard (see Table 10). The other subskills are likewise located along the line according to their difficulty, from left to right, for the students who did not reach the standard.

The position of the subskills from the bottom to the top of the figure also reflects information about the subskill. The vertical position of each subskill is based on its discrimination index. The higher on the page a subskill is located, the more highly it discriminates between the students who make the standard and those who do not. For example, Subtraction (O2) is the most discriminating subskill with a discrimination of .335 (see Table 10); therefore, it is the highest on the page of any of the subskills. This discrimination means that 35.5 percent more students who meet the standard answered the subtraction items correctly than did the students who did not meet the standard. Counting (C1), on the other hand, is the least discriminating subskill (.076); hence, it is located at the bottom of the page.

Figure 6 is very useful in identifying subskills for which students who do not make the grade level standard may need assistance. Subtraction (O2), Addition (O1), and Place Value (C4) stand out as the most difficult and most discriminating subskills. Each of these subskills was examined in close detail. The students who failed to make the mathematics standard did not have difficulty with basic subtraction facts involving 1-digit numbers less than 10. However, these students did have difficulty with basic facts involving numbers between 10 and 20 and with subtraction of a 1- or 2-digit number from a 2-digit number. No regrouping was required in grade 1 subtraction problems.

In the addition operation, the students who did not meet the standard had difficulty with addition of a 2-digit number with a 1- or 2-digit number. They did slightly better when adding three 1-digit numbers, although the horizontal format for such problems created some difficulties. Predictably, students generally did better on addition than subtraction.

Concepts Subskill 4, Determining Place Value, was a difficult and discriminating subskill. Students were reasonably capable when the item required going from the expanded form to a number, for example, "Which number below is the same as 2 tens and 7 ones?" Students had far more difficulty when given the number 27 and asked to identify the number in the tens place. This format was a difficult type of item for all students.

The subskills for Counting (C1) and Establishing Relationships (C3), which involved counting and matching sets, were the easiest and least discriminating subskills.

Objective	Subskill	Subskill Name	Discrimination Index (Discrimination Rank*)	Proportion Answering Correctly by Group +R+M	-R+M	+R-M	-R-M	Proportion of Those Failing Mathematics Answering Correctly (Difficulty Rank**)	Overall Proportion Answering Correctly (Overall Difficulty Rank**)
Operations	4	Division (2 digit by 1 digit)	.403(1)	.781	.657	.386	.343	.360 (2)	.646 (2)
Concepts	1	Counting	.344(2)	.911	.825	.611	.517	.554 (8)	.802 (11)
Problem Solving	-	No Subskills Listed	.342(3)	.864	.778	.528	.497	.510 (5)	.751 (7)
Operations	2	Subtraction (3 digits, regroup, fractions)	.339(4)	.844	.757	.537	.464	.493 (4)	.729 (4)
Measurement	4	Operations and Conversions	.338(5)	.783	.689	.464	.412	.432 (3)	.671 (3)
Operations	3	Multiplication (1 x 3 digits, regrouping)	.318(6)	.888	.822	.615	.525	.561 (9)	.781 (10)
Concepts	4	Determining Place Value	.312(7)	.877	.788	.608	.515	.552 (7)	.768 (9)
Operations	1	Addition (3 digits, regrouping, fractions)	.296(8)	.825	.766	.560	.491	.520 (6)	.733 (6)
Concepts	3	Establishing Relationships	.275(9)	.471	.337	.152	.195	.178 (10)	.385 (1)
Measurement	2	Estimation	.231(10)	.806	.764	.596	.550	.569 (10)	.732 (5)
Concepts	2	Identifying Equivalencies	.224(11)	.900	.816	.732	.619	.664 (12)	.819 (12)
Measurement	3	Using Measuring Devices	.204(12)	.946	.912	.773	.712	.737 (13)	.886 (14)
Measurement	1	Identifying Units of Measurement	.179(13)	.980	.924	.886	.730	.792 (16)	.917 (16)
Geometry	1	Identification	.165(14)	.820	.788	.661	.644	.651 (11)	.767 (8)
Concepts	5	Interpreting Tables and Graphs	.149(15)	.929	.909	.823	.746	.777 (14)	.882 (15)
Geometry	2	Labeling	.142(16)	.929	.908	.813	.765	.784 (15)	.884 (13)

*1 = most discriminating **1 = most difficult

Table 11. Subskill analysis for grade 3, subskills ranked from most to least likely to differentiate those passing from those failing mathematics

