DOCUMENT RESUME

ED 404 180 SE 059 722

AUTHOR Burstein, Leigh; And Others
TITLE The Validity of Interpretations of the 1992 NAEP Achievement Levels in Mathematics.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 31 Aug 93
CONTRACT RS90159001
NOTE 214p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Mathematical Concepts; *Mathematics Achievement; Mathematics Skills; *Test Construction; *Test Results
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This report evaluates the degree to which the achievement level descriptions adopted by the National Assessment Governing Board (NAGB) for the 1992 National Assessment of Educational Progress (NAEP) assessment in mathematics accurately represent what students at a given achievement level can do. Three different analytical approaches were used in the investigation and resulted in the following conclusions: (1) judged in terms of actual student performance, many of the items selected as exemplars of the achievement levels are misleading; (2) the 1992 NAEP mathematics assessment did not measure some of the attributes included in the descriptions of the achievement levels and measured some other attributes only poorly; (3) frequently, many of the students at a given level did not successfully answer items linked to certain aspects of the descriptions at that level; (4) the definitions of the levels overlap considerably and frequently differ only in subtle nuances; and (5) the characteristics of items that differentiate among achievement levels suggest descriptions of performance that differ substantially from the current achievement level descriptions. It was concluded that the analyses did not support the validity of the published content descriptions as characterizations of what students within specified scoring ranges can do. (JRH)

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

National Center for Research on
Evaluation, Standards, and Student Testing

Technical Review Panel for Assessing the Validity of the National Assessment of Educational Progress

The Validity of Interpretations of the 1992 NAEP Achievement Levels in Mathematics

- UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- University of California, Santa Barbara
- University of Southern California
- The RAND Corporation

National Center for Research on
Evaluation, Standards, and Student Testing

Technical Review Panel for Assessing the Validity of
the National Assessment of Educational Progress

The Validity of Interpretations of the 1992 NAEP Achievement Levels in Mathematics

**Leigh Burstein, Daniel M. Koretz, Robert L. Linn,
Brenda Sugrue, John Novak, Elizabeth Lewis & Eva L. Baker**

University of California, Los Angeles, The RAND Corporation, and
The University of Colorado, Boulder

August 31, 1993

U.S. Department of Education
National Center for Education Statistics
Grant RS90159001

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part under the National Center for Education Statistics Contract No. RS90159001 as administered by the U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the National Center for Education Statistics or the U.S. Department of Education.

Table of Contents

Executive Summary

The National Center for Education Statistics (NCES), concerned that the reporting of the National Assessment of Educational Progress (NAEP) should be accurate and informative, asked the NAEP Technical Review Panel (TRP) to evaluate the degree to which the achievement level descriptions adopted by the National Assessment Governing Board (NAGB) for the 1992 assessment in mathematics accurately represent what students at a given achievement level can do. This report presents the results of that inquiry.

Descriptions of the skills and knowledge represented by the achievement levels are necessary to guide inferences about performance on the NAEP and to provide signals about needed improvements. However, to serve these functions, and to avoid generating unwarranted and incorrect interpretations of student performance, these descriptions must meet certain standards. In particular, the descriptions should provide valid indications of what students who perform at a given achievement level did on the NAEP assessment in mathematics.

It is important at the outset to state that the analyses and results that are presented here are focused on the issue of whether performance on NAEP validates the proposed content descriptions. In addressing this issue, we examine **the degree to which the descriptions of the levels provide valid indications of what students who perform at a given achievement level did on the NAEP assessment in mathematics**. There are other important questions that are not addressed in this report. The report is mute with regard to issues such as the appropriateness of the achievement level setting process and the appropriateness of the levels that were adopted. Nor does it evaluate the utility of the achievement levels for various audiences.

Three general approaches were used to investigate the degree to which the descriptions of the NAEP achievement levels provide a valid indication of the actual performance of students at each of the achievement levels:

1. The statistical properties of the items which had been selected as exemplars in the descriptions of each achievement level were reviewed.

2. The NAGB descriptions of the levels were used to form a list of statements about what students at a given level should be able to do. Judges (mathematics educators familiar with the curriculum at the target grade levels) then used those statements (without being told the level from which the statement was taken) to identify items that called for the knowledge, skill, or understanding contained in the descriptor-based statements, and the performance of students on the identified items was summarized for each level.

3. Items were classified by achievement level in terms of a number of statistical indices, such as the extent to which performance on each item differentiated between students at different levels, and the content of the items (as identified by curriculum experts) was compared to the descriptions of the corresponding achievement levels.

The following five major conclusions are based on the results of the three analytical approaches just described.

1. **Judged in terms of actual student performance, many of the items selected as exemplars of the achievement levels are misleading.** In some instances, less than half the students performing within the range of a given achievement level correctly answered an exemplar item for that level. In other cases, more than 75% of the students performing at a given level correctly answered an item intended to be an exemplar for the next higher achievement level. Presenting such items as exemplars of a given level provides a misleading impression of what students performing at a given level are actually able to do.

2. **The 1992 NAEP mathematics assessment did not measure some of the attributes included in the descriptions of the achievement levels and measured some other attributes only poorly.** That is, the 1992 item pool provided sparse coverage of some attributes and no coverage of others. This sparse coverage is especially problematic for the grade 4 basic and advanced levels and for the grade 12 advanced level. Thus, it is impossible to say with any confidence whether students scoring at the level in question can do what those aspects of the descriptions describe.

3. **Frequently, many—in some cases, a majority—of the students at a given level did not successfully answer items linked to certain aspects of the**

descriptions at that level. Among students whose performance reached a given level, performance on items linked to that level (by the second of the approaches noted above) varied and was in many cases lower than many people would consider reasonable. For example, in some instances, the median percentage of students answering correctly was less than 50% on items associated with that level. Low percent correct values were especially frequent for items in the Basic range. This variation in performance is greatest for items corresponding to Basic level descriptions.

4. **The definitions of the levels overlap considerably and frequently differ only in terms of subtle nuances**. Consequently, the association of items with a given level was often found to be ambiguous. Experienced mathematics educators were generally unable to make such distinctions reliably without specific and detailed training. Thus, it is unlikely that general populations of mathematics specialists, professional educators or the lay public could be any more successful at interpreting correctly the intended differences among levels.

5. **The characteristics of items that differentiate among achievement levels suggest descriptions of performance that differ substantially from the current achievement level descriptions.** Differentiating items were identified on the basis of statistical properties (i.e., high probability of correct response for students at that level and a relatively low probability of correct response for students scoring below that level), and judges ascertained the attributes of these items. Judging from this empirical evidence, the primary bases for differentiating the performance of students across levels appear to be the extensiveness and quality of curriculum exposure and potentially associated degrees of language facility.

In sum, then, our analyses do not support the validity of the published content descriptions as characterizations of what students within specified score ranges can do. Some of the attributes of the descriptions could not be mapped to the NAEP items; those that could be mapped to NAEP did not consistently show performance patterns that would support the validity of the descriptions; and the exemplars as a set do not accurately characterize the performance of groups in question.

In our judgment, descriptions of the achievement levels are not informative unless they accurately portray what students at the various levels can do. Characterizations of the levels should align with the actual performance of students on the NAEP, and empirical evidence of that alignment should meet reasonable standards. The likelihood that these goals can be met depends not only on the processes used to set the levels and establish descriptions, but also on the characteristics of the NAEP itself. For example, the item pool must be rich at each of the levels, and it must represent adequately the skills and knowledge that are the basis for setting the levels and that are used to describe them. Neither of these criteria was consistently satisfied in the establishment of the 1992 achievement levels in mathematics.

The task in mathematics (and perhaps in other areas) is all the more difficult because the field is still in the early stages of major curriculum reform where there is considerable variability in the penetration and extent of reform at the classroom level. Under such circumstances defining achievement levels based on what students can do now may differ markedly from what is deemed desirable that they be able to do if the reform takes hold. This creates the natural tension between building the assessment and associated achievement levels around the desired new curriculum frameworks to capture what we want students to be able to accomplish versus grounding them accurately in the current prevailing conditions based on assessment frameworks and associated item pools that no longer represent the full range of desired learning goals. **While the flaws in the content descriptions identified in our work can be attributed in part to insufficient attention to examining their validity,** it may well be that the shortcomings of the current achievement level effort are inextricably tied to the mismatch between the natural desire to move beyond the current horizon with an assessment design and associated data that are not appropriate to the task.

The Validity of Interpretations of the 1992 NAEP Achievement Levels in Mathematics

Leigh Burstein, Daniel M. Koretz, Robert L. Linn, Brenda Sugrue, John Novak, Elizabeth Lewis and Eva L. Baker

Background

Over the past several years, the reporting of student performance on the National Assessment of Educational Progress (NAEP) has been changing in response to evolving expectations for the assessment. Until recently, reporting of NAEP results was intended only to describe what students know and can do. Judgments about what students should be able to do were left to readers; no effort was made to incorporate such judgments into the actual reporting of NAEP results.

The 1988 legislation reauthorizing the NAEP, however, focused attention on standards for what students should know. That statute established the National Assessment Governing Board (NAGB) and gave it responsibility for setting "appropriate achievement goals." In an effort to meet that responsibility, NAGB has set performance standards, called achievement levels, for the 1990 and 1992 assessments. The achievement levels set three standards of performance on NAEP at each grade level: Basic, Proficient, and Advanced.

The achievement levels have been controversial since the first effort in 1990-91 (General Accounting Office, 1993; Linn, Koretz, Baker, & Burstein, 1991; NAGB, 1991; Stufflebeam, Jaeger, & Scriven, 1991). The points of controversy have been diverse, pertaining to both the process by which levels were established and the meaningfulness of the final standards. Because of the controversy regarding the achievement levels that were established in 1990-91 in mathematics, those initial levels were viewed as preliminary.[1] Rather than treating them as a baseline for future assessment, a new effort was undertaken by NAGB and their contractor, the American College Testing Program (ACT), to set achievement levels for the 1992 assessment. The pending release of the achievement levels for the 1992 mathematics assessment that were adopted by NAGB based on the standard setting work conducted by ACT has already provoked disagreements regarding the interpretation of the levels (GAO, 1993).

A primary controversy about the new achievement levels in mathematics is simply whether the descriptions of achievement levels prepared by ACT and NAGB provide a reasonable depiction of the performance of students who reach the achievement levels. The descriptions of the levels are phrased for the most part in terms of what students **should** know and be able to do, but ACT and NAGB initially maintained that the wording describes what students **can do** and that the interpretations of results should be rephrased accordingly. These demurrers notwithstanding, we believe that the levels are widely interpreted as statements about what students can do regardless of the use of "should" in the descriptions or in text providing interpretations of results. The very logic of the achievement levels foreordains their interpretation in this way: Judges think about what students should do;

[1] Despite their purported preliminary status, the reports of the 1992 mathematics results adjusted the 1990 levels to permit trend analysis!

those expectations are mapped to NAEP; and finally, NAEP reports how many students **actually do** meet these expectations.

The National Center for Education Statistics (NCES), concerned that the reporting of NAEP should be accurate and informative, asked the NAEP Technical Review Panel (TRP) to evaluate the degree to which the achievement level descriptions adopted by NAGB for the 1992 assessment in mathematics validly represent what students at a given achievement level can do. This report presents the results of that inquiry.

It is important to stress at the outset that the analyses and results that are presented here are focused on the issue of **whether performance on NAEP validates the proposed content descriptions**. In addressing this issue, we examine **the degree to which the descriptions of the levels provide valid indications of what students who perform at a given achievement level did on the NAEP assessment in mathematics**. There are other important questions that are not addressed in this report. The report is mute with regard to issues such as the appropriateness of the achievement level setting process and the appropriateness of the levels that were adopted. Nor does it evaluate the utility of the achievement levels for various audiences.[2] The levels are taken as a given in this report.

The 1992 Mathematics Achievement Levels

Achievement levels in mathematics were established in both 1990 and 1992. Achievement levels in reading have also been established by NAGB. The focus of this report, however, is limited to the 1992 mathematics achievement levels.

[2] The utility of the 1990 achievement levels for writers in the popular media was the subject of another TRP report (Koretz and Deibert, 1993).

NAGB provided simple, "policy-based" definitions of the three levels (see Figure 1) that served as the basis for panels of judges to develop grade and subject matter specific descriptions of the levels. In addition to the policy-based definitions, the panels used the NAEP mathematics frameworks and their experience with the NAEP assessments in arriving at content-based descriptors of each achievement level. The latter descriptions were then used by panels of judges consisting of teachers, non-teacher educators and non-educators who reviewed the NAEP items. The panelists were asked to provide their best judgment of the percentage students at the borderline of each of three achievement levels who would respond correctly to the items. The average judgments on a final set of ratings were mapped onto the NAEP scale (see

Figure 1

The NAEP Policy Level Definitions of the Achievement Levels*

"Basic. This level, below proficient, denotes partial mastery of knowledge and skills that are fundamental for proficient work at grade —4, 8, and 12. For 12th grade, this is higher than minimum competency skills (which normally are taught in elementary and junior high schools) and covers significant elements of standard high-school-level work."

"Proficient. This central level represents solid academic performance for each grade tested—4, 8, and 12. It reflects a consensus that students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. At grade 12, the proficient level encompasses a body of subject-matter knowledge and analytical skills, of cultural literacy and insight, that all high school graduates should have for democratic citizenship, responsible adulthood, and productive work."

"Advanced. This higher level signifies superior performance beyond proficient grade-level mastery at grades 4, 8, and 12. For 12th grade, the advanced level shows readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, it may be based in part on international comparisons of academic achievement and may also be related to Advanced Placement and other college placement exams."

* Phillips et al., 1993, Interpreting NAEP scales, p. 38.

American College Testing, 1993, for a more complete description of the rating process).

The mapping established 9 scale points corresponding to the minimal scores for the Basic, Proficient, and Advanced achievement levels at each of grades 4, 8, and 12. The final levels were set by NAGB to be one standard error below the scale points identified by the panelists. This adjustment in the levels ranged from approximately 2 to 6 points on the NAEP scale depending on the grade and achievement level (Mullis et al., 1993, p. 361).

Refinements in the descriptions of the final achievement levels were also made by panelists. The final descriptions that were adopted by NAGB are reproduced in Figures 2 through 4.[3] A set of exemplar items was also selected for each level and grade. To qualify as an exemplar, an item had to meet the criteria that are listed in Figure 5 and it had to be selected by a panel based on the quality of the item, the coverage of content for the set of exemplars as a whole, and the grade appropriateness for items that were used at more than one grade.

As was previously indicated, the development of achievement levels has focused on the question of what students should be able to do in order to be considered to be at the Basic, Proficient, or Advanced levels. For example, the

[3] The empirical studies of content descriptions that follow are based on the final descriptions of the 1992 Mathematics Achievement Levels which appear as Figures 1.6-1.8 in Interpreting NAEP Scales (Phillips et al. 1993), Figures 1.1-1.3 in NAEP 1992 Mathematics Report Card for the Nation and the States (Mullis et al., 1993), and Figure 1 in Bourque (1993). These are not the content descriptions developed by the Mathematics Level-Setting Panel with its 69 members representing mathematics teachers, non-teacher educators, and members of the general public (Figure 2 in Bourque (1993)) nor are they the so-called Revised Draft Descriptions of the Achievement Levels Recommended by the Follow-up Validation Panel (Figure 3 in Bourque (1993)). The final descriptions were revised by the entire group of Validation Panel members "to provide more within- and across-grade consistency and to align the language of the description more closely with the language of the NCTM Standards" (Bourque, 1993, p. 12).

description of the grade 4 Proficient level states that students at that level "should consistently apply integrated procedural knowledge."

Because it is stated that students at a given level "**should**" do something or have a given level of understanding, it does not necessarily follow that students at that level **actually do** that activity or possess the stated level of understanding. The grade 4 Proficient level, for example, ranges from 248 to 279 on the NAEP scale. It is essential to ask whether or not grade 4 students who score in that range have a substantial probability of correctly answering items on NAEP that are selected as exemplars of that level or that correspond to the description of the grade 4 Proficient level.

Study Description

Three general approaches were used to investigate the degree to which the descriptions of the NAEP achievement levels provide a valid indication of the actual performance of students at each of the achievement levels:

1. **Exemplar Items Analysis** – The statistical properties of the items that were selected as exemplars were reviewed.

2. **Classification of Items Based on Statements in the Levels Descriptions** – The NAGB descriptions of the levels were used to form a list of statements about what students at a given level should be able to do. Judges (mathematics educators familiar with the curriculum at the target grade levels) then used those statements (without being told the level from which the statement was taken) to identify items that called for the knowledge, skill, or understanding contained in the descriptor-based statements, and the performance of students at each level on the identified items was summarized.