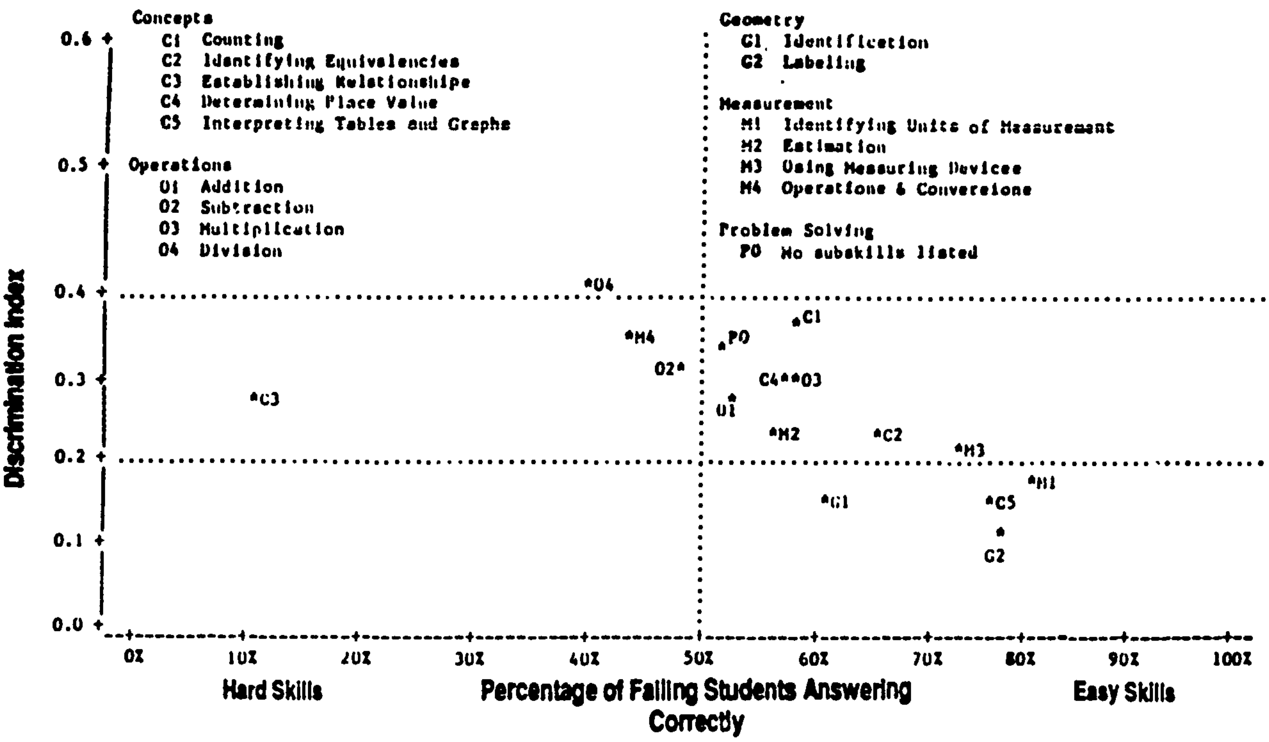

Figure 7. Plot of subskill discrimination by difficulty for grade 3

Grade 3 subskill deficiencies.

Information pertaining to the analysis of grade 3 subskill deficiencies is given in Table 11 (opposite) and Figure 7 (above). In Figure 7, two subskills stand out dramatically. These subskills are Establishing Relationships (C3) and Division (O4). Subskills M4, Operations and Conversions, and (O2), Subtraction, also fall below the 50 percent difficulty point for the students who did not meet the grade-level standard. Before discussing these particular subskills, it is useful to note that, in general, the grade 3 subskills are more difficult and discriminating than the grade 1 subskills.

At grade 3, Establishing Relationships (C3) requires students to identify a true statement of the form "4 + 2 5." The problems involve evaluating an inequality after performing an arithmetic operation. The students clearly do not know how to do these items.

Division problems (O4) appear on the BSAP tests for the first time in Grade 3. Students can do problems involving basic division facts, but have difficulty in dividing a 2-digit number by a 1-digit number.

Measurement Subskill 4 involves operations and conversions of measurement units within the same system and type of measure. This subskill includes units of money, time, and length. These items are relatively difficult for all students. The conversion of meters to centimeters is particularly difficult and discriminating.

Subtraction (O2) continues to be a problem at grade 3, as it had been in grade 1 (and grade 2). The students, even those who do not make the mathematics standard, can do subtraction of 1-, 2-, and 3-digit numbers if regroup-

Objective	Subskill	Subskill Name	Discrimination Index (Discrimination Rank*)	Proportion Answering Correctly by Group +R+M	-R+M	+R-M	-R-M	Proportion of Those Failing Mathematics Answering Correctly (Difficulty Rank**)	Overall Proportion Answering Correctly (Overall Difficulty Rank**)
Measurement	4	Operations and Conversions	.479(1)	.695	.570	.242	.179	.198 (4)	.418 (5)
Concepts	3	Establishing Relationships	.459(2)	.621	.523	.253	.157	.185 (2)	.385 (4)
Concepts	5	Interpreting Tables and Graphs	.412(3)	.762	.575	.408	.290	.324 (7)	.508 (10.5)
Operations	4	Division (mixed numbers, fractions, percents)	.408(4)	.692	.565	.306	.250	.266 (5)	.451 (6)
Measurement	5	Scale Drawings	.408(5)	.817	.720	.516	.345	.396 (14)	.579 (14)
Measurement	3	Using Measuring Devices	.378(62)	.726	.666	.380	.322	.339 (10)	.508 (10.5)
Concepts	4	Determining Place Value	.377(7)	.772	.712	.439	.362	.386 (13)	.562 (13)
Operations	1	Addition (mixed numbers, decimals, integers, regrouping)	.373(8)	.856	.762	.535	.441	.469 (17)	.639 (17)
Operations	3	Multiplication (mixed numbers, percents)	.363(9)	.828	.752	.512	.430	.455 (16)	618 (16)
Operations	2	Subtraction (mixed numbers, decimals, integers, regrouping)	.360(10)	.813	.728	.493	.418	.441 (15)	.598 (15)
Concepts	2	Identifying Equivalencies	.356(11)	.558	.405	.202	.171	.181 (1)	.341 (2)
Problem Solving	-	No Subskills Listed	.348(12)	.705	.570	.399	.311	.338 (8)	.492 (9)
Geometry		Identification 1	.344(13)	.720	.599	.438	.324	.359 (12)	.512 (12)
Measurement	1	Identifying Units of Measurement	.328(14)	.683	.563	.407	.313	.339 (9)	.482 (8)
Geometry	2	Comparison	.278(15)	.632	.568	.382	.329	.345 (11)	.469 (7)
Geometry	3	Application	.268(16)	.480	.378	.225	.188	.197 (3)	.318 (1)
Measurement	2	Estimation	.133(17)	.459	.431	.301	.329	.32[illegible] (6)	.383 (3)