3. **Classification of Items Based on Statistically Differentiating Student Performance** – Items that discriminate among achievement levels were identified by statistical criteria and the content of the items (as identified by curriculum experts) was compared to the descriptions of the corresponding achievement levels.

Review of Exemplars

The review of the statistical properties of items selected as exemplars focused on the proportion of students performing at each achievement level that correctly answered the exemplar items. It was reasoned that students who perform at, say, the Proficient level should have a reasonably high probability of correctly answering an exemplar item for that level and an even higher probability of correctly answering items selected as exemplars for the Basic level. Similarly, one would not expect students at one level to have a very high success rate on items that are used to exemplify a higher level. Informed observers, of course, may differ about what rate of success is "reasonably high."

Based on advice from their Technical Advisory Committee on Standard Setting, ACT used a minimum of .501 for the percent of borderline students **expected** to correctly answer a given item. The criterion used in the past for selection of anchor items has been a minimum proportion correct of .65, and the initial screen used by ACT prior to public comment forums was .80. Regardless of the minimum value that is used, it is important to evaluate the actual proportion correct for the exemplars because the expected proportion correct based on ratings of judges is not necessarily the same as the actual proportion of students who correctly answered the items. We also thought it was important to evaluate the proportion correct for the complete range of students performing at a given achievement level and not just those at the cut point between levels.

Classification of Items Based on Statements in the Descriptions of Levels

Reckase (1992) argued that the description of a skill such as those contained in the NAGB Achievement level descriptions "defines a domain of items." Using the example of the skill "perform operations involving polynomials," Reckase went on to note that

> there are a very large number of items that match that description, and they vary in difficulty and discrimination over a fairly wide range. What is meant when someone says that students at the proficient level can perform the necessary operations is that if students at that level were given a random sample of items from that domain, they would answer a high proportion of them correctly. However, it does not mean that they would be able to answer the hardest one correctly with high probability. (Reckase, 1992, p. 1)

We are in agreement with this statement by Reckase. Indeed, our second approach is based on essentially the same logic. We would add the qualification, however, that the proportion of items answered correctly by students scoring below the level associated with the descriptive statement should be substantially lower than that for the level with which the statement is associated.

Although we could not create random samples of items from domains corresponding to the descriptive statements, we were able to use the statements to define subsets of items in the pool of items administered in 1992. That is, the descriptions of the levels were used to create a series of statements about students' performance and those statements were used by judges to classify the NAEP items according to whether or not they called for the activity, skill, or understanding in question. Once classified, a variety of item statistics were summarized for students scoring within the range at each achievement level.

The task of creating statements ("descriptors") to map to test items is not a straightforward one. There are a number of issues in the design of this activity that can affect the results of the study and hence the inferences based on it. The issues considered in the design were:

1. How to present the text of NAGB's descriptions—Present entire content descriptions for each level, or present elements of the descriptions that represent particular types of knowledge or skills.

2. How to describe the task to judges—Tell judges that the goal is to determine which level an item represents or have them match the items to elements of the NAGB descriptions without being aware of the levels.

3. How to extract elements from the paragraph descriptions—Decompose phrases such as "conceptual and procedural" by separating "conceptual" from "procedural" or leave them combined.

4. How to preserve the language used in the content descriptions.

5. How to maintain the independence of judgments made by different judges.

The decisions made with regard to these points are contained in the description of the instrument development and of the sample of judges that follows.

Development of instruments. A number of possible approaches to mapping the achievement levels to sets of test items were considered. For example, the paragraph descriptions could have been left intact and judges could have been asked to sort the items into three groups, each group consisting of the items that represented the knowledge and skills described in one of the achievement level descriptions. However, if this strategy had been used, the basis for the classification by judges would have been unclear. Two judges using very different criteria for assignment might make the same classification decision. Likewise, judges who interpret an attribute (e.g., "real-world problem solving") differently, but weight the importance of that attribute differentially might make the same classification decision. Moreover, we believed that the judges' task would be less difficult and that more reliable ratings would be obtained if the paragraphs were "decomposed" so that judges could map distinguishable elements of the descriptions to items. Additionally, to avoid biasing or confounding our results, judges were not given any information about either the existence of achievement levels or the identity of the achievement level from which each description element was taken.[4] Finally, the language of the ACT/NAGB descriptions was altered as little as possible in decomposing them into individual elements.

The actual task presented to judges for mapping test items to elements of the NAGB achievement-levels descriptions was constructed as follows.

1. The paragraph descriptions from the three achievement levels at each grade (Figures 2-4) were parsed into clauses that represented distinct

[4] The type of mapping we are investigating should be informed by the descriptive statement about the content rather than by the particular level from which the statement is taken or by the label given that level.

mathematical knowledge, understandings or skills that could be required to answer an individual test item.

Figure 2

Description of Mathematics Achievement Levels for Basic, Proficient, and Advanced Fourth Graders*

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Fourth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra and functions. At the fourth-grade level, algebra and functions are treated in informal and exploratory ways, often through the study of patterns. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 211	Fourth-grade students performing at the basic level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content areas.

Fourth graders performing at the level should be able to estimate and use basic facts to perform simple computations with whole numbers; show some understanding of fractions and decimals; and solve some simple real-world problems in all NAEP content areas. Students at this level should be able to use—though not always accurately—four-function calculators, rulers, and geometric shapes. Their written responses are often minimal and presented without supporting information.

* SOURCE: Figure 1.1, Mullis, I.V.S. et al., (1993) NAEP 1992 Mathematics Report Card for the Nation and the States, p. 44.

Figure 2 (continued)

Proficient 248	Fourth-grade students performing at the proficient level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth graders performing at the proficient level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the proficient level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved.

Advanced 280	Fourth-grade students performing at the advanced level should apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth graders performing at the advanced level should be able to solve complex and nonroutine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. These students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely.

Figure 3

Description of Mathematics Achievement Levels for Basic, Proficient, and Advanced Eighth Graders

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Eighth Graders*

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 256	Eighth-grade students performing at the basic level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.

Eighth graders performing at the basic level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level should also be able to use fundamental algebraic and informal geometric concepts in problem solving.

As they approach the proficient level, students at the basic level should be able to determine which of available data are necessary and sufficient for correct solutions and use them in problem solving. However, these 8th graders show limited skill in communicating mathematically.

* SOURCE: Figure 1.2, Mullis, I.V.S. et al., (1993), NAEP 1992 Mathematics Report Card for the Nation and the States, p. 51.

Figure 3 (continued)

Proficient 294	Eighth-grade students performing at the proficient level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.

Eighth graders performing at the proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical solutions.

Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs; apply properties of informal geometry; and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced 331	Eighth-grade students performing at the advanced level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.

Eighth graders performing at the advanced level should be able to probe examples and counter-examples in order to shape generalizations from which they can develop models. Eighth graders performing at the advanced level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

Figure 4

Description of Mathematics Achievement Levels for Basic, Proficient, and Advanced Twelfth Graders*

Description of Mathematics Achievement Levels for
Basic, Advanced, and Proficient Twelfth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 287	Twelfth-grade students performing at the basic level should demonstrate procedural and conceptual knowledge in solving problems in the five NAEP content areas.

Twelfth-grade students performing at the basic level should be able to use estimation to verify solutions and determine the reasonableness of results as applied to real-world problems. They are expected to use algebraic and geometric reasoning strategies to solve problems. Twelfth graders performing at the basic level should recognize relationships presented in verbal, algebraic, tabular, and graphical forms; and demonstrate knowledge of geometric relationships and corresponding measurement skills.

They should be able to apply statistical reasoning in the organizations and display of data and in reading tables and graphs. They should be able to generalize from patterns and examples in the areas of algebra, geometry, and statistics. At this level, they should use correct mathematical language and symbols to communicate mathematical relationships and reasoning processes; and use calculators appropriately to solve problems.

* SOURCE: Figure 1.3, Mullis, I.V.S. et al., (1993), NAEP 1992 Mathematics Report Card for the Nation and the States, p. 56.

Figure 4 (continued)

Proficient 334	Twelfth-grade students performing at the proficient level should consistently integrate mathematical concepts and procedures to the solutions of more complex problems in the five NAEP content areas.

Twelfth-grade students performing at the proficient level should demonstrate an understanding of algebraic, statistical, and geometric and spatial reasoning. They should be able to perform algebraic operations involving polynomials; justify geometric relationships; and judge and defend the reasonableness of answers as applied to real-world situations. These students should be able to analyze and interpret data in tabular and graphic form; understand and use elements of the function concept in symbolic, graphical, and tabular form; and make conjectures, defend ideas, and give supporting examples.

Twelfth-grade students performing at the advanced level should understand the function concept; and be able to compare and apply the numeric, algebraic, and graphical properties of functions. They should apply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous and discrete mathematics.

They should be able to formulate generalizations and create models through probing examples and counter examples. They should be able to communicate their mathematical reasoning through the clear, concise, and correct use of mathematical symbolism and logical thinking.

An illustration of the statements that were abstracted from the achievement level descriptions for judges to use in classifying items is given below.

> The item calls for use of basic number facts to perform simple computations with whole numbers.

This statement is based on the NAGB description of the grade 4 Basic level:

> Specifically, 4th grade students performing at the basic level should be able to estimate and use basic facts to perform simple computations with whole numbers.

An effort was made to make all the statements about what an item "calls for" correspond as closely as possible to the wording in the NAGB achievement level descriptions. Hence, in the above example, except for the insertion of the word "of," the phrase "use basic facts to perform simple computations with whole numbers" appears both in the NAGB achievement level description and the statement that was created for this study for judges to use in classifying items.

The above quote from the NAGB achievement level descriptions not only says that students should be able to "use," but states that they should be able to "estimate." This was accommodated by a second statement that judges used to classify items:

The item calls for estimation of simple whole number results.

2. The resulting lists of statements (henceforth called "descriptors") were rearranged so that descriptors that related to similar content areas were grouped together, regardless of achievement level. Descriptors that related to aspects of the same content, for example, whole numbers or geometry, were subsumed as sub-descriptors of a higher level descriptor which asked whether the item involved that content area. Similarly, descriptors that related to aspects of written responses and problem solving were presented as sub-descriptors. For example, in the 8th-grade instrument, the written-response descriptors were as follows:

"If the item requires a written response, check any of the following descriptions that apply:

The item calls for:

22(a) making conjectures

22(b) defending ideas

22(c) giving supporting examples

22(d) explaining the reasoning process underlying conclusions

22(e) conveying underlying reasoning skills beyond the level of arithmetic."

3. A number of versions of the instruments were piloted and revised before arriving at the final versions which are in Appendix B[5] along with the parsed versions of the achievement level descriptions from which they were derived. The final version of the instruments maintained the exact language of the NAGB achievement-level descriptions, unless there were semantic difficulties in leaving parsed clauses intact but separate.[6] When a clause had the connector "and" (depicting intersection of knowledge and skill types), it was typically switched to "or" so that an item requiring either knowledge or skill would be matched to that descriptor.

The final instruments covered the knowledge and skills mentioned in the NAGB descriptions nearly completely. The attributes that were not included in the instruments were of several specific types. One exception was references to the use of calculators, rulers and geometric shapes.[7] A second category of omissions were phrases that could not be viewed as a characteristic

[5] In the final instruments, a question mark was placed after each descriptor so that judges could indicate uncertainty in their mapping of a particular descriptor to an item. However, there appears to be no systematic benefits from taking the reported uncertainty data into account.

[6] Some verbs were converted to gerunds (for example, "apply" became "applying").

[7] The items that require students to use calculators are grouped in particular blocks and so it is obvious which items call for the use of calculators; it is also obvious which items require use of geometric shapes and rulers.

of a single item. For example, a number of phrases referred to demonstrating a skill "...in the five NAEP content areas." Finally, a few phrases referred to qualities of student performance rather than to skills or knowledge; for example, that students should be able to "use ... appropriately" or "display mastery in the use of ...."

Table 1 indicates the number of descriptors that related to each achievement level description in each grade level.

Sample of judges. For each grade level, a group of six mathematics educators (teachers or former teachers), who were familiar with the content of the mathematics curriculum at that grade level, were recruited and trained to examine each test item and select the descriptors that described the knowledge or skills that "the item called for." A summary of the background characteristics of these judges is presented in Appendix C.[8]

Table 1

Number of Descriptors Abstracted from NAGB Descriptions by Achievement Level and Grade

	Achievement level			
Grade	Basic	Proficient	Advanced	Total
4	5	9	4	18
8	8	17	6	31
12	14	14	7	35

[8] The summary of the teacher background questionnaire data was prepared by Audrey McEvans.

Data collection. Each judge received a binder that contained fourteen blocks of 1992 NAEP mathematics test items at one grade level. These binders contain all the item blocks administered to the main NAEP sample; blocks used only for trend analysis purposes or other special studies were not included. Judges working at the 4th-grade level had 178 items to judge; judges working at the 8th-grade level considered 211 items; and judges working at the 12th-grade level covered 208 items. At each grade level, half of the judges received the blocks of items in reverse order.

Judges were told that they were participating in a study whose purpose was to determine the mathematics knowledge and skills that are being assessed by the NAEP mathematics test items. They were asked to **"use their own professional judgment in deciding which descriptors applied to each item and to interpret the descriptors in light of their experience of 4th-, 8th- or 12th-grade mathematics content and students."** The judges were told that there were no right or wrong decisions regarding which descriptors mapped to any item and were encouraged to select as many of the descriptors as applied to each item. However, to ensure independence in judgment, they were told not to discuss the descriptors or test items with any other judge.

The task of mapping items to descriptions took the judges an average of seven hours to complete. On completion of the task, all judges were asked to write their impressions of the activity and approximately half of the judges were interviewed.[9]

---

[9] While the entire authorship team contributed to the design of this substudy, Brenda Sugrue and John Novak were operationally responsible for the preparation of study materials and supervision of the data collection. Interviews were conducted by Reggie Stites who also observed and prepared field notes for portions of the data collection. A more detailed description of the design and results of this substudy is contained in Sugrue et al. (forthcoming).

Mapping items to descriptors and levels. A critical decision was what constituted a match between an item and a descriptor; that is, how to determine whether an item mapped to a descriptor and, through the descriptor's location in the NAGB content descriptions, to an achievement level. We considered several possible decision rules (requiring that at least 4, 5, or all 6 judges map the item to the descriptor) and examined their empirical consequences.[10] In the end, a criterion that **at least four of the six judges assign the descriptor to the item to consider it a mapping** was chosen.[11]

With the chosen decision rule on item-descriptor mapping, each item was initially classified as representing an achievement level if at least four of the six judges assigned at least one descriptor from the particular achievement level to the item. Thus, an item could be assigned to more than one achievement level or, indeed, to no achievement level if there was no descriptor that was assigned to it by at least four judges.

To obtain single level classifications of items, **each item was then assigned to the highest achievement level from which even one descriptor was mapped to the item by four or more judges**. This approach assumes that if an item calls for multiple skills, then it is the most advanced of those skills that limits performance on the item. In the analyses that follow, we examined the results from both of these classifications.

[10] The choice of either 4, 5, or 6 judges agreeing that the descriptor mapped to a given item emphasizes a consensus judgment arrived at independently and without training, albeit with different levels of stringency. Note that the judgment is a symmetric one. That is, saying that 0, 1, or 2 judges concluded that a descriptor did not characterize an item represents a consensus that there was no match.

[11] We were concerned that a more stringent cutoff of complete (6) or almost complete (5) agreement would be too stringent. These cutoffs might lead to too few items being mapped to any descriptor or too few descriptors with more than a handful of items mapped to them.

Classification of Items Based on Statistically Differentiating Student Performance

In the third approach, items were classified according to item statistics for students scoring within the range of a given achievement level. Subsets of items for which students at that given achievement level had a high probability of answering the item correctly, and which students at the next lower achievement level were substantially less likely to answer correctly, were then reviewed by mathematics teachers and subject matter experts to see how well they corresponded to the NAEP achievement level descriptions.

Statistical criteria for differentiation. The type of item statistics used in the analyses followed closely the procedure used to define items that correspond to NAEP anchor points.[12] First, the proportion of students at each achievement level who correctly answered an item was obtained for each item. These proportion correct values are referred to as p-values. There are four such p-values for each item administered at a given grade (one each for students with proficiencies that were classified as Below Basic, Basic, Proficient, and Advanced). From the above description of the grade 4 Basic level in mathematics, for example, one would expect to find relatively high p-values for students performing at the Basic level or higher on items requiring students to perform simple computations with whole numbers.