*1 = most discriminating **1 = most difficult

Table 12. Subskill analysis for grade 8, subskills ranked from most to least likely to differentiate those passing from those failing mathematics

ing is not involved. However, subtraction problems involving regrouping are very difficult for the low-scoring students.

Grade 8 subskill deficiencies

The five BSAP mathematics objectives contain 17 subskills assessed at grade 8. Information describing student performance on these subskills is shown in Table 12 (above) and Figure 8 (opposite).

As the discussion of students' overall performance would suggest, eighth grade students have found the BSAP subskills rather difficult. At grade 8, the easiest subskill for all students was Addition (O1) with an overall difficulty of .639. In contrast, at grade 1, the hardest subskill was Determining Place Value (C4), with an overall difficulty of .72. First graders, therefore, were doing better on their most difficult subskill than eighth graders were doing on their easiest subskill. Clearly, the subskills are getting more difficult for students as they progress through the grades.

The eighth grade students who do not make the mathematics standard are clearly struggling. The easiest subskill for these students is Addition (O1), which has a difficulty of .469. Addition in the eighth grade involves mixed numbers, decimals, integers, and regrouping. On the other end of the scale there are five subskills that are so difficult for the students who do not reach the mathematics standard that they are operating at less than a chance level. These are, in order of discrimination, Operations and Conversions (M4), Establishing Relationships (C3), Division (O4), Identifying Equivalencies (C2), and Geometry Applications (G3).

Measurement Subskill 4, Operations and Conversions, requires multiplication or subtraction with regrouping to convert one unit of measure to another (e.g., calculating the number of inches in 3 feet). M4 problems involving time as the unit of measurement were especially difficult.

Establishing Relationships (C3) requires students to compare fractions with different denominators and to determine which fraction is smallest or largest. The students' incorrect responses suggest that they were comparing numerator to numerator and denominator to denominator and were not using a common denominator to make the comparison.

Division (O4) problems include decimals, mixed numbers, proper fractions, and percentages. Students found problems that asked them to determine what percentage one number was of another number especially difficult. This