[12] The procedure used in the past to define items that correspond to the NAEP anchor points applies the following rules. A 25 point score range that is centered at each anchor point (e.g., 237.5 to 262.5 for the 250 anchor point) is used to compute probabilities of a correct response for students with proficiencies near the anchor points. An item that is an anchor at a given level must meet specific criteria. For example, an item that anchors level 250

1. must have a p-value of at least .65 for students at the 250 level (i.e., 237.5 to 262.5),
2. the p-value for students at the 200 level must be at least .30 less than the p-value for students at the 250 level,
3. at least 50 percent of the students at the 200 level must get the item wrong,
4. there must be at least 100 students at 200 and 250 levels.

The NAEP rules were then adapted to identify items which differentiate between achievement levels. For an item to be said to differentiate at a Basic level, it had to satisfy the following conditions.

1. The p-value for students at the Basic level must be at least .65.
2. The p-value for students from the Below Basic level must be at least .30 less than the p-value for Basic students.
3. At least 50% of the Below Basic students must get the item wrong.

The requirement of a minimum sample size of 100 was not used. Such a requirement would always be met for the Below Basic, Basic, and Proficient categories, but would be problematic at the Advanced level since relatively few students perform at that level. To determine whether the application of the above three criteria was too stringent, we later relaxed the requirement that there be at least 30% difference in percent correct between the higher and the lower group. This relaxation essentially meant that there was a minimum difference of greater than 15% between the two groups while less than a majority of the lower group answered correctly. Parallel definitions were used for the Proficient and Advanced achievement levels.

Characterization of the differentiating test items.[13] Items that met the statistical criteria were then reviewed by mathematics experts[14] who were asked to identify all the content that the item measured, a task which is termed

[13] The data collection from the mathematics experts was supervised by John Novak.

[14] The mathematics experts used at this stage in the item review were primarily advanced graduate students with appropriate prior teaching experience in mathematics. All experts were familiar with the NCTM Standards and most had advanced training in assessment design, psychometrics, and cognitive psychology in addition to their mathematics education expertise. The item reviews described in the text were conducted on several occasions with at minimum three-person teams of experts whose item coding decisions were based on consensus judgments.

identifying the item's "signature" in the parlance of the Third International Mathematics and Science Study (TIMSS; Report #57, Survey of Mathematics and Science Opportunities, 1993). The experts were not required to restrict their coding to the descriptor statements from the achievement level descriptions but were encouraged to indicate all appropriate attributes that fit the item. In all, the experts identified 45 separate content descriptors that applied to at least one test item.

In addition to identifying the mathematics content tapped by each differentiating item, a coding system was developed to examine the linguistic features of the NAEP mathematics items that differentiated between achievement levels.[15] The coding system used was largely adapted from linguistic feature categories developed by Spanos, Rhodes, Dale, and Crandall (1988). Table 2 lists the linguistic features coding categories. The actual coding instrument used is contained in Appendix D. The instrument was applied by a subset of five of the mathematics experts in conjunction with their coding of mathematics content. The assignment of a code to a particular item was based on consensus judgment.

---

[15] The coding system used in this work was developed initially by Cesar Larriva. Larriva and Novak supervised the linguistic coding activity, and Larriva prepared initial summaries of the results.

Table 2

Linguistic Features Coded for Test Items

1a.	NUMBER OF COMPARATIVES
1b.	LOGICAL CONNECTORS
2a.	MATHEMATICAL VOCABULARY
2b.	NATURAL LANGUAGE VOCABULARY WITH DIFFERENT OR SPECIALIZED MEANING IN MATHEMATICS
2c.	COMPLEX STRING OF WORDS OR PHRASES
2d.	WORDS WHICH SIGNAL OPERATIONS
3.	CONCEPTS REQUIRING EXPERIENCE OR KNOWLEDGE
4a.	WORDS WHICH FUNCTION AS UNITS OR HAVE QUANTITATIVE ATTRIBUTES
4b.	QUANTITIES EXPRESSED IN WRITTEN TEXT

In all, then, three different methods of characterizing the items that differentiate among the achievement levels were used; namely,

1. **Achievement Level Descriptors**—Content descriptors from the achievement level descriptions that were consistently mapped to the items by the judges in our second approach.

2. **Item Signatures** —Descriptors identified by the mathematics experts as contributing to each item's signature.

3. **Linguistic Features**—Linguistic features of each test item as identified by the mathematics experts.

Various combinations of these coding systems were used to examine how the items which differentiated among achievement levels might be characterized.

Results

Review of Exemplar Items

How adequate are the items that were selected as exemplars? To answer this question, we reviewed the item statistics of the exemplar items presented in the November 9, 1992 ACT draft final report.

The selection of exemplar items for each achievement level is described by ACT (1993) on pages 52 to 55, by Bourque (1993) on pages 9 to 11, by Phillips et al. (1993) on pages 49 and 50 and by Mullis et al. (1993) on pages 42 and 43. The draft ACT final report and the NAEP 1992 Mathematics Report Card for the Nation and the States (Mullis et al., 1993, pp. 45-63) listed a total of 10 exemplars for the three grade 4 achievement levels (2 for Basic, 5 for Proficient, and 3 for Advanced). At grade 8 there were 8 exemplar items (3, 3, and 2 at the Basic, Proficient and Advanced levels, respectively) and at grade 12, there were 11 exemplar items, the majority (7) of which were at the Basic level with 2 each at the Proficient and Advanced levels. The criteria used to select exemplar items for the achievement levels are listed in Figure 5.

Item statistics for the exemplar items are displayed in Tables 3, 4, and 5 for grades 4, 8, and 12, respectively. The Tables list the proportion of students in each of four score ranges, Below Basic, Basic, Proficient, and Advanced, who correctly answered each exemplar item. For example, the first Basic item referred to here as exemplar B1 (NAEP ID number M022801) at grade 4, was answered correctly by .21 of the students who scored below the cut score for the Basic level, by .64 of the students who scored in the Basic achievement level range of scores, by .92 of the students in the Proficient range, and by .99 of the students scoring at or above the minimum score for the Advanced achievement level.

Figure 5

Criteria for the Selection of Exemplar Items for the NAGB Achievement Levels

"... for an item to be chosen as a possible exemplar for the Basic achievement level,

(1) The expected p-value for students at the cut point for the Basic level of achievement had to be greater than 0.51.

(2) The content of the item had to match the content of the operationalized description of Basic; and

(3) The empirical p-value for the item had to be higher than the empirical p-value for items selected as exemplars for the Proficient level."*

For items to be chosen as a possible exemplar at the Proficient and Advanced levels, the items had to meet parallel requirements.

* Bourque (1993), pp. 9-10.

Table 3

Proportion Correct for Students Performing in Each Achievement Level Range for Items Selected as Exemplars for the NAEP Achievement Levels at Grade 4

ID #	NAEP ID #	Exemplar item for achievement level	Proportion correct for students performing at the achievement level			
			Below basic	Basic	Proficient	Advanced
B1	(M022801)	Basic	.21	.64	**.92**	.99
B2	(M044601)	Basic	.22	.49	**.70**	.89
P1	(M022001)	Proficient	.18	.19	.54	**.97**
P2	(M022802)	Proficient	.31	**.75**	.92	.87
P3	(M022901)	Proficient	.25	.35	.60	**.91**
P4	(M044201)	Proficient	.26	.45	**.74**	.95
P5	(M048701)	Proficient	.04	.21	.48	**.75**
A1	(M022401)	Advanced	.31	.53	**.75**	.95
A2	(M023101)	Advanced	.13	.18	.48	**.90**
A3	(M049001)	Advanced	.00	.08	.29	.59

Note. Number in parenthesis is the NAEP identification number. Numbers in bold represent the level at which percent correct first exceeds statistical criterion of 65%.

Table 4

Proportion Correct for Students Performing in Each Achievement Level Range for Items Selected as Exemplars for the NAEP Achievement Levels at Grade 8

ID #	NAEP ID #	Exemplar item for achievement level	Proportion correct for students performing at the achievement level			
			Below basic	Basic	Proficient	Advanced
B1	(M045101)	Basic	.35	.64	**.83**	.94
B2	(M054701)	Basic	.61	**.83**	.91	.97
B3	(M023701)	Basic	.11	.37	.62	.81
P1	(M049601)	Proficient	.37	**.67**	.90	.97
P2	(M054801)	Proficient	.25	.54	**.73**	.82
P3	(M054901)	Proficient	.16	.19	.36	**.65**
A1	(M049401)	Advanced	.09	.21	.48	**.79**
A2	(M049801)	Advanced	.01	.07	.16	.42

Note. Number in parenthesis is the NAEP identification number. Numbers in bold represent the level at which percent correct first exceeds statistical criterion of 65%.

Table 5

Proportion Correct for Students Performing in Each Achievement Level Range for Items Selected as Exemplars for the NAEP Achievement Levels at Grade 12

ID #	NAEP ID #	Exemplar item for achievement level	Proportion correct for students performing at the achievement level			
			Below basic	Basic	Proficient	Advanced
B1	(M024801)	Basic	.36	**.83**	.98	.99
B2	(M057401)	Basic	**.83**	.93	.98	.95
B3	(M057402)	Basic	.48	**.76**	.88	.96
B4	(M057403)	Basic	.56	**.81**	.93	.88
B5	(M057404)	Basic	.33	.64	**.86**	.96
B6	(M060901)	Basic	.46	**.79**	.92	.99
B7	(M061001)	Basic	.26	.56	**.85**	.93
P1	(M024701)	Proficient	.01	.30	**.89**	.98
P2	(M057301)	Proficient	.52	**.83**	.97	.98
A1	(M057901)	Advanced	.18	.29	.55	**.84**
A2	(M061501)	Advanced	.09	.13	.57	**.92**

Note. Number in parenthesis is the NAEP identification number. Numbers in bold represent the level at which percent correct first exceeds statistical criterion of 65%.

Even using a relatively lenient criterion, a number of the exemplar items showed low enough rates of success that it is hard to defend their use as exemplars. For example, if one used the lenient criterion that half of all the students performing at a given achievement level should answer the item correctly to qualify as an exemplar, two items at grade 4 (Basic item 2 and Proficient item 5) and three at grade 8 (Basic item 3, Proficient item 3 and Advanced item 2) would fail to qualify as exemplars for their designated levels. Using a more stringent criterion of, say, 65% (consistent with the NAEP anchor item criteria) would disqualify an additional seven exemplar items across the three grade levels. In other words, approximately one-third of the exemplar items across the three grades failed to satisfy a requirement that two-thirds or more of the students at a given achievement level actually answered correctly the item intended to illustrate what students at that level can do.

The results indicate that a substantial number of exemplars show too low a rate of success among students at the relevant level to be reasonable as exemplars. Consequently, it is often impossible to distinguish between items identified as exemplars for one achievement level from those selected as exemplars for another level in terms of actual student performance. This may not be surprising, given that the statistical criteria were only one basis for selecting exemplars: **"Although a statistical filter was used to select the items for consideration, the primary criterion was a good match between the content of an item and the description of the level it represented"** (Phillips et al., 1993, p. 49).

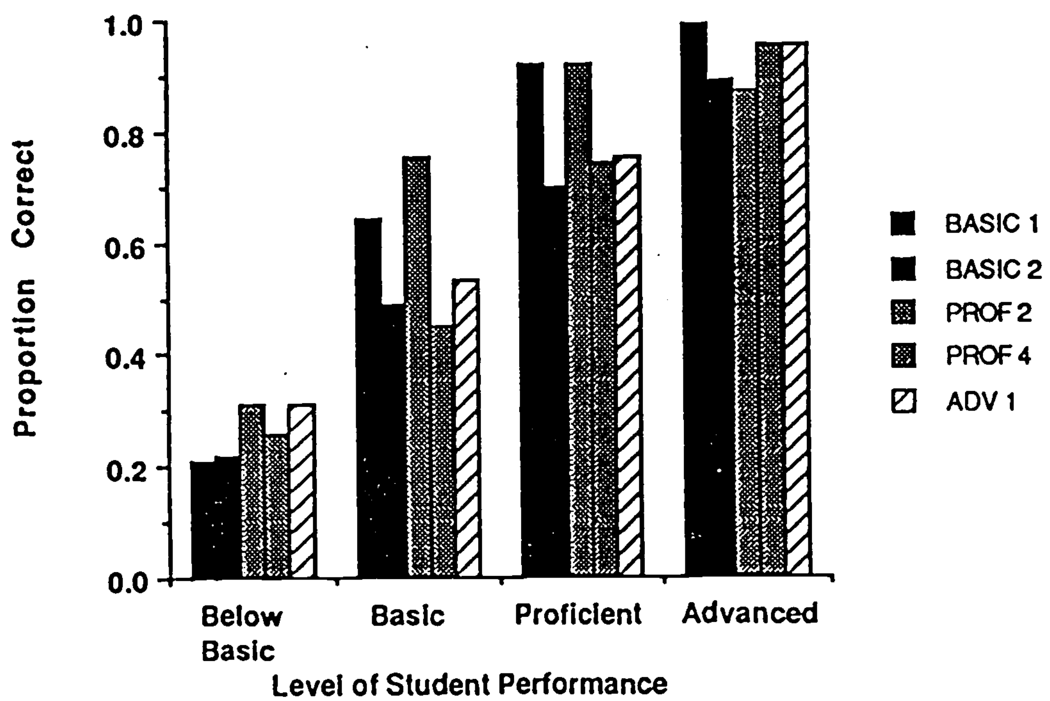
In three cases, items selected as exemplars of the Proficient and Advanced levels are actually easier for students at the Basic level than are some of the Basic exemplars (Table 3). Indeed, the easiest item for grade 4 students performing at the Basic level is the second exemplar item (P2)[16] for the Proficient level. In order of increasing difficulty, the five easiest items for grade 4 students performing at the Basic level and their associated p-values are P2 (.75), B1 (.64), A1 (.53), B2 (.49), and P4 (.45).

The p-values for the five just identified items are plotted in Figure 6 as a function of the achievement level of the students. On the basis of actual student performance, it is apparent in Figure 6 that the three items that are classified as exemplars for the Proficient and Advanced levels could just as well be classified as exemplars of the Basic level.

[16] In the tables and text B refers to Basic exemplars, P to Proficient, and A to Advanced. The number that follows corresponds to exemplar item numbers as used in Mullis et al. (1993, pp. 45-63).

Figure 6

Proportion Correct by Achievement Level for Grade 4 Exemplar Items Selected to Illustrate Proficient and Advanced Exemplars that Are Statistically Similar to Basic Exemplars



A plot of grade 4 Proficient exemplars P1 and P5 and Advanced exemplars A1 and A2 in Figure 7 reveals an equally confusing picture in terms of the actual performance of students at the different achievement levels. As can be seen, grade 4 students performing at the Proficient level are more likely to answer exemplar item A1 correctly (.75) than they are Proficient exemplars P1 (.54), or P5 (.48).

Based on the item statistics in Table 3 that are displayed graphically in Figures 6 and 7, a case could be made that Basic exemplar B2 would make a better exemplar of either the Proficient or Advanced levels than of the Basic level. On the other hand, Proficient exemplar P2 would make a better exemplar of the Basic level than of the Proficient level, while Advanced exemplar A1 might better serve as an exemplar of the Proficient level.

An inspection of Tables 4 and 5 reveals inconsistencies at grades 8 and 12 that are similar to those that were found for the grade 4 exemplars. Note, for example, that at grade 8 (Table 4), exemplar P1 has a p-value of .67 for students scoring at the Basic level. The p-values for two Proficient exemplar items (P1 and P3) are plotted together with the corresponding p-values for Basic exemplar item B1 and Advanced exemplar item A1 in Figure 8. An inspection of Figure 8 suggests that, in terms of actual student performance, Basic exemplar item B1 and Proficient exemplar item P1 should be classified at the same level (either Basic or Proficient). Based on actual student performance, Proficient exemplar item P3 and Advanced exemplar item A1 should also be classified at the same level (Advanced).

Figure 7

Proportion Correct by Achievement Level for Grade 4 Exemplar Items Selected to Illustrate Advanced Exemplars that Are as Easy or Easier than Selected Proficient Exemplars

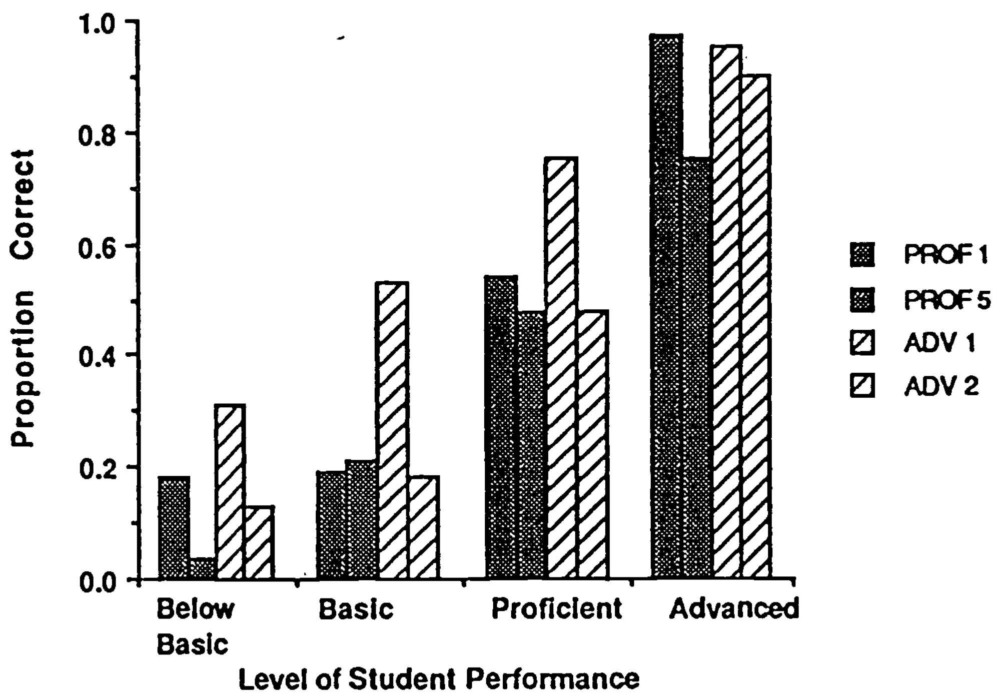
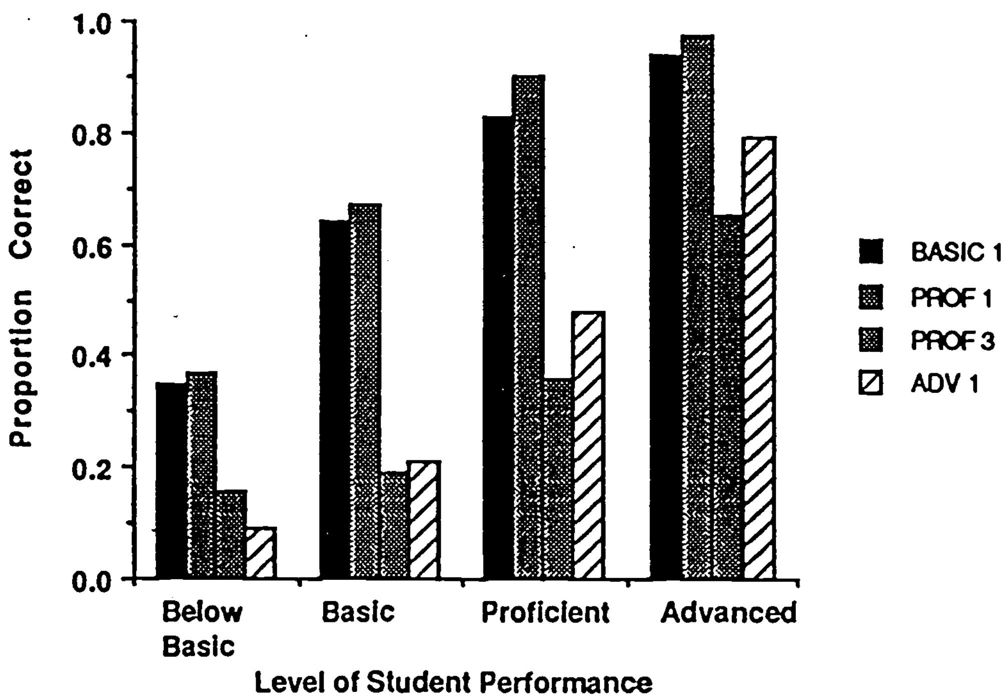


Figure 8

Proportion Correct by Achievement Level for Selected Grade 8 Exemplar Items



As can be seen in Table 5, Proficient exemplar item P2 at grade 12 has a p-value of .83 for students scoring at the Basic level. That p-value is as high or higher for those students than the p-values for all but one of the seven exemplar items for the grade 12 Basic level. Figure 9 provides a graphic comparison of the p-values of Proficient exemplar P2 with four of the Basic exemplar items for students performing at each of the achievement levels. As can be seen, P2 has a higher p-value than any of the four Basic exemplars for students performing at the Below Basic, the Basic, or the Proficient levels. Based on these item statistics, it is quite unclear why items B3, B5, B6, and B7 should be exemplars for the grade 12 Basic level while item P2 is an exemplar of the Proficient level.

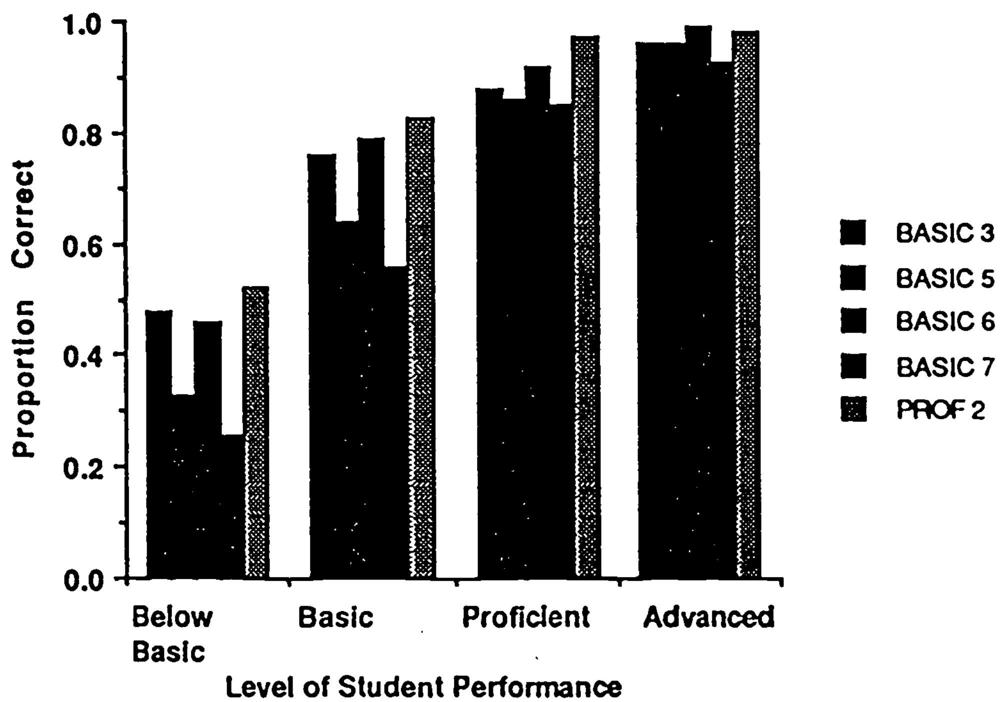
The exemplars should provide elaboration of the descriptions of the knowledge, skills, and understandings that students at a given level have achieved. To do so, they must show reasonably high rates of success among students at the appropriate levels. In addition, the probability of a correct response should be substantially lower for students performing at a lower achievement level because it is implicit that students performing below the level being exemplified generally lack the knowledge or skill that the item requires. Judged in terms of actual student performance, a substantial number of the items that were selected as exemplars are poorly suited for that role.

Item Classification From Levels Descriptions

Analysis of mapping of items to descriptors and levels. The mapping of items to descriptors and levels produced a considerable amount of information from each judge and across judges. Essentially, the 6 judges at grade 4 each made 3204 (178 items x 18 descriptors) decisions mapping items to

Figure 9

Proportion Correct by Achievement Level for Selected Grade 12 Exemplar Items



descriptors. The corresponding numbers for judges at grades 8 and 12 were 6541 (211 x 31) and 7280 (208 x 35), respectively. Analysis criteria that focus on the critical features of the data were applied to avoid getting bogged down in minor details of the data.[17]

Applying the criterion that at least 4 judges assigned the descriptor to the item to consider it a mapping, there were 28 (out of 178) 4th-grade items, 2 (out of 211) 8th-grade items, and 34 (out of 208) 12th-grade items that were not mapped to any descriptor. Hence, these items could not be mapped to any of the achievement levels.

The number of items mapped to each descriptor by at least 4 judges at each grade level are presented in Tables 6-8. In these tables, the descriptors have been sorted by achievement level and then by the number of items mapped to the descriptor.[18] The number of items where the judges' opinions were evenly divided (3 yes, 3 no) about whether it mapped to the descriptor is also

[17] We conducted a series of large-scale generalizability analyses of the mappings of descriptors to the items by the judges. The purpose was to examine the variability (technically, the variance components) associated with judges, descriptors, assessment items (classified variously by content and by item format and type) and their interactions. These generalizability analyses are reported briefly in Appendix E. More details of the analyses can be found in Novak, Burstein, and Sugrue (forthcoming). Generally, large sources of variability, especially at grades 4 and 8, were descriptors and interactions of judges with descriptors. There was considerable variability in judges' interpretations of some clusters of descriptors.

[18] At grade 4, descriptor D3 (understanding of mathematical concepts or mathematical procedures) and at grade 8, descriptor D15 (conceptual understanding or procedural understanding) were excluded because they were mapped to almost every item by at least four judges. In retrospect the decision to leave mathematics concepts and mathematics procedures combined and conceptual and procedural understanding combined in a single descriptor was an unfortunate decision. Judges rightly concluded that virtually every item at these grades involved either conceptual or procedural understanding and responded accordingly. As a consequence these descriptors could not inform the mapping of items to achievement levels and thus were excluded.

Table 6

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, and by 3 Judges, Grade 4

Descriptor ID number	Keywords: Content	Keywords: Process	Level	# of items (4 or more judges)	# of items (3 judges)
D1b	whole numbers	estimating	B	14	10
D1a	whole numbers	simple computation	B	10	39
D2a	fractions	some understanding	B	4	14
D6a	simple real-world problems	problem solving	B	1	13
D4	integrated procedural, conceptual knowledge	application, problem solving	P	87	39
D7	strategies	problem solving	P	70	62
D1c	whole numbers	computation	P	56	20
D6b	real-world problems	problem solving	P	46	18
D2b	fractions, decimals	conceptual understanding	P	13	3
D1d	whole numbers	estimation	P	12	8
D1e	whole numbers	determination of reasonableness of results	P	10	26
D8b		explanation (how)	P	8	1
D8a		giving supporting information	P	6	4
D8d		clear, concise communication	A	11	12
D8c		explanation (why)	A	3	4
D5	complex, nonroutine real-world problems & integrated procedural and conceptual knowledge	problem solving	A	1	2
D6c	complex nonroutine, real-world problems	problem solving	A	1	8

Table 7

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, and by 3 Judges, Grade 8

Descriptor ID number	Keywords: Content	Keywords: Process	Level	# of items (4 or more judges)	# of items (3 judges)
D9	strategies	problem solving (selecting and using strategies)	B	89	51
D8	problems, diagrams, charts, graphs	problem solving	B	81	25
D1	arithmetic operations, whole numbers, decimals, fractions, or percents	understanding, estimation	B	78	55
D6a	informal geometric concepts	problem solving	B	48	16
D10	technological tools (calculators, computers, and geometric shapes)	problem solving	B	42	21
D4	fundamental algebraic concepts	problem solving	B	15	12
D12	data	determining necessity and sufficiency of data	B	2	16
D2	basic-level arithmetic operations	understanding, problem solving	P	108	18
D7	quantity or spatial relationships	problem solving or reasoning	P	77	39
D6b	properties of informal geometry	application	P	49	11
D14a	statistics or probability	calculating results	P	13	2
D3	fractions, percents, decimals	understanding connections	P	10	11
D13a	data or graphs	making inferences	P	9	10
D22e	beyond arithmetic	reasoning	P	9	1
D22b		defending ideas	P	8	7
D22c		giving supporting examples	P	6	4
D19		generating examples	P	6	4

Table 7 (continued)

Descriptor ID number	Keywords: Content	Keywords: Process	Level	# of items (4 or more judges)	# of items (3 judges)
D14b	statistics or probability	evaluating results	P	6	6
D16	concepts and procedures	application, problem solving	P	6	12
D5	algebra and functions	understanding connections	P	3	3
D13b	process of gathering and organizing data	understanding	P	3	4
D22a		making conjectures	P	2	0
D14c	statistics or probability	communicating results	P	2	4
D18	mathematical ideas	comparing and contrasting	P	1	0
D21	number sense	considering reasonableness of answers	A	50	45
D6c	geometry	considering reasonableness of answers	A	16	23
D22d		explaining reasoning process	A	15	2
D11		abstract thinking, creation of problem-solving techniques	A	2	11
D17	mathematical rules, concepts and principles	reaching beyond recognition, identification and application, generalizing, synthesizing	A	1	1
D20	examples and counterexamples	generalizing, developing models	A	1	7

Table 8

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, and by 3 Judges, Grade 12

Descriptor ID number	Keywords: Content	Keywords: Process	Level	# of items (4 or more judges)	# of items (3 judges)
D11	procedural knowledge or conceptual knowledge	problem solving	B	53	57
D1b	geometric relationships and corresponding measurement skills		B	46	13
D2a	algebraic reasoning strategies	problem solving	B	36	23
D1a	geometric reasoning strategies	problem solving	B	22	21
D6	verbal, algebraic, tabular or graphical forms of presentation	recognizing relationships	B	19	30
D4b	tables or graphs	applying statistical reasoning	B	9	4
D14b	mathematical language and symbols	communication of reasoning processes	B	8	5
D14a	mathematical language and symbols	communication of mathematical relationships	B	8	14
D9	real-world problems	estimating to determine reasonableness of results	B	7	8
D2d	algebra	generalizing from patterns or examples	B	5	4
D8	real-world problems	estimation to verify results	B	5	13
D4e	data analysis or statistics	generalizing from patterns or examples	B	2	10
D1f	geometry	generalizing from patterns or examples	B	1	1
D4a	organization and display of data	applying statistical reasoning	B	1	2
D4d	data in tabular or graphical form	analyzing and interpreting	P	18	5
D2b	algebra	understanding, reasoning	P	17	18
D14d		defending ideas	P	16	2
D1d	spatial reasoning	understanding	P	13	9

Table 8 (continued)

Descriptor ID number	Keywords: Content	Keywords: Process	Level	# of items (4 or more judges)	# of items (3 judges)
D1c	geometric reasoning	understanding	P	12	13
D1e	geometric relationships	justifying	P	10	7
D2c	algebraic operations involving polynomials	performing	P	9	9
D14f		giving supporting examples	P	8	6
D4c	statistical reasoning	understanding	P	7	2
D10	real-world situations	judging or defending reasonableness of answers	P	7	14
D3a	elements of the function concept in symbolic, graphical, or tabular form	understanding	P	3	4
D14e		making conjectures	P	1	2
D3c	elements of the function concept in symbolic, graphical, or tabular form	using	P	1	6
D12	concepts and procedures, complex problems	interpreting, problem solving	P	0	0
D14c	mathematical symbolism	clear and concise use, logical thinking, communicating mathematical reasoning	A	5	9
D3e	numeric, algebraic, or graphical properties of functions	application	A	4	1
D3b	the function concept	understanding	A	2	3
D3d	numeric, algebraic, or graphical properties of functions	comparing	A	1	0
D5	continuous and discrete mathematics	problem solving	A	1	1
D13	procedural and conceptual knowledge	integration, synthesis of ideas	A	1	4
D7	examples and counterexamples	formulating generalizations, creating models	A	0	0

reported.[19] Short keyword versions of the descriptors are included in these tables to facilitate interpretation.

Although some aspects of the achievement-levels descriptions are well-represented in the test items, some aspects are hardly represented at all. At grade 4, 7 of the 18 descriptors were mapped to fewer than 9 items; the same was true for 16 of 31 grade 8 descriptors and 24 of 35 grade 12 descriptors.

Relatively few descriptors—scattered among grades and achievement levels—mapped unambiguously to a large number of items. For this purpose, a descriptor had to map to 9 or more items and it should **not** map also to a large number of items which evenly divided the judges. The only descriptors to meet these conditions at grade 4 come from the Proficient level. Only one of the Advanced level descriptors at any level had substantially more items mapped than received ambiguous mappings (Grade 8, descriptor 22d "Explaining the reasoning process underlying conclusions"); the selection of this descriptor was essentially automatic for all extended constructed response items which asked students to "explain your reasons for your answer." Even though there were many more descriptors at grade 12 than at grade 4, there were only a few descriptors that were consistently mapped to a large number of items. The descriptors that mapped involved primarily applying straightforward topic/content terms (geometric relationships and corresponding measurement skills, algebraic reasoning strategies, reading tables and graphs, analyzing and interpreting data in tabular or graphical form); the descriptor involving explicit requests to defend one's ideas in written responses was also consistently mapped.