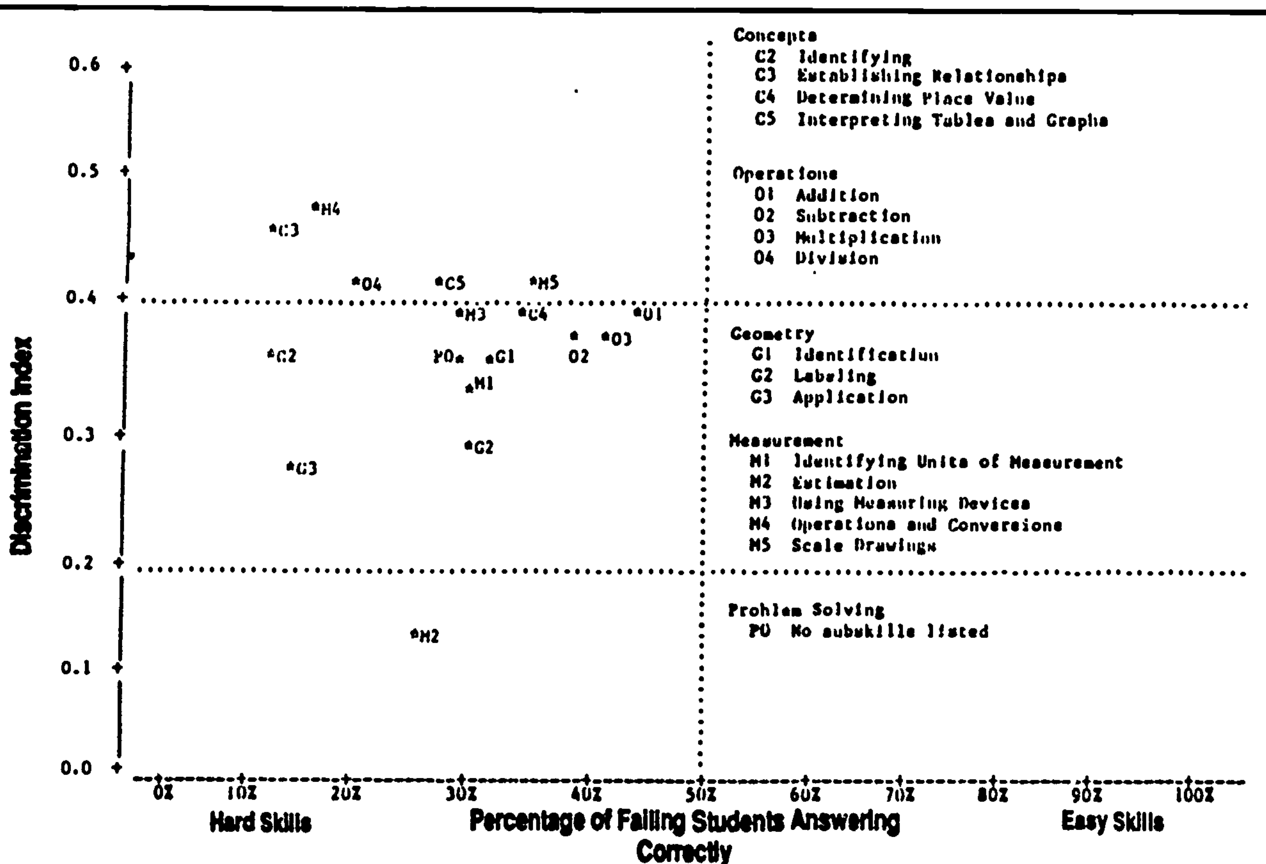

Figure 8. Plot of subskill discrimination by difficulty for grade 8.

type of item is a multistep problem requiring division and conversion to percentage.

Identifying Equivalencies (C2) involves problems with decimals, ratios, proportions, and percentages. Students had particular difficulty converting mixed numbers to percentages. Few students did well on these problems; students who did not make the mathematics standard were performing well below chance on these items. The +R-M and -R-M students had serious difficulty with all the items assessing the C2 subskill.

Geometry Applications (G3) requires students to calculate perimeter, area, and volume. Areas of triangles, parallelograms, and circles were particularly difficult for all students, with item difficulties below the chance level. Geometry Applications was the most difficult subskill for all eighth graders as a group, with a subskill difficulty of .318.

Three other subskills are notable because of their discrimination values. Interpreting Tables and Graphs (C5) and Scale Drawing (M5) are highly discriminating, whereas Estimating (M5) is conspicuous in its low discrimination power. Tables and Graphs (C5) problems requiring students to determine the percentage of a total figure represented by a shaded area are clearly more difficult than other types of problems measuring the same subskill. Estimation from scale drawings is a relatively easy subskill, which sharply differentiates those who meet the mathematics standard from those who do not. Problems involving meters are more difficult than problems involving miles. Measurement Subskill 2 involves estimation of area and volume. For all students combined, it is the third most difficult subskill; for the students who do not make the mathematics standard, it is the sixth most difficult subskill. The -R-M students actually do better on this subskill than the +R-M students. One possible explanation for the low discrimination power of this subskill is that the lower-achieving students may be using estimation procedures to solve the problems, whereas the more able students may actually be trying to calculate the exact answers. The calculating procedure may be a more difficult way to solve the problems than the estimation procedure.

Reanalyses at the item level

Reanalysis at the item level provides highly specific information about the characteristics of items that students find particularly difficult. The item-level analysis is generally performed separately for each subskill. Information from the item-level analyses has already been presented in the examples of subskill-level analyses for grades 1, 3, and 8. Procedures for the item-level analyses are very straightforward and basically involve two components. The first is the presentation of basic descriptive information for each item within a subskill. The second component involves studying the actual test items in conjunction with the basic descriptive information.

An example of the basic-item information for three hypothetical items is contained in Table 13 (opposite). The first level of information contains the name of the item, the overall item difficulty, the item difficulty for students not reaching grade-level standard, the item discrimination, and the item difficulty for the four categories of students. The second level of information is the proportion of students in each category who select each option. In actual prac-

Item	Overall Item Difficulties	Difficulty for Students Not Reaching Standards	Item Discrimination	Item Difficulty for Students in the 4 Categories +R+M	-R+M	+R-M	-R-M
C 3.17	.89	.51	.46	.97	.94	.53	.49
C 3.21	.71	.38	.43	.83	.67	.40	.37
C 3.42	.55	.36	.25	.63	.55	.37	.38

Proportion Selecting Each Option by Group

	+R+M			-R+M			+R-M			-R-M		
Item	A	B	C	A	B	C	A	B	C	A	B	C
Item 1 (C)	.02	.03	.95	.03	.03	.94	.22	.25	.53	.24	.27	.49
Item 2 (A)	.83	.10	.07	.66	.20	.14	.40	.36	.24	.36	.38	.26
Item 3 (C)	.23	.14	.63	.26	.19	.55	.38	.25	.37	.55	.10	.35

Table 13. Hypothetical item analysis for three option multiple choice items

tice, the table might also include a brief one-line description of the skill the item is measuring.

The data indicate that the three items vary in difficulty. Item 1 is easiest, Item 2 is of modest difficulty, and Item 3 is the most difficult. If item descriptors were attached to the items, the nature of the hard and easy items would be apparent.

The difficulties of items 2 and 3 for the students who do not reach grade-level standard are .38 and .36 respectively. Since there are three options per item a difficulty of .33 is expected by chance if students simply fill in their answer sheets randomly. Thus, students who do not reach grade-level standard have achievement on Items 2 and 3 that is barely better than chance performance. Performance that is close to the chance level is important to note because it may indicate an absence of instruction.

Items 1 and 2 discriminate sharply between those who do and do not reach grade-level standard. Item 3 also discriminates in the appropriate direction, but to a lesser degree.