[19] We interpret the even division among judges on mapping of items to descriptors as evidence of serious disagreements in interpreting a particular descriptor when it occurs frequently, especially when compared to the number of items where at least 4 judges agree on mapping. Otherwise, attributing the diversity of opinion to properties of specific items seems warranted.

The number of items assigned to each achievement level by at least 4 judges is reported in Table 9. Items that were assigned to more than one level are included in the counts for each of the levels to which they were assigned. A considerable number of items were mapped to descriptors from multiple levels; over half the grade 8 items were mapped to descriptors at both Basic and Proficient levels.

The classification of items to single achievement levels that resulted from assigning each item to the highest achievement level from which even one descriptor was mapped to the item by four or more judges is provided in Table 10. Very few 4th-grade items were distinguished as representative of either the basic (4%) or advanced (7%) achievement level descriptions. Very

Table 9

Number of Items Classified to Single or Multiple Achievement Levels

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	6	13	88
Proficient	109	11	24
Basic & Proficient	22	110	49
Advanced	1	0	2
Basic & Advanced	0	10	1
Proficient & Advanced	12	7	3
Basic & Proficient & Advanced	0	8	7
Total	178	211	208

Note. Items were classified to a given level if at least 4 out of 6 judges mapped at least one descriptor from the particular achievement level to the item.

Table 10

Number of Items Classified to Highest Single Achievement Level

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	6	13	88
Proficient	131	121	73
Advanced	13	75	13
Total	178	211	208

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges

few 8th-grade items were classified as involving descriptors from any of the basic achievement levels (6% Basic), and very few 12th-grade items were designated as involving the advanced level descriptors (6%). 42% of 12th-grade items were classified as Basic. Higher percentages of 4th-grade and 8th-grade items (74% and 57% respectively) than 12th grade items (35%) were classified as Proficient.

The task of mapping descriptors to test items was made difficult by the fact that (a) many of the descriptors (taken directly from the NAGB descriptions) were ambiguous, and (b) many descriptors from different levels were very similar. Judges (in post-task interviews and written comments) reported having difficulty deciding which descriptors applied to particular items when the descriptor was ambiguous or when there were multiple descriptors containing similar phrases. The large number of descriptors which were either not chosen very frequently or yielded a substantial number of evenly divided judgments reported in Tables 6-8 lend support to the concern about ambiguity. Table 11 contains several instances of similar wordings from

Table 11

Descriptors With Similar Phrases From Different Levels

Level	Descriptor ID #	Phrase
Grade 4		
B	D1a	(use basic facts to perform simple) computations with whole numbers
P	D1c	(use) whole numbers to compute results
B	D1b	estimate with whole numbers
P	D1d	(use) whole numbers to estimate
B	D2a	(show some) understanding of fractions and decimals
P	D2b	(have a conceptual) understanding of fractions and decimals
B	D3	understanding (the mathematical) concepts and procedures
P	D4	procedural and conceptual understanding (to problem solving)
B	D6a	solve (some simple) real-world problems
P	D6b	solve real-world problems
B	D3	understanding (the mathematical) concepts and procedures
A	D5	procedural and conceptual understanding (to complex and nonroutine real-world problem solving)
B	D6a	solve (some simple) real-world problems
A	D6c	solve (complex and nonroutine) real-world problems
P	D6b	solve real-world problems
A	D6c	solve (complex and nonroutine) real-world problems
P	D8b	explanations of how solutions were achieved
A	D8c	explaining (why, as well as) how, answers (and solution processes) were achieved
P	D4	procedural and conceptual understanding (to problem solving)
A	D5	procedural and conceptual understanding (to complex and nonroutine real-world problem solving)
B	D3	understanding (the mathematical) concepts and procedures
P	D4	procedural and conceptual understanding (to problem solving)
A	D5	procedural and conceptual understanding (to complex and nonroutine real-world problem solving)
B	D6a	solve (some simple) real-world problems
P	D6b	solve real-world problems
A	D6c	solve (complex and nonroutine) real-world problems

Table 11 (continued)

Level	Descriptor ID #	Phrase
Grade 8		
B	D1	understanding of arithmetic operations
P	D2	(thorough) understanding of (basic-level) arithmetic operations
B	D15	conceptual and procedural (understanding)
P	D16	(applying mathematical) concepts and procedures
P	D22e	(convey) underlying reasoning (skills)
A	D22d	(explain) the reasoning (process) underlying their conclusion
Grade 12		
B	D1a	(using) geometric reasoning (strategies)
P	D1c	(an understanding of) geometric reasoning
B	D2a	(using) algebraic reasoning
P	D2b	(an understanding of) algebraic reasoning
B	D9	reasonableness of results as applied to real-world (problems)
P	D10	reasonableness of answers as applied to real-world (situations)
B	D4a,b	(apply) statistical reasoning
P	D4c	(an understanding of) statistical reasoning
B	D11	procedural and conceptual knowledge
P	D12	mathematical concepts and procedures
B	D11	procedural and conceptual knowledge
A	D13	procedural and conceptual knowledge
B	D14b	use (correct) mathematical (language) and symbols to communicate mathematical reasoning (processes)
A	D14c	communicate their mathematical reasoning through (clear, concise, and correct) use of mathematical symbolism
P	D3a	understand (elements of) the function concept
A	D3b	understand the function concept
P	D12	mathematical concepts and procedures
A	D13	procedural and conceptual knowledge
B	D11	procedural and conceptual knowledge
P	D12	mathematical concepts and procedures
A	D13	procedural and conceptual knowledge

different levels. Descriptors that were least consistently mapped by judges were those that were not content-specific, contained terms such as "problem solving," "reasoning," "reasonableness of answers," "conceptual or procedural knowledge," and referred to "clear and concise" written responses.

Analysis of student performance on items mapped to descriptors and levels. The performance of students classified by achievement level was obtained for the sets of items assigned to single achievement levels. More specifically, the percentage of students who answered each item correctly (p-values) for students classified as Below Basic, Basic, Proficient, or Advanced were obtained. The median p-values across the set of items mapped to each descriptor (using the "at least four out of six judges" criterion for a mapping) are provided in Tables 12-14 for all descriptors to which at least 9 items were mapped.

As discussed earlier, different statistical criteria might be chosen to judge whether the performance of students on items mapped to descriptors at a given achievement level was consistent with the descriptors' classifications. In order for the pattern of student performance to be consistent with the mapping of items to a descriptor, we chose as one of our standards a variant of the NAEP anchor item criteria; namely, the median p-values on the subset of items to which the descriptor was mapped should be at least .65 for students classified at the achievement level from which the descriptor was abstracted, and the p-values should be less than .5 for students classified at the next lowest level.

Table 12

Median P-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 4

Descriptor ID number	Keywords		Level of descriptor	# of items mapped to descriptors	Level of students			
	Content	Process			Below basic	Basic	Proficient	Advanced
D1a	whole numbers	simple computation	B	10	.289	.472	.738	.933
D1b	whole numbers	estimating	B	14	.351	.468	.667	.898
D1c	whole numbers	computation	P	56	.264	.526	.792	.942
D1d	whole numbers	estimation	P	12	.287	.456	.706	.922
D1e	whole numbers	determination of reasonableness of results	P	10	.332	.496	.707	.900
D2b	fractions or decimals	conceptual understanding	P	13	.184	.275	.533	.898
D4	integrated procedural and conceptual knowledge	application and problem solving	P	87	.208	.418	.660	.898
D6b	real-world problems	problem solving	P	46	.235	.449	.682	.922
D7	strategies	problem solving	P	70	.212	.410	.675	.890
D8d		clear, concise communication	A	11	.067	.228	.534	.680

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

Table 13

Median P-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 8

Descriptor ID number	Keywords: Content	Keywords: Process	Level of descriptor	# of Items mapped to descriptor	Level of students: Below basic	Level of students: Basic	Level of students: Proficient	Level of students: Advanced
D1	arithmetic operations, whole numbers, decimals, fractions, or percent	understanding, estimation	B	78	.343	.633	.829	.940
D4	fundamental algebraic concepts	problem solving	B	15	.195	.373	.744	.956
D6a	informal geometric concepts	problem solving	B	48	.248	.512	.710	.906
D8	problems, diagrams, charts, graphs	problem solving	B	81	.271	.541	.764	.930
D9	strategies	problem solving (selecting and using strategies)	B	89	.256	.523	.763	.930
D10	technological tools (calculators, computers, and geometric shapes)	problem solving	B	42	.248	.406	.685	.831
D2	basic-level arithmetic operations	understanding, problem solving	P	108	.265	.542	.811	.946
D3	fractions, percents, decimals	understanding connections	P	10	.205	.335	.685	.868

Table 13 (continued)

Descriptor ID number	Keywords		Level of descriptor	# of Items mapped to descriptor	Level of students			
	Content	Process			Below basic	Basic	Proficient	Advanced
D6b	properties of informal geometry	application	P	49	.231	.482	.693	.904
D7	quantity or spatial relationships	problem solving or reasoning	P	77	.358	.623	.845	.944
D13a	data or graphs	making inferences	P	9	.381	.654	.804	.954
D14a	statistics or probability	calculating results	P	13	.179	.329	.721	.943
D22e	beyond arithmetic	reasoning	P	9	.004	.037	.160	.442
D6c	geometry	considering reasonableness of answers	A	16	.213	.287	.502	.740
D21	number sense	considering reasonableness of answers	A	50	.459	.717	.896	.962
D22d		explaining reasoning process	A	15	.038	.232	.481	.762

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

Table 14

Median P-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 12

Descriptor ID number	Keywords: Content	Keywords: Process	Level of descriptor	# of Items mapped to descriptor	Level of students: Below basic	Level of students: Basic	Level of students: Proficient	Level of students: Advanced
D1a	geometric reasoning strategies	problem solving	B	22	.264	.361	.701	.892
D1b	geometric relationships and corresponding measurement skills		B	46	.228	.354	.774	.940
D2a	algebraic reasoning strategies	problem solving	B	36	.193	.356	.818	.947
D4b	tables or graphs	applying statistical reasoning	B	9	.329	.359	.647	.916
D6	verbal, algebraic, tabular or graphical forms of presentation	recognizing relationships	B	19	.470	.679	.854	.949
D11	procedural knowledge or conceptual knowledge	problem solving	B	53	.340	.575	.868	.972
D1c	geometric reasoning	understanding	P	12	.311	.530	.778	.859
D1d	spatial reasoning	understanding	P	13	.325	.661	.864	.949

Table 14 (continued)

Descriptor ID number	Keywords		Level of descriptor	# of Items mapped to descriptor	Level of students			
	Content	Process			Below basic	Basic	Proficient	Advanced
D1e	geometric relationships	justifying	P	10	.437	.804	.898	.961
D2b	algebra	understanding, reasoning	P	17	.153	.289	.714	.929
D2c	algebraic operations involving polynomials	performing	P	9	.204	.498	.860	.971
D4d	data in tabular or graphical form	analyzing and interpreting	P	18	.484	.726	.850	.936
D14d		defending ideas	P	16	.048	.170	.504	.690

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

When these criteria are applied,[20] the 4th-grade results in Table 12 indicate that both Basic descriptors should be Proficient and one Proficient descriptor (D2b) should be Advanced. At grade 8 (Table 13), five of the six Basic descriptors should be Proficient, two of the six Proficient descriptors should be Basic, and one Proficient descriptor (D22e) does not even have a median p-value of at least .65 for students classified as Advanced! One of the three advanced descriptors (D21) should be Basic. The results at grade 12 (Table 14) are much the same. Five of the six Basic descriptors should be Proficient; three of the seven Proficient descriptors (D1d, D1e, and D4d) should be Basic, and there were no Advanced 12th-grade descriptors to which at least nine items were mapped.

Taken as a whole, the pattern of performance reflected in Tables 12-14 raises questions about the soundness of the mapping of descriptors to achievement levels. Of the 38 descriptors to which at least 9 items were mapped, less than half (17) exhibited a pattern of student performance that was consistent with the achievement level statements from which the descriptors were derived.

If one looks at the entire distributions of p-values across items for each descriptor, there are patterns that further call the achievement level descriptions into doubt. Illustrative distributions of the p-values across the items are displayed in Figures 10-13 from 4 descriptors at grade 8. These show that the items within each group (either descriptor or achievement level) vary substantially in terms of percent correct. The fact that the percent correct varies within any given descriptor is in itself neither surprising nor

[20] In some cases we had to relax the less than .5 criterion slightly in our interpretations. Otherwise certain descriptors could not have been classified according to levels.

Figure 10
P-Values for Groups of Students on Subset of 49 Items Mapped to Proficient Descriptor Number D6b, Grade 8

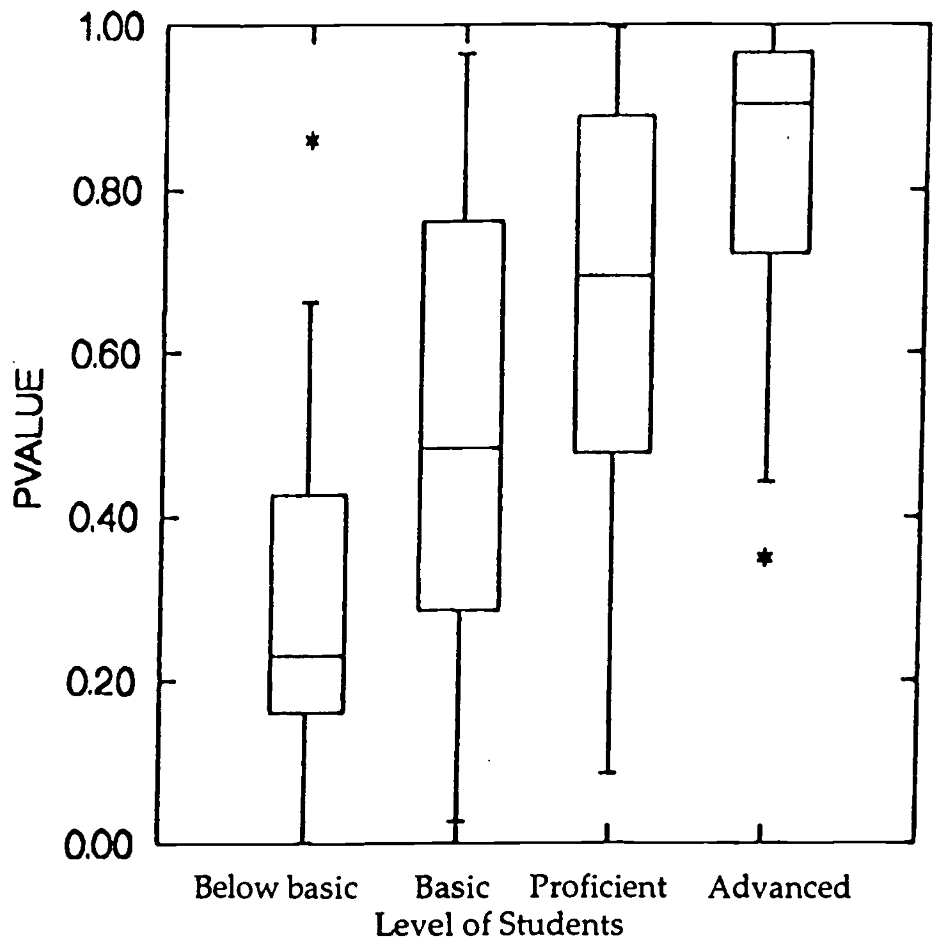


Figure 11
P-Values for Groups of Students on Subset of 78 Items Mapped to Basic Descriptor Number D1, Grade 8

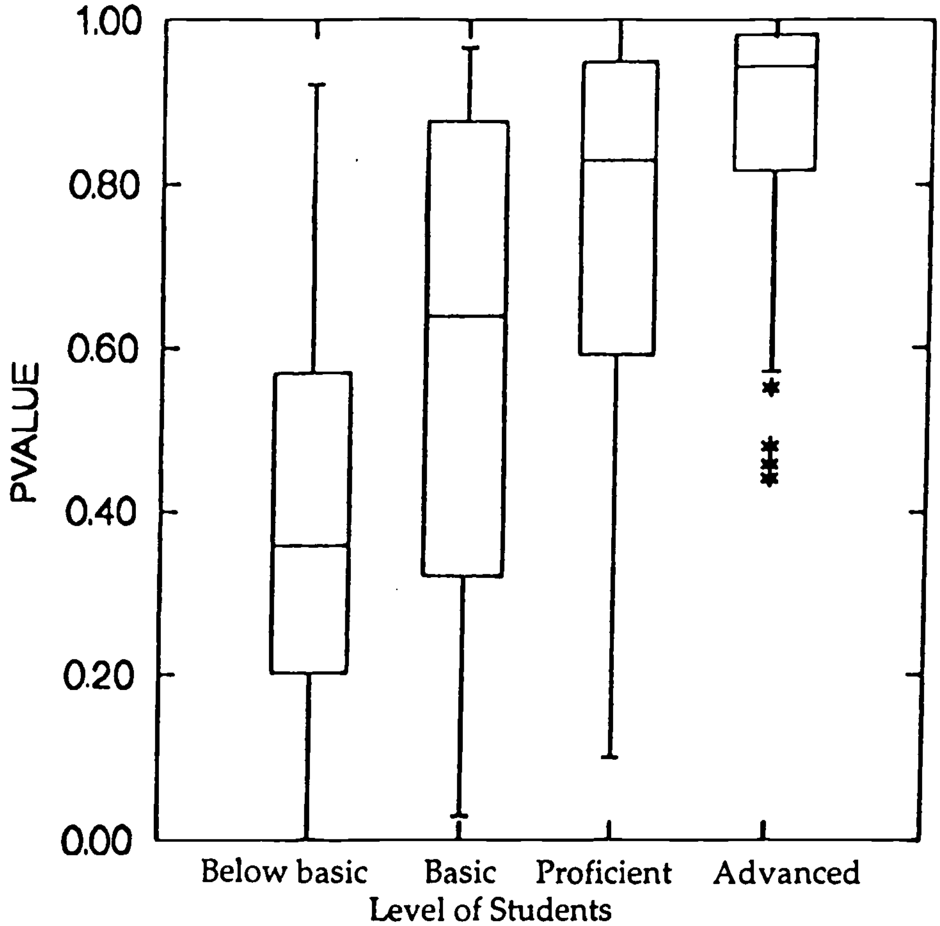


Figure 12
P-Values for Groups of Students on Subset of 15 Items Mapped to Basic Descriptor Number D4, Grade 8

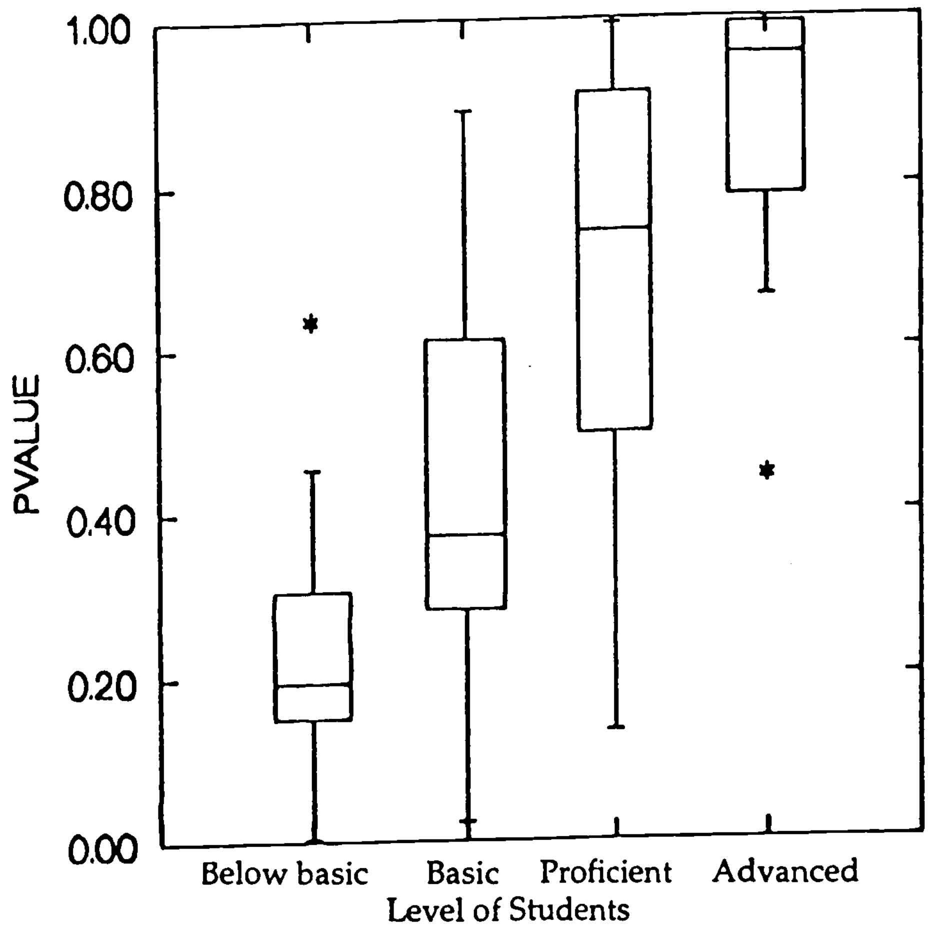
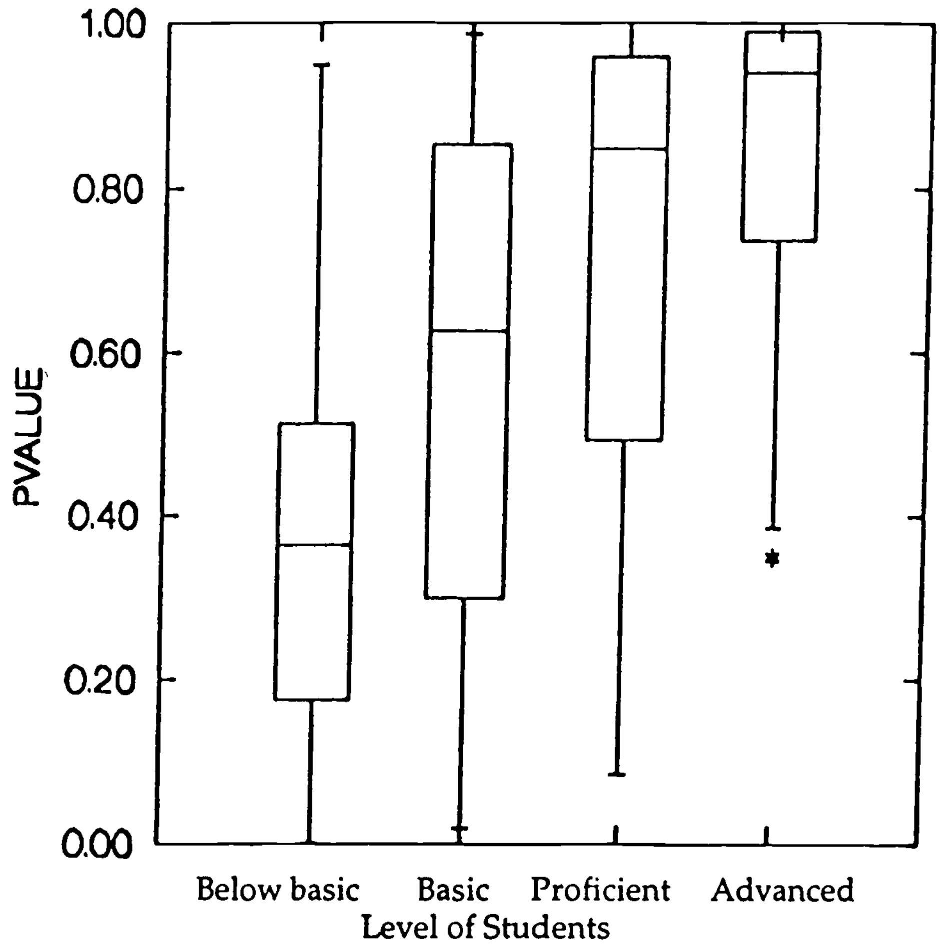


Figure 13
P-Values for Groups of Students on Subset of 17 Items Mapped to Proficient Descriptor Number D7, Grade 8



undesirable, but the specific patterns shown here are nonetheless troubling. First, the distributions differ markedly across descriptors within the same achievement level (e.g. D1 versus D4). Second, the distributions of p-values overlap considerably across the levels. Third and most important is the fact that some of the distributions are so low. The negative conclusions above are not dependent on the choice of .65 as the standard for percent correct. **In fact, in some cases, the distributions are so low that it is hard to choose any reasonable criterion by which one can say that students at the level exhibit the skills implied by the descriptor with any degree of accuracy.**

If the pool of NAEP assessment items adequately represents the domains associated with specific descriptors (which it may not), the plots also serve to highlight what may be either misassignment of descriptor statements to achievement levels or simply flawed descriptions of the skills purportedly associated with certain levels. For example, the descriptor D4, "using fundamental algebraic concepts in problem solving," was drawn from the Basic level description; yet more than 75% of the 15 items mapped to this descriptor had percent correct values less than the threshold of .65 (Figure 10). Conversely, the performance of the students scoring at the Basic level on items mapped to descriptor D7, "familiarity with quantity or spatial relationships in problem solving or reasoning," from the Proficient level description was distributed fairly evenly around .65 (Figure 13).

The lack of consistent separation among performances on the items mapped to descriptors from different levels can also be seen when the items are pooled across the descriptors for each level. Tables 15-17 contain the mean p-values for Below Basic, Basic, Proficient, and Advanced students on the sets of items assigned to single achievement levels. Performance on sets of items at

Table 15

Mean P-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 4

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	28	.362	.602	.792	.887
Basic	6	.356	.523	.715	.838
Proficient	131	.307	.500	.698	.874
Advanced	13	.136	.336	.560	.746

Table 16

Mean P-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 8

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	2	.601	.774	.895	.953
Basic	13	.508	.714	.830	.914
Proficient	121	.339	.563	.750	.883
Advanced	75	.379	.578	.734	.855

a particular level should be high for students classified at that level or higher, but lower for students classified at lower levels. This is so for all three grade levels. However, in some cases, the performance of students classified at one level on items that represent a higher level is similar to their performance on items that represent the level at which they (the students) are classified. For example, Figure 14 shows that at grade 4, students classified as Basic score

Table 17

Mean P-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 12

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	34	.358	.589	.780	.893
Basic	88	.387	.607	.805	.926
Proficient	73	.325	.519	.720	.854
Advanced	13	.137	.272	.549	.777

almost as well on Proficient items (p-value = .500) as they do on Basic items (p-value = .523); Figure 15 shows that 8th-grade students classified at the Proficient level score almost as well on Advanced items (p-value = .734) as they do on Proficient items (p-value = .750). Also, once again, these figures reveal considerable variability and overlap among the performances across the levels.

The analyses above of student performance on the items classified by the achievement levels to which their descriptors belong indicate that **the descriptions do not provide a clear indication of which items students at a given level are likely to be able to answer correctly**. Among students at a given level, performance on items linked to that level by judges varied and was in many cases lower than many people would consider reasonable. For example, in some instances, the median percent correct for students was less than 50% on items associated with that level. Low percent correct values were especially frequent for students in the Basic range. This variation in performance is greatest for items corresponding to Basic level descriptions.

Figure 14
Distribution of Item Percents Correct (P-Values) for Subsets of Items Not Classified, or Classified as Basic, Proficient, and Advanced Based on Judges' Mappings, Grade 4

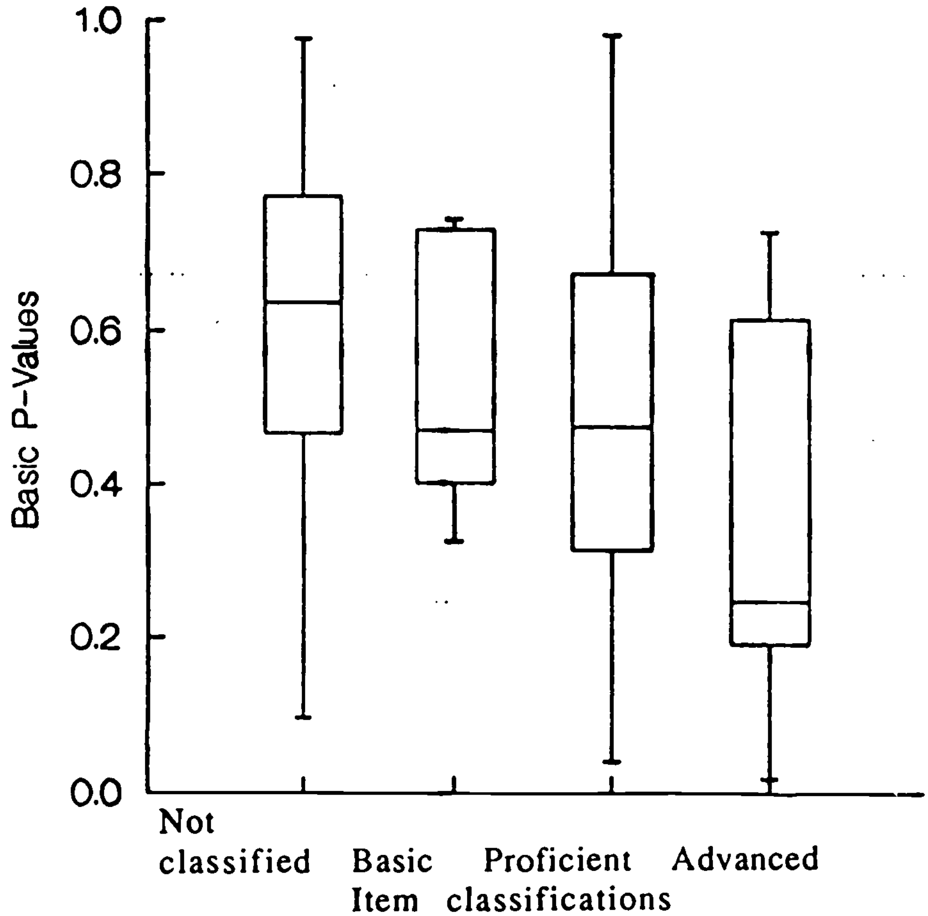
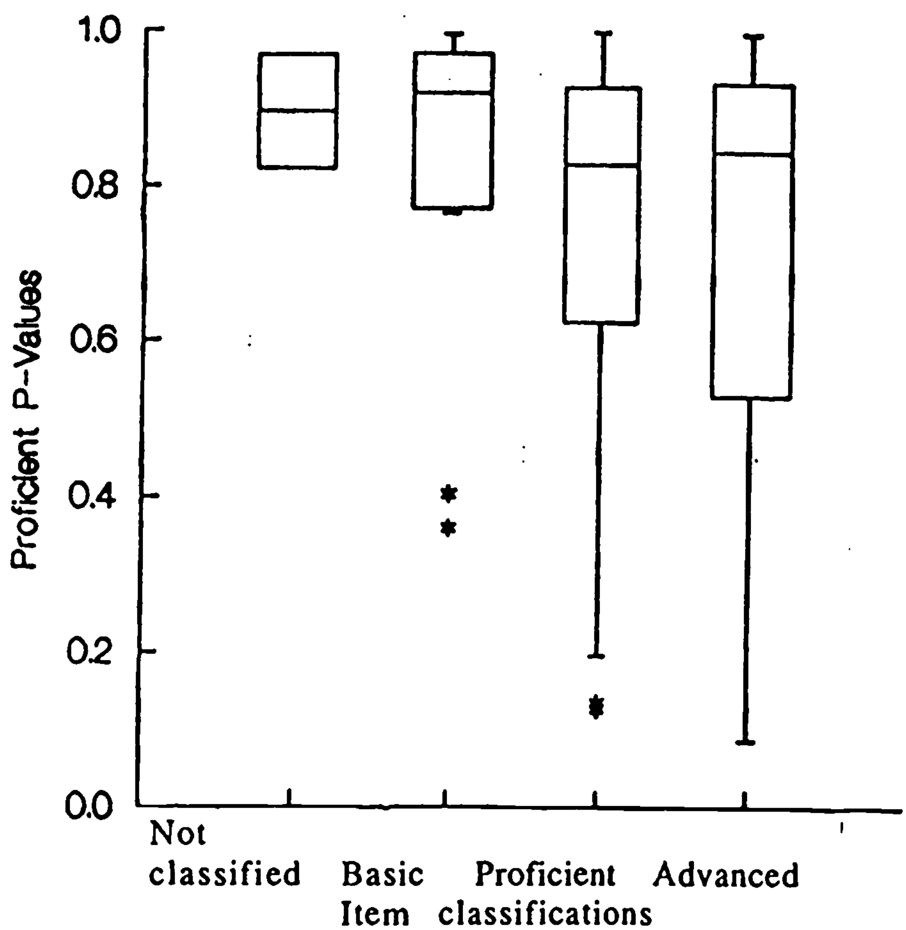


Figure 15
Distribution of Item Percents Correct (P-Values) for Subsets of Items Not Classified, or Classified as Basic, Proficient and Advanced Based on Judges' Mappings, Grade 8



Item Classification Based on Statistically Differentiating Student Performance

The analyses reported in this section examine the content characteristics of items that successfully differentiate among students at different achievement levels. The work proceeded in two stages. First, items that students at a given achievement level have a high probability of answering correctly while students at the next lower level had a substantially lower probability of answering correctly were identified. We termed the items that met the statistical criteria we used **differentiating items**. Second, the resulting sets of differentiating items at each level and grade were described using a variety of classification and coding schemes with the intent of characterizing the tasks which students at different achievement levels were able to perform with high probability.