The proportion of students in the four categories who select each option is revealing. For Item 1, most students in all four groups select the correct answer. Options A and B, wrong answers, are chosen equally often. For Item 2, Option A is the correct answer. Option B, however, is a popular incorrect answer. In the -R-M group, more students pick Option B than pick the correct answer. For item 3, the correct answer is Option C. Option A, however, is chosen frequently by all types of students. The +R-M students select Option A more often than they pick the correct answer. This pattern is even stronger for the -R-M students. Fifty-five percent of the -R-M students select Option A, which is incorrect, while only 35 percent select the correct answer (Option C).

The tendency of students who do not reach grade-level standard to pick a particular wrong answer is highly informative. In such situations, the test item and options should be studied very carefully. The situation illustrated by Item 3 suggests that students are making a type of error that is reflected in Option A. An analysis of Option A can provide a very specific description of a particular student deficiency. An example of this type of item is illustrated by the situation found in grade 8 with the Concepts Objective, Subskill 3, Establishing Relationships. In this subskill, students must identify the largest fraction from a particular set of fractions. For example, imagine that the options were A) 50/100, B) 20/50 and C) 6/10. The choice of Option A by the students who do not reach grade-level standard suggests that students are merely comparing the numerators to the numerators and the denominators to the denominators. Since 50 is greater than 20 or 6 and since 100 is greater than 50 or 10, students conclude that 50/100 is the largest fraction. This type of interpretation is somewhat speculative, rather than definitive. Nonetheless, it is highly likely that the students who do not reach grade-level standard and pick Option A are not comparing the fractions using a common denominator.

The preceding discussion of item-level analysis was designed to illustrate the two major components involved in examining specific test items, the first of which is the presentation of basic descriptive information. The major task for this activity is formatting computer output in convenient, readable, and well-organized tables and charts (e.g., Table 13). The second component of the item-level analysis involves the careful study of each item in conjunction with the basic descriptive information. This type of analysis requires expertise in mathematics and, especially, mathematics education. Teachers and other mathematics educators are well qualified to perform this content analysis.

As a final note, it is important to recognize that item-level analysis as it is described in the context of reanalyzing statewide test data is not the same as item analysis in the context of test development and measurement in general. Item analysis in the more general measurement sense is a set of procedures used to examine test items to determine whether they should be discarded, revised, or used on an operational test form. Item-level analysis in the context of reanalyzing statewide test data is a set of procedures used to extract as much information as possible about students based on their item responses.

Summary of the procedures for reanalyses

The reanalyses described in this section were designed to examine student achievement at the level of total test scores, objectives, subskills, and items. The purposes of the reanalyses, the intended audiences, and the structure of the tests involved will greatly influence which analyses would be useful and informative. The reanalyses performed in the South Carolina study will not necessarily be useful, informative, or even appropriate in all situations, and they should not be considered as a definitive list of reanalysis procedures. The South Carolina reanalysis procedures are described to illustrate some of the major types of reanalysis that can be employed and to illustrate the issues involved in choosing and interpreting the various reanalysis procedures.

The reanalysis procedures that have been described involve calculating means for item and subskill difficulties, taking the differences between means to get item and subskill discriminations, sorting the data into various groups, and preparing a variety of tables and charts. These are obviously not technically complicated procedures, and they can be easily carried out on a variety of microcomputers using a wide range of currently available software.

PROCEDURES FOR INTERPRETING REANALYSIS RESULTS IN TERMS OF FACTORS THAT AFFECT ACHIEVEMENT

Interpretation procedures

The major purposes of the reanalyses were to describe general trends in students' mathematics achievement, to identify specific areas of mathematics in which students had deficiencies, and to identify changes in curriculum, instruction, and teacher training that would improve student achievement in the deficiency areas. The reanalysis procedures illustrated in the preceding section can be used to describe general trends in student achievement and to identify areas of student deficiencies. The reanalysis procedure, of course, cannot provide information about changes in curriculum, instruction, and teacher training that would improve achievement in the deficiency areas. The relevant information from any reanalysis must be translated or interpreted in order to determine how curriculum, instruction, and teacher training can be best modified to improve student learning.

Researchers who conduct studies are traditionally the people who interpret the research results and make recommendations implied by the interpretation. This approach was not used in the South Carolina project. Rather, a Project Study Group composed of experienced mathematics educators was assembled to interpret the results of the reanalysis and to make recommendations. The use of such a group was extremely valuable since a major purpose of the project was to formulate recommendations for modification in the areas of curriculum, instruction, and teacher training. Members of the project study group were very well qualified to offer such recommendations. In addition, these mathematics educators were able to offer interpretations and recommendations that were helpful and meaningful to other mathematics educators. Research specialists, on the other hand, are not necessarily expert in mathematics curriculum, instruction, and teacher training. In addition, researchers are generally trained to present research results in a way that is meaningful to other researchers but not necessarily to practical-minded educators. The role of teachers and other educators as the interpreters of data analysis differs from the traditional research perspective, which generally views educators as consumers of completed research conducted and interpreted by researchers.

The project study group

The project study group was composed of 37 educators carefully chosen to represent all levels of mathematics education and all areas of the state. The group included elementary, middle, and high school teachers; Chapter 1 coordinators; district mathematics coordinators; assistant superintendents; superintendents; faculty from teacher training institutions; staff from the State Department of Education; and members of the research team. The group was selected to be representative of other demographic characteristics, such as gender and race.