Identifying differentiating items. As indicated in an earlier section, the starting point for defining what constituted a differentiating item involved satisfying three criteria (derived from the NAEP anchor item identification procedures) for the proportions of correct responses (p-values) on the items at the various achievement levels. To qualify as a differentiating item at a given level (Basic, Proficient, or Advanced), an item had to

1. Have a p-value for students at that level of at least 0.65.
2. Have a p-value for students at the next lower level of less than 0.50.
3. Have a difference of at least 0.30 between the p-value at the differentiating level and the p-value at the next lower level.

Any item that satisfied all three criteria for a given level was identified as a differentiating item at that level. Because we were concerned that this set of criteria might be viewed as too restrictive, the set of items that satisfied only the first two criteria (dropping the 30% difference between levels criterion) were also identified. The counts of items satisfying all three criteria and just two criteria broken down by grade and by the levels at which they differentiate are reported in Table 18.

The first point about these data is that there are considerably more items at each grade that do not differentiate among the levels than do differentiate. Using all three criteria, the proportion of differentiating items ranges from approximately 25% (42 of 178 Grade 4 items) to 32% (67 of 208 Grade 12 items). Second, relaxing the criteria increases the number of differentiating items by a considerable number (42 to 75 at grade 4; 53 to 78 at grade 8; and 67 to 84 at grade 12) and raises the proportion of differentiating items to roughly 40% at all three grades. Third, dropping the 30% difference criterion adds a large number of items that differentiate at the Basic level at all three grades. Also,

Table 18

Breakdown of Differentiating Items by Grade and Level Using All Three Statistical Criteria and the First Two Statistical Criteria

		Level				% of all items at grade
Grade	Criteria	Basic	Proficient	Advanced	Total	
4	All three	16	14	12	42	25%
	First two	25	32	18	75	42%
8	All three	25	17	11	53	25%
	First two	35	21	22	78	37%
12	All three	29	26	12	67	32%
	First two	38	31	15	84	40%

the number of Proficient level differentiating items increases substantially at grade 4 and the number of Advanced differentiating items increases noticeably at grade 8.

Clearly, the choice of criteria matters with regard to the size and nature of the pool of items that are defined to be differentiating. There is the obvious tradeoff between the sharper distinctions among levels with more restrictive criteria versus a larger pool of items to help characterize a given level. Despite the substantial increase in the pool of items that would be available for further study, we decided to highlight the analysis of those items that satisfied all three criteria because of their linkage with current NAEP anchor item criteria. Most of the analyses that follow were carried out both ways and any major differences in results associated with our choice will be noted.[21]

Correspondence between mapped achievement level descriptors and level at which items differentiate. Once items that do differentiate among the achievement levels are identified, the correspondence between the level at which differentiation occurred and the level at which judges mapped these items to NAGB descriptor statements can be examined. Tables 19-21 provide the data for these comparisons.

It is evident from these tables that the assignment of items to levels based on judges' mappings of descriptors to items is not consistent with the assignment of items based on differentiation of student performance. For example, only 11 of the 42 grade 4 differentiating items (all at the Proficient

[21] As part of our routine descriptive analyses, we generated the distributions of the differentiating items across achievement levels for each block of exercises at all three grades. There was substantial variability across blocks and grades in both the number of differentiating items per block and the distribution of items by the levels at which they differentiate in a given block. See Appendix F for a discussion of these results.

Table 19

Assignment of Differentiating Items to Levels Based on Judges Mappings, Grade 4

Level to which item was mapped	Basic differentiating items	Proficient differentiating items	Advanced differentiating items
Not classified	4	2	0
Basic	0	1	0
Proficient	11	11	12
Advanced	1	0	0
Total	16	14	12

Table 20

Assignment of Differentiating Items to Levels Based on Judges Mappings, Grade 8

Level to which item was mapped	Basic differentiating items	Proficient differentiating items	Advanced differentiating items
Not classified	0	0	0
Basic	2	0	1
Proficient	14	14	7
Advanced	9	3	3
Total	25	17	11

level) were mapped to descriptors drawn from the levels at which they differentiated. At grade 8, 19 out of 53 mapped consistently, also mainly at the Proficient level (14). The match was somewhat better at the Basic level in Grade 12 where 15 out of 29 mapped consistently, with an overall 24 out of 67 consistent mappings.

Table 21

Assignment of Differentiating Items to Levels Based on Judges Mappings, Grade 12

Level to which item was mapped	Basic differentiating items	Proficient differentiating items	Advanced differentiating items
Not classified	4	3	1
Basic	15	12	6
Proficient	9	8	4
Advanced	1	3	1
Total	29	26	12

Tables 22-24 provide further indication of just how problematic the assignment of the descriptor statements to levels may be when judged from the perspective of student performance. The majority of parsed descriptor statements from the NAGB content descriptions were not mapped to any differentiating items (12 of 18 at grade 4, 21 of 31 at grade 8, and 26 of 35 at grade 12). Only five descriptors overall (at grade 4, D1B "estimating with whole numbers"; at grade 12, D1d "understanding of spatial reasoning," D2b "understanding of algebraic reasoning," D2c "performing algebraic operations involving polynomials," and D4a "applying statistical reasoning in the organization and display of data") mapped solely to items differentiating at their NAGB designated level. Five other descriptors (at grade 4, D2b "using basic number facts to perform simple computations with whole numbers" and at grade 8, D6b "applying the properties of informal geometry," D7 "quantity and spatial relationships in problem solving or reasoning," D14a "Calculating results within the domain of statistics or probability," and D21 "Use number sense to consider reasonableness of an answer") failed to map to any items at their NAGB identified level but did map at other achievement levels.

Table 22

Mapping of Descriptors to Differentiating Items, Grade 4

Level of descriptor	Descriptor ID number	Number of differentiating items		
		Basic	Proficient	Advanced
B	D1A			
B	D1B	8		
P	D1C		6	3
P	D1D			
P	D1E			
B	D2A			
P	D2B			3
B	D3	16	14	12
P	D4	5	7	8
A	D5			
B	D6A			
P	D6B	5	7	7
A	D6C			
P	D7	8	7	9
P	D8A			
P	D8B			
A	D8C			
A	D8D			

Table 23

Mapping of Descriptors to Differentiating Items, Grade 8

Level of descriptor	Descriptor ID number	Number of differentiating items		
		Basic	Proficient	Advanced
B	D1	5	9	
P	D2	10	17	6
P	D3			
B	D4			
P	D5			
B	D6A	6		3
P	D6B	7		3
A	D6C			
P	D7	14		4
B	D8	11	4	5
B	D9	8	8	9
B	D10	4	4	3
A	D11			
B	D12			
P	D13A			
P	D13B			
P	D14A			3
P	D14B			
P	D14C			
B	D15			
P	D16			
A	D17			
P	D18			
P	D19			
A	D20			
A	D21	6	3	
P	D22A			
P	D22B			
P	D22C			
A	D22D			
P	D22E			

Table 24

Mapping of Descriptors to Differentiating Items, Grade 12

Level of descriptor	Descriptor ID number	Number of differentiating items		
		Basic	Proficient	Advanced
B	D1A	4	4	3
B	D1B	8	8	6
P	D1C			
P	D1D		3	
P	D1E			
B	D1F			
B	D2A	8	12	
P	D2B		5	
P	D2C		3	
B	D2D			
P	D3A			
A	D3B			
P	D3C			
A	D3D			
A	D3E			
B	D4A	5		
B	D4B			
P	D4C			
P	D4D			
B	D4E			
A	D5			
B	D6	3	3	
A	D7			
B	D8			
B	D9			
P	D10			
B	D11	6	9	3
P	D12			
A	D13			
B	D14A			
B	D14B			
A	D14C			
P	D14D			
P	D14E			
P	D14F			

Characterizing items by item signatures. The analyses presented thus far lend little empirical support for achievement level descriptions reported by NAGB. There was only limited and spotty correspondence between student performance on items that differentiate among the levels and the NAGB achievement levels from which the descriptor statements assigned to them by our judges were drawn. **If the descriptor mapping was inconsistent, the question then is whether it is possible to characterize the content differences among assessment items that differentiate among the levels in some other manner.**

Our attempts at characterizing the differentiating items were predicated on the judgments of their content by mathematics education experts. These experts were asked to identify the "item signature" for each of the differentiating items. The item signature concept was developed during the Survey of Mathematics and Science Opportunities Study (SMSO, 1993) conducted in connection with the Third International Mathematics and Science Study (TIMSS), as a means of applying the TIMSS multi-aspect, multi-dimensional curriculum framework to the characterization of the full array of content assessed by individual test items.

In the present study, item signatures were determined by having math experts code all the content measured by each item for the full set of differentiating items.[22] The content categories were generated by the experts working together in groups of three to five, one grade level at a time. The experts examined each item from the set that satisfied all three statistical criteria, and listed the relevant content attributes of that item and

[22] A more detailed description of the coding of differentiating items according to their item signatures is contained in Appendix G.

accumulated the list of attributes and item assignments to attributes. All decisions about attributes were made by consensus.

Our analyses of these data were carried out on the full set of content attributes identified by the experts.[23] Tables 25-27 report the item counts and percent of total number of content codes at each achievement level for each of the content attributes at each grade level. There were 45 different attributes in all but some attributes did not occur at all grades. The attributes are grouped in these tables roughly according to content similarity.

To facilitate closer examination of the items that satisfy the differentiating criteria at each level, all 1992 NAEP items from public release blocks[24] which differentiated student performance are displayed in Appendix I. For each differentiating item, we have also included its NAEP ID, block and item number; the counts of the number of judges (out of 6) who mapped the item to each descriptor derived from the NAGB content descriptions; and the p-values overall and for those students scoring at the Below Basic (PPLUS1), Basic (PPLUS2), Proficient (PPLUS3), and Advanced (PPLUS4) levels.

---

23 In an effort to provide a more parsimonious characterization, the descriptors were also assigned to appropriate categories in the content aspect of the TIMSS curriculum framework in mathematics. The results of this classification are discussed in Appendix H. A complete description and discussion of the study that examined the content and linguistic characteristics of the "differentiating" items can be found in Novak, Burstein, and Larriva (forthcoming).

24 Limiting examples of differentiating items to those available from NAEP public release blocks makes it difficult to illustrate some of the differences in the item characteristics across the levels. Moreover, the restriction may result in a misleading impression of the actual pool of differentiating items and probably affected the selection of exemplar items. The concern about the consequences of current NAEP item release guidelines is well-known but it warrants more attention than it currently receives.

Table 25

Mapping of Differentiating Items to Content Categories (Item Signatures), Grade 4

Category Frequency Col Pct	Level			Total
	Basic	Proficient	Advanced	
arithmetic operations	7 15.91	9 18.75	6 15.79	22
decimals	1 2.27	1 2.08	0 0.00	2
fractions	0 0.00	0 0.00	4 10.53	4
money	2 4.55	0 0.00	2 5.26	4
number sense	0 0.00	2 4.17	0 0.00	2
place value	1 2.27	0 0.00	0 0.00	1
estimation	1 2.27	3 6.25	0 0.00	4
measurement	2 4.55	0 0.00	2 5.26	4
metric units	1 2.27	1 2.08	0 0.00	2
use of rulers / tools	2 4.55	1 2.08	0 0.00	3
geometry	1 2.27	1 2.08	1 2.63	3
proportional reasoning	0 0.00	1 2.08	0 0.00	1
number sentences	1 2.27	1 2.08	0 0.00	2
pattern recognition	0 0.00	2 4.17	0 0.00	2
probability	0 0.00	0 0.00	2 5.26	2
tables / graphs / charts	4 9.09	2 4.17	0 0.00	6

Table 25 (continued)

Category Frequency Col Pct	Level Basic	 Proficient	 Advanced	Total
explain reasoning	0 0.00	0 0.00	1 2.63	1
logical reasoning	1 2.27	0 0.00	0 0.00	1
real-world problems	5 11.36	8 16.67	6 15.79	19
story problems	2 4.55	3 6.25	2 5.26	7
alternative symbol systems	2 4.55	1 2.08	0 0.00	3
diagram	4 9.09	3 6.25	3 7.89	10
calculator	1 2.27	0 0.00	0 0.00	1
complex problem solving	2 4.55	3 6.25	2 5.26	7
multi-step problems	2 4.55	4 8.33	4 10.53	10
recall of definition	0 0.00	1 2.08	2 5.26	3
spatial reasoning	2 4.55	1 2.08	1 2.63	4
Total	44	48	38	130

Table 26

Mapping of Differentiating Items to Content Categories (Item Signatures), Grade 8

Category Frequency Col Pct	Level			
	Basic	Proficient	Advanced	Total
algebra of integ	0 0.00	2 4.26	0 0.00	2
arithmetic operations	3 4.48	4 8.51	0 0.00	7
conversions	0 0.00	1 2.13	0 0.00	1
conversions / %, decimals, fractions	0 0.00	0 0.00	1 2.08	1
decimals	1 1.49	1 2.13	2 4.17	4
fractions	5 7.46	4 8.51	0 0.00	9
number line	1 1.49	1 2.13	0 0.00	2
number sense	3 4.48	7 14.89	1 2.08	11
percentage	0 0.00	3 6.38	0 0.00	3
place value	1 1.49	0 0.00	1 2.08	2
square root	0 0.00	1 2.13	0 0.00	1
estimation	0 0.00	1 2.13	2 4.17	3
measurement	4 5.97	0 0.00	3 6.25	7
metric units	0 0.00	0 0.00	2 4.17	2
use of rulers / tools	1 1.49	0 0.00	0 0.00	1

Table 26 (continued)

Category Frequency Col Pct	Level			Total
	Basic	Proficient	Advanced	
geometric properties	0 0.00	0 0.00	1 2.08	1
geometry	6 8.96	1 2.13	3 6.25	10
proportional reasoning	3 4.48	1 2.13	0 0.00	4
algebraic operations	0 0.00	1 2.13	0 0.00	1
algebraic reasoning	1 1.49	1 2.13	1 2.08	3
pattern recognition	2 2.99	0 0.00	2 4.17	4
substitution	0 0.00	1 2.13	0 0.00	1
probability	2 2.99	2 4.26	2 4.17	6
statistics	0 0.00	0 0.00	1 2.08	1
tables / graphs / charts	6 8.96	3 6.38	3 6.25	12
explain reasoning	0 0.00	0 0.00	1 2.08	1
logical organization	0 0.00	0 0.00	1 2.08	1
logical reasoning	1 1.49	0 0.00	1 2.08	2
real-world problems	5 7.46	0 0.00	0 0.00	5
story problems	4 5.97	6 12.77	1 2.08	11
written response	1 1.49	0 0.00	0 0.00	1

Table 26 (continued)

Category Frequency Col Pct	Level			Total
	Basic	Proficient	Advanced	
alternative symbol systems	1 1.49	1 2.13	0 0.00	2
diagram	7 10.45	0 0.00	4 8.33	11
calculator	1 1.49	2 4.26	1 2.08	4
complex problem solving	0 0.00	0 0.00	2 4.17	2
multi-step	1 1.49	0 0.00	2 4.17	3
recall of definition	6 8.96	2 4.26	5 10.42	13
recall of rule / formula / property	0 0.00	0 0.00	1 2.08	1
spatial reasoning	0 0.00	1 2.13	4 8.33	5
visualization	1 1.49	0 0.00	0 0.00	1
Total	67	47	48	162

Table 27

Mapping of Differentiating Items to Content Categories (Item Signatures), Grade 12

Category / Frequency Col pct	Level			
	Basic	Proficient	Advanced	Total
algebra of integers	2 2.17	3 3.03	1 1.22	6
arithmetic operations	6 6.52	5 5.05	1 1.22	12
conversions / %, decimal, fraction	2 2.17	1 1.01	0 0.00	3
fractions	2 2.17	1 1.01	2 2.44	5
number line	2 2.17	0 0.00	0 0.00	2
number sense	6 6.52	3 3.03	5 6.10	14
estimation	4 4.35	1 1.01	3 3.66	8
coordinate geometry	2 2.17	6 6.06	1 1.22	9
geometric properties	3 3.26	6 6.06	7 8.54	16
geometry	8 8.70	8 8.08	8 9.76	24
proportional reasoning	1 1.09	1 1.01	3 3.66	5