All members of the project study group received a copy of the project report through the section describing the results of the reanalysis. They were asked to read and study the report for several weeks and to note any comments, suggestions, or recommendations that occurred to them as they examined the results of the reanalysis, incorporating insights based on their own experiences into their comments about the reanalysis results.

The study group gathered for a one-day meeting. After a general presentation in which the reanalysis results were reviewed and questions were answered, the participants were divided into 3 groups of approximately 12 people. The work of these groups was coordinated by a member of the research team, and each group was asked to respond to five tasks.[1] These tasks were to:

1) Identify useful BSAP information.

2) Identify effective ways to report BSAP information.

3) Identify strategies, procedures, and materials to support the mathematics curriculum.

4) Identify strategies, procedures, and materials to support mathematics instruction.

5) Identify strategies, procedures, and materials to support teachers.

Tasks 1 and 2 deal with the information component of the Conceptual Model (Figure 1, page 3). Tasks 3, 4, and 5 deal with factors that affect achievement.

The activities of each group were organized around a worksheet for each of the five tasks. The worksheets are shown in the Appendix. The worksheet for Task 3 dealing with curriculum is shown on the next page.

A five-step procedure organized around the worksheets was used by the groups in responding to each of the five tasks. The procedure will be illustrated using the third task, dealing with curriculum, as an example. In Step 1, each member of the committee listed three to five ways to modify or support mathematics curriculum to facilitate student achievement. In Step 2, after all members had written their Step 1 suggestions on the worksheets, the group coordinator asked each person to read his or her suggestions, and the suggestions were written on a chart so that everyone could see them. Repetitious suggestions were not listed. The process continued until everyone had participated and all unique suggestions had been listed. All suggestions were numbered on the chart. In Step 3, people selected or voted for the three suggestions on the list that they thought would be most useful for supporting mathematics curriculum. The number corresponding to each choice was written at the bottom of the worksheets. In Step 4, the worksheets were collected, the votes tallied, and the suggestions for curriculum development were listed in the order of their popularity. In Step 5, the recommendations obtained in Steps 1 through 4 from all three groups were combined and organized under a set of major headings.

Using this procedure, the study group had seven major recommendations dealing with curriculum. These were:

(1) *These tasks and the worksheets used to guide the efforts of the Project Group in responding to the tasks were developed by Dr. Jeanne Miyasaka, South Carolina Department of Education.*

Name ____________

BSAP Mathematics Project Meeting
February 15, 1985

Task 3: Identifying Strategies, Procedures and Materials to Support the Mathematics Curriculum

The study considered certain aspects of the curriculum and curricular materials that may be related to students' BSAP performance.

A. Based on the results of the study and your own experience list three (3) to five (5) ways to modify or support the mathematics curriculum to facilitate students' achievement.

1. ____________
2. ____________
3. ____________
4. ____________
5. ____________

B. Select three (3) ideas from the study group list that you think would be the most useful for supporting the mathematics curriculum. Write the corresponding numbers of the ideas in the spaces below.

________ ________ ________

Figure 9. Worksheet for Task 3

(1) Identify supplementary materials related to BSAP curriculum and instruction.
(2) Provide a list of cross-references that shows the connections among BSAP subskills, BSAP item types, *Teaching and Testing Our Basic Skills*, state-adopted textbooks, and supplementary curriculum materials.
(3) Organize or facilitate the organization of a centralized clearinghouse for BSAP-related materials dealing with curriculum, instruction, and diagnostic tests.
(4) Develop and provide nonsecured diagnostic tests assessing BSAP subskills. Organize the diagnostic tests by objectives over the grades.
(5) Review and consider restructuring BSAP objectives with respect to a more broadly defined mathematics curriculum and other instructionally related issues.
(6) Review and consider revising *Teaching and Testing Our Basic Skills*.
(7) Review and consider revising BSAP item specifications.

Teaching and Testing Our Basic Skills, referenced in these recommendations, is a curriculum/test guide prepared by the South Carolina Department of Education to explain the basic skills program to teachers. The process illustrated for Task 3 dealing with curriculum was also used to formulate recommendations for instruction, teacher training, and reporting BSAP information.

The recommendations of the project study group obtained through the five-step procedure were incorporated in the project's final report. The recommendations were not used as the basis for formulating a final recommendation by the researcher, but constituted the final recommendations themselves. Thus, in effect, the project study group composed of mathematics educators wrote the final recommendations based on the reanalysis. The researchers served as facilitators in the process of formulating and organizing the group's recommendations, but the researchers did not edit or modify the substance of the recommendations made by the mathematics educators.

The involvement of teachers and other educators as the interpreters of the reanalysis is seen as a special feature of the project. It had several major benefits. First, the interpretation of the data analysis and the recommendations based on the interpretations turned out to be very practical and useful. Second, teachers and other educators who have read or heard reports from the study seem to have viewed the study as being especially useful since the interpretations and recommendations seem relevant to practical problems they have encountered. Third, the educators were able to incoporate their own experiences and insights into their recommendations and were not limited to test data alone. Lastly, the common gap between educational research and educational practice was narrowed by this study since school-based educators were involved directly in interpreting university-based research that was designed to have practical implications.

PROJECT OUTCOMES

Introduction

The South Carolina project had a wide range of outcomes including a variety of dissemination activities and the development of curriculum and instructional materials for statewide use. The dissemination activities were extremely important in helping to gather support for the recommendations of the project study group. The curriculum and instructional materials developed in response to the study group's recommendations are currently being used in South Carolina, and evaluation studies designed to assess the impact of these materials are currently underway.

Dissemination activities

The results of the reanalyses and the recommendations of the project study group were disseminated across the state through written materials, oral presentations, and television broadcasts. The written materials included the project final report (108 pages), the Executive Summary (22 pages), and a set of notes that abstracts the Executive Summary (4 pages). All district superintendents and curriculum or mathematics coordinators received a set of all three documents. In addition, every principal in the state (approximately 1,200) received the Executive Summary and the abstract. Copies of the Executive Summary were sent to a variety of professional organizations in the state, including the Association of Mathematics Supervisors and the Association of Teachers of Mathematics.

More than 15 oral presentations concerning the project were made around the state to a variety of interested groups. These presentations were organized by the State Department of Education, different professional organizations, various policy groups, individual school districts, and school districts working together. The Executive Summary was the basis for these presentations. Emphasis was given to evidence of major progress at all grades and the identification of specific areas of mathematics on which students were still having difficulty. The recommendations of the project study group were presented and described in detail. The presentation also demonstrated how teachers could relate areas of deficiency (e.g., Concepts, Subskill 3) to the BSAP curriculum guide *Teaching and Testing Our Basic Skills*. The purpose of the demonstration was to show teachers how they could obtain clarification of the deficiency areas.

A major dissemination device was a television production, which described the study, the major results, and the recommendations of the project study group. An announcement of the broadcast was sent to every principal in the state along with the Executive Summary and the abstract of the summary. This abstract contained the major tables and charts used in the broadcast along with a listing of the project study group's recommendations. These "TV Notes" were designed to minimize the need for teachers to make copious notes during the broadcast. The original telecast occurred in August, the week before school began, when schools were providing in-service workshops for teachers. The telecast was broadcast to be used as the basis for an in-service program or to be videotaped for use in a future in-service program. The production was also broadcast on six later occasions.

The dissemination activities were a crucial component of the reanalysis project because the purpose of the project was to bring about changes in curriculum, instruction, and teacher training. The dissemination procedures were designed to achieve this purpose. The traditional academic approach of publishing the results in various journals was not viewed as especially useful since it would introduce considerable delay in disseminating results without necessarily communicating them to the intended audience. The extensive use of oral presentations in the schools, at various regional meetings, and at meetings of appropriate policy and professional groups was very effective in delivering the information to the people who could initiate the needed changes. The television broadcasts to the schools also appear to have been very effective in delivering the reanalysis information and recommendations to educators who could act upon them.

Experience with the South Carolina project indicates that the dissemination procedures should be given as much attention and support as the reanalysis activities themselves. The value of the project is not in merely identifying student deficiencies, but in encouraging and enabling educators to respond in ways that can correct the problems. To do this, educators in the schools need to have the descriptions of the student deficiencies. The recommendations of the project study group were also extremely helpful to educators around the state.

Examples of project outcomes

The project's written materials, oral presentations, and television broadcasts all contained a strong message urging educators in schools, school districts, the State Department of Education, and all teacher training institutions to study the report and to use the recommendations of the project study group to:

1) Specify priorities from among the recommendations.
2) Specify procedures for responding to the recommendations given priority.
3) Specify a schedule for implementing procedures responsive to the recommendations.
4) Consider procedures for documenting, evaluating, and sharing information about these implementation activities.

It is, nonetheless, very difficult to assess all the outcomes of the project, since no systematic survey of school and school districts' activities has been conducted. At the state level, however, there have been four specific outcomes that represent the responses of the State Department of Education to the recommendations of the project study group. First, the annual report of BSAP test results has been changed to include a "right-response summary." In this summary, each item on the test is identified in terms of the specific subskill it is designed to measure. Then, for each item, the proportion of students in the school, the school districts, and the state who answered the item correctly is reported. Differences between the school and school district and between the district and state proportions are also reported. The previous reporting procedures provided information about students at the objective level. The objective could contain several subskills, and students' performance was reported simply as adequate or deficient at the objective level. The change in reporting has been very useful to teachers and other educators in identifying specific weaknesses of students in their school or school district. This change was specifically recommended by the project study group and resembles the type of analysis conducted as part of the project.

The second major response of the State Department of Education has been the publication of a set of curriculum materials entitled *Planning for Improvement in Basic Skills Mathematics*. Three sets of materials have been produced: Grades 2-4, Grades 5-7, and Grades 7-9. The materials focus on the five subskills that the reanalyses indicated were the most problematic for students who did not reach grade-level standard at grades 3, 6, and 8. Each subskill considered in these materials is examined in considerable detail; the nature of the deficiency is described, along with a summary of the reanalysis results. The description of instructional implications includes a listing of instructional strategies, teaching activities, teaching materials (which can be copied from the publication), and teaching games. Annotated references to other materials, videotapes, computer software, and published written materials that could be used to provide remediation of the deficiency are also provided.

A third major response of the State Department of Education to the recommendations of the project study group has been the development of an instructional television series that demonstrates effective teaching strategies for remediation of deficiencies in the subskills identified in *Planning for Improvement in Basic Skills Mathematics*. There are 24 lessons for grade 3, 15 lessons for grade 6, and 14 lessons for grade 8. A Teacher Guide has been prepared to accompany this series. The units in the series provide an overview of the subskills, general and specific guidelines for teaching the subskills, and models of some specific teaching strategies.