Table 27 (continued)

Category / Frequency Col pct	Level			
	Basic	Proficient	Advanced	Total
algebraic operations	2 2.17	6 6.06	4 4.88	12
algebraic reasoning	2 2.17	8 8.08	4 4.88	14
pattern recognition	2 2.17	2 2.02	0 0.00	4
substitution	5 5.43	5 5.05	2 2.44	12
permutations/ combinations	1 1.09	0 0.00	0 0.00	1
probability	1 1.09	0 0.00	0 0.00	1
statistics	1 1.09	0 0.00	1 1.22	2
tables / graphs / charts	3 3.26	1 1.01	3 3.66	7
explain reasoning	1 1.09	0 0.00	0 0.00	1
real-world problems	6 6.52	2 2.02	4 4.88	12
story problems	3 3.26	0 0.00	5 6.10	8
diagram	8 8.70	9 9.09	7 8.54	24
calculator	3 3.26	5 5.05	3 3.66	11

Table 27 (continued)

Category / Frequency / Col pct	Level			
	Basic	Proficient	Advanced	Total
multi-step	3 3.26	7 7.07	8 9.76	18
recall of definition	6 6.52	6 6.06	2 2.44	14
recall of rule / formula / property	5 5.43	9 9.09	7 8.54	21
spatial reasoning	2 2.17	4 4.04	1 1.22	7
Total	92	99	82	273

Generally, there is considerable scatter of items across content attributes with limited distinguishing clustering. At grade 4, being able to read tables, charts, and graphs differentiates Basic from Below Basic students but has limited effect at other levels. Estimation items differentiate Proficient from Basic students as do multi-step problems (some multi-step problems also differentiate Advanced from Proficient students). As might be expected, fraction items, largely new material at this grade, differentiate the Advanced students from all others but do not have similar impact at lower achievement levels. The prevalence of arithmetic operations across the levels is understandable and expected given their centrality to subject matter at this grade. Also prominent across achievement levels are word problems, whether short, artificial ones (which we term "story problems") or more realistically grounded, perhaps longer ones (the "real world problems" attribute). In addition, items involving diagrams of various kinds (these may be non-verbal representations of the problem situation or alternatively serve as concrete aids to framing the questions being asked) differentiated at all three levels. The topics prevalent at all levels suggest that to some degree, verbal fluency and facility with visual representations are pertinent features of the achievement levels at grade 4.

In some respects the patterns at grade 8 are less clear. At the Basic level, differentiating items are drawn mainly from advanced topics in arithmetic operations with whole numbers (e.g., embedded in simple story problems and tables/charts/graphs), operations with fractions, application of measurement formulas and low level geometry topics, the latter with accompanying diagrams. The Basic level differentiating items from the public release blocks (see Appendix I) illustrate the relatively straightforward nature of the word problems and the measurement and geometry applications that Basic students can do with high probability but Below Basic students were less likely to answer correctly. Indeed, there are 4 Basic level public release items dealing with applications involving rectangles (three on drawing rectangles which meet certain conditions and one on finding the length of a specific side)! Proficient level items involved more complex story problems and tables/graphs/charts (again see Appendix I for examples). Items on percentage were more prevalent than at the Basic level and content on measurement, geometry, real world problems, and recall of definitions were less prevalent. Virtually none of the arithmetic operations items of any kind differentiated at the Advanced level. More advanced measurement, geometry, tables/charts/graphs, and recall of definition items surfaced at the Advanced level. In addition, items involving spatial reasoning or diagrams differentiated at the Advanced level.

At grade 12, topics normally covered in first-year algebra and geometry at the high school level appear among the differentiating items. Topics in geometry, items with diagrams (connected with geometry here), recall of rules, formulas or properties and multi-step problems were prevalent at all three levels. Among the Basic differentiating items are a number that measure geometry and geometric properties, algebraic substitution, recall of formulae, rules, and properties, real world problems, and story problems that employ lower level geometry and algebra topics (see Appendix I for illustrations). The sophistication of the algebra and geometry content increases for the Proficient differentiating items. Coordinate geometry, algebraic operations, algebraic reasoning, recall of formulae and properties, multistep problems, and spatial reasoning are all prevalent. At this level, the mathematics is less likely to be placed in a real world or story problem context than at either the Basic or Advanced levels. The Advanced level differentiating items require yet even more sophisticated command of algebra and geometry (e.g., systems of equations, quadratic equations, and volumes of cylinder problems) than the Proficient level, as well as the ability to work multi-step word problems embedded in often real world contexts.

To summarize the content characterization of the achievement levels based on the coding by our math experts, the prevalent descriptors drawn from the content attributes coded in Tables 25-27 are presented in Table 28 in a manner that highlights contrasts across levels and grades. Brief cell labels were included that capture our overall impression of the content of the items that differentiated at each level.[25]

[25] These labels were inferred from a close examination of the individual items to which both content attributes and descriptors were mapped. The items contained in Appendix I serve as one source as do the non public release items and the additional pool of items that satisfied the two statistical criteria described earlier.

Table 28. Characteristics of differentiating items based on item signatures

Grade	Achievement Level: Basic	Proficient	Advanced
4	arithmetic operations / whole numbers; a few real-world problems; tables / charts / graphs; diagrams "Concrete"	arithmetic operations; estimation; real-world problems non-routine / multi-step diagrams "Non-routine whole number problems"	conceptual understanding of fractions / decimals; arithmetic operations probability; diagrams; real-world problems / more multi-step "Beyond whole numbers"
8	arithmetic operations in story problems; fractions; / number sense; measurement/geometry simple tables /charts /graphs; recall definitions diagrams "General math"	arithmetic operations with fractions/ percents; number sense; tables/charts/graphs story problems "Pre-algebra"	probability / statistics; tables/charts/graphs; spatial reasoning; geometry/ measurement; diagrams; recall of definitions; logical reasoning "Enriched"
12	arithmetic operations rational numbers; number sense; estimation; simple algebra (substitution) real-world and story problems; calculators; multi-step problems recall definitions, rules, formula; geometry properties and recall; diagrams; probability / statistics; tables/charts/graphs; "Partial mastery of algebra/geometry"	algebraic reasoning and operations; geometry (coordinate geometry, geometric properties); calculators; multi-step problems; recall of definitions, rules, formula; spatial reasoning diagrams "Algebra/geometry"	Number sense; estimation; proportional reasoning; multi-step geometry and algebra problems; recall of rule/formula/property; diagrams; "High end eclectic"

While the picture is far from perfect, what appears to come through in Table 28's portrayal of the empirical evidence is that **a primary basis for differentiating the performance of students across levels appears to be the extensiveness and quality of curriculum mastery or exposure.**[26] The items at the different achievement levels tend to represent content covered at different levels of the mathematics curriculum. As one moves across the levels, the differentiating items call upon a wider repertoire of content, and students are asked to apply this content in a wider array of circumstances. These circumstances seem to entail short word problems, either artificial or grounded in realistic situations, problems with visual representations (diagrams), or involve multiple steps where command of rules, formulae, and recall of definitions are essential components.

The curriculum exposure characterization is suggested by several patterns within grade levels. First, the appearance of fractions and decimals as content for the Advanced level at grade 4 could well reflect a faster curricular pace in classrooms and schools where these topics are taught. Grades 4 and 5 are the first major transition to operations with fractions and decimals as opposed to whole number arithmetic.

Second, the topics at the various levels at grade 8 appear to mirror several of the curriculum tracks offered to students at this grade level. The content associated with the Basic level appears to be the final onslaught at mastering arithmetic operations with whole numbers, fractions, and decimals, basic measurement formulas and simple geometry and table

---

[26] Strictly speaking, the results provide direct evidence associated with curriculum mastery. However, the parallels to curriculum distinctions associated with different courses and tracks strongly suggest that differences in opportunities to learn the test content (curriculum exposure) contribute to the pattern in the results. This interpretation is necessarily speculative rather than definitive, however.

reading. The Proficient students have moved on to pre-algebra material with ratio, proportion, and percent problems and operations with fractions fully mastered and more applications in the context of tables, charts, and graphs and in simple story problems. The prevalent Advanced content suggests a more enriched curriculum exposure with items measuring spatial reasoning, logical reasoning, and probability and statistics entering the picture.

Finally, if anything, the curriculum exposure and curriculum mastery patterns are even more distinctive at grade 12. Performance at the Basic level suggests exposure to introductory topics that might be found in the first year of high school algebra and high school geometry, but mastery is spotty at best. The differentiating items at the Proficient level cover most of the contents of these two courses and suggest that the student is prepared to move on to more sophisticated and complicated material as multi-step problems and graphical representations become more prevalent. The Advanced students at grade 12 appear to be able to handle just about anything that the NAEP item pool allows them to tackle. The mixture of multi-stepness combined with the recall of rules, formulas, and properties in the domains of algebra and geometry is a strong indication that these students have studied, and may have mastered, the traditional college preparatory mathematics curriculum by grade 12.

Characterizing items by linguistic features. In addition to content considerations, the possibility that linguistic features of the NAEP mathematics items might contribute to their success in differentiating among the achievement levels was explored. While this issue has been raised by others before (Spanos et al., 1988), here the decision to pursue possible linguistic feature influences was predicated on some concern that students' difficulties in solving certain types of problems may have more to do with their understanding of the questions than with their knowledge of the content being tapped. The prevalence of word problems of a variety of types over a range of content as differentiating items also suggested a perhaps subtle verbal component of task complexity contributing to performance patterns.

A coding instrument was developed for examining the linguistic features of NAEP mathematics test items (see Appendix D). Seven of the nine linguistic features constituting the coding instrument, represent (except for minor interpretive differences) categories developed by Spanos, Rhodes, Dale and Crandall of the Center for Applied Linguistics (1988). The other two features (quantitative attributes (4a) and quantities expressed in written text (4b)) attempt to measure apparent recurring linguistic characteristics that preliminary review of the test items suggested.[27]

Features 1a (comparative) and 1b (logical connectors) represent standard syntactic characteristics. Features 2a (mathematical vocabulary), 2b (natural language vocabulary), 2c (complex strings) and 2d (words which signal operations) are semantic attributes that reflect either general or math-specific understanding. Feature 3 (concepts requiring experience or knowledge) captures a pragmatic dimension.

A group of five former mathematics teachers jointly coded the items by reaching consensus on each. The actual data involve counts of instances of each feature for each differentiating item. Tables 29 and 30 present, respectively, the mean number of instances of occurrence at each level and

[27] Descriptor 5, multi-stepness, was included in an attempt to quantify complexity but was deleted during the coding session since the teachers had difficulty agreeing on the number of steps involved in the test items. Although multi-stepness really is more a cognitive than a strictly linguistic category, it was attempted in lieu of a better measure of linguistic complexity. Also, in a preliminary review of test items, the number of words in each test item was designated as a descriptor, again as a way of measuring linguistic complexity, but produced no meaningful results and was dropped.

Table 29

Mean Number of Instances of Linguistic Features in Sets of Items at Each Level in Grades 4, 8 and 12

		Descriptors									
Grade	Level	1a	1b	2a	2b	2c	2d	3	4a	4b	# of items
4	Basic	0.75	0.44	0.63	0.69	0.50	0.19	0.44	1.63	0.38	16
4	Proficient	0.57	0.71	0.36	0.64	0.86	0.21	0.21	1.36	0.36	14
4	Advanced	0.25	0.50	0.67	0.50	0.50	0.08	0.16	1.33	0.08	12
8	Basic	0.44	0.68	0.60	1.20	0.44	0.32	0.48	0.97	0.48	25
8	Proficient	0.76	0.53	0.71	0.47	0.47	0.35	0.24	0.71	0.35	17
8	Advanced	1.55	0.64	1.45	2.00	0.82	0.09	0.36	1.27	1.09	11
12	Basic	0.59	0.76	1.17	1.31	0.38	0.21	0.24	0.72	0.31	29
12	Proficient	0.42	0.73	1.38	2.42	0.58	0.15	0.08	0.31	0.50	26
12	Advanced	0.41	0.50	1.83	1.75	1.25	0.08	0.25	1.58	0.08	12

Table 30

Percentage of Items Containing at Least One Instance of Each Linguistic Feature at Each Achievement Level in Grades 4, 8 and 12

		Descriptors									
Grade	Level	1a	1b	2a	2b	2c	2d	3	4a	4b	# of items
4	Basic	38	38	44	50	31	19	44	**75**	25	16
4	Proficient	50	50	29	43	**71**	14	21	**71**	36	14
4	Advanced	25	42	42	42	33	08	17	**83**	08	12
8	Basic	24	60	48	64	36	24	40	48	24	25
8	Proficient	41	41	59	35	41	29	12	53	12	17
8	Advanced	**73**	55	**82**	**82**	55	09	36	64	55	11
12	Basic	45	52	62	**76**	31	17	17	45	21	29
12	Proficient	31	58	**77**	**85**	54	15	08	19	35	26
12	Advanced	33	42	**83**	**83**	**92**	08	25	58	08	12

grade and the percentage of items from a given level containing at least one instance of the feature. The results were examined for evidence that might indicate trends across performance levels.

Given the exploratory nature of this analysis, we concentrate on what appear to be significant differences across the levels in the prevalence of linguistic features and on whether at least two-thirds of the items at a given level contained a specific linguistic feature. Using these criteria, there are no features that are highly prevalent or differentiate among the levels at all three grades. In fact several features (logical connectors (1b); words which signal operations (2d); concepts requiring experience or knowledge (3); quantities expressed in written text (4b)) were not prevalent across items at any level or grade. Moreover, at grade 4 there were no significant differences in the mean frequencies on any of the linguistic features.

The main evidence of possible impact of linguistic features appears in the semantic characteristics, primarily natural language vocabulary with special mathematics use (2b) and mathematical vocabulary (2a). At grades 8 and 12, everyday words that have a different or specialized meaning in mathematics (2b) were prevalent and differentiated across the levels. Over 80% of the items that differentiate at the Advanced level at grades 8 and 12 included at least one instance of this feature. The prevalence of mathematics specific vocabulary (2a) differed significantly across levels at grade 8 with over 80% of the Advanced level items containing such words while at grade 12, most items differentiating at both Proficient and Advanced levels included such terms. In fact virtually all of the Advanced level items at grade 12 included specific mathematics vocabulary and special mathematics uses of natural vocabulary embedded in complex strings of words (feature 2c). The prevalence of expressions denoting quantitative attributes (4a; e.g., weight, mph) also differed significantly at grade 12 with the Advanced items containing them more often.

It is important to reiterate that our attempt at examining linguistic features of differentiating test items is preliminary and the results are at best suggestive.[28] Nevertheless, based on the evidence in hand, there are indications that at least those linguistic features of test items associated with either specialized mathematics terminology or special mathematical meanings attributed to standard vocabulary may affect whether an item differentiates performance at a specific achievement level. These differentiating features of items might be attributed to the level of curricular exposure (e.g., had the student had the coursework where the specific terminology or use was introduced or emphasized). Alternatively, it could be that students less facile with standard English usage experienced additional difficulties in coping with the specialized meanings of otherwise common vocabulary or with the sheer volume of new mathematics terminology they are expected to learn in their mathematics coursework. With the current data we

---

[28] The linguistic features coding was applied only to those items meeting the differentiation criteria; the prevalence of various features among the remaining items is unknown. Moreover, the linguistic coding process was difficult. It was sometimes unclear which descriptor best applied to a particular word or whether words or phrases which fit particular descriptor definitions really merited inclusion because of their relatively uncomplicated nature. And, at times, the initial descriptor definitions were too restrictive and boundaries had to be loosely interpreted to account for test item linguistic features which were deemed linguistically significant.

These problems might be partially corrected if the descriptor definitions were fine tuned using the knowledge gained from this coding session and if other descriptors are added. For example, it became evident during the coding session that other descriptors that can measure overall sentence structural or syntactic complexity were necessary (i.e., beyond the two syntactic categories already included, 1a and 1b). Counting clauses may be a difficult and tedious yet viable option. Also a descriptor which measures the prevalence of idiomatic phrases such as "run out of" or "how long does it take?" should also be considered.

are unable to decipher further which if any of these explanations can account for our results.

Taking these analyses as a set, we are prepared to say that **the characteristics of items that differentiate statistically among achievement levels are not accurately reflected by the current achievement level descriptions**. Judging from this empirical evidence, the primary bases for differentiating the performance of students across levels appears to be the extensiveness and quality of curriculum mastery (implying exposure) and potentially associated degrees