The fourth State Department activity that evolved from the project is the reanalysis of the student BSAP reading achievement from 1981 to 1985. The procedure described for the mathematics project was employed, with minor modifications, in the reading study.

SUMMARY AND DISCUSSION

The purpose of this publication is to describe the South Carolina Project as a model for other educators interested in taking full advantage of test data that have already been collected. The project was organized around the conceptual model shown in Figure 1 (page 3). The starting point in the Conceptual Model is student achievement. Information about student achievement can be obtained by using the reanalysis procedures employed in the South Carolina project. These procedures are designed to supply information about different types of students (e.g., +R+M, +R-M, -R+M, -R-M) in terms of different levels of performance (e.g., total test scores and performance at the objective, subskill, and item level).

The information about student achievement obtained from the reanalysis then needs to be interpreted and disseminated. The interpretation and dissemination steps are shown in the Conceptual Model and have been described in some detail. The interpretation of the reanalysis information focuses on factors that affect achievement, such as curriculum, instruction, and teachers' preservice and in-service training. Dissemination is directed toward mathematics educators who can help bring about the changes in these areas that will improve student achievement. A number of project outcomes at the state level dealing with reporting achievement information, curriculum, and instruction have been described. These and similar activities at the school and school district level are designed to have a positive impact on student achievement. Student achievement is thus both the starting point and the ending point in the diagnosis and prescription cycle.

As mentioned in the Introduction, the South Carolina Project should not be copied directly in the reanalysis of school, school district, or statewide test data. Rather, educators in these settings should use the South Carolina experiences as a starting point and modify the process to meet the needs of their own situation. In this process, the active involvement of local educators in all phases of the project is essential. It is not merely a "good idea" to involve mathematics educators, it is essential to the success of the project. The reanalysis of test data is not the purpose or end product of the project. The end product is changes in curriculum, instruction, and teacher training that will have a positive impact on student achievement, especially in those areas where reanalysis shows students to be deficient. Experts in measurement and statistics can help reanalyze test data. The experience and expertise of mathematics educators, however, is needed to translate or interpret the reanalysis results into curriculum, instruction, and teacher training activities. In the final analysis, the South Carolina Project is primarily concerned with curriculum, instruction, and teacher training, not with the reanalysis of test data.

50

Worksheets used by the project study group

Name________________

BSAP Mathematics Project Meeting
February 15, 1985

Task 1: Identifying Useful BSAP Information

The project has examined BSAP test data in an effort to provide information that might be helpful in developing curriculum materials and in planning instructional strategies. Three types of information are described in the report. These are: (1) information about BSAP scores for students in each of four groups, depending on their Mathematics and Reading scores, (2) information about how close or how far students were from the mathematics standards, and (3) information about students' performance on BSAP subskills.

Please comment briefly in the space provided on the usefulness of these three types of information.

(1) Information about the four groups.

____

____

(2) Information about how close or how far students were from the standard.

____

____

Information about subskill performance.

____

____

A. Other type of information can be obtained by examining BSAP test data in various ways. List three (3) to five (5) other types of BSAP information you think would be helpful.

1. ______________________________

2. ______________________________

3. ______________________________

4. ______________________________

5. ______________________________

B. Select three (3) ideas from the study group list that you think would be most useful. Write the corresponding numbers of the ideas in the space below.

__________ __________ __________

51

Name________________

BSAP Mathematics Project Meeting
February 15, 1985

Task 2: Identifying Ways to Report BSAP Information

As a result of the study, several types of information about students' BSAP mathematics achievement were provided. In addition, in Task 1, other types of useful BSAP information were identified.

A. List three (3) to five (5) ways to enhance the reporting of the BSAP information.

1. ____
____

2. ____
____

3. ____
____

4. ____
____

5. ____
____

B. Select three (3) ideas from the study group list that you think would be most useful. Write the corresponding numbers of the ideas in the space below.

____________ ____________ ____________

Name ______________

BSAP Mathematics Project Meeting
February 15, 1985

Task 4: Identifying Strategies, Procedures, and Materials to Support Mathematics Instruction

A variety of instructional strategies can be used for teaching and reteaching mathematics skills. Many of these strategies are supported by diagnostic testing.

A. Based on the project results and your experience, list three (3) to five (5) strategies, procedures, and/or materials, which facilitate instruction.

1. ______________________________

2. ______________________________

3. ______________________________

4. ______________________________

5. ______________________________

B. Select three (3) ideas from the study group list that you think would be most useful. Write the corresponding numbers of the ideas in the space below.

__________ __________ __________

Name________________

BSAP Mathematics Project Meeting
February 15, 1985

Task 5: Identifying Strategies, Procedures, and Materials to Support Teachers

Teacher effectiveness is based on many characteristics including their preservice and in-service training and the ongoing instructional support they receive.

A. List three (3) to five (5) ways to prepare or support teachers in order to enhance the effectiveness of their mathematics instruction.

1. ____
2. ____
3. ____
4. ____
5. ____

B. Select three (3) ideas from the study group list that you think would be most useful. Write the corresponding numbers of the ideas in the space below.

__________ __________ __________ __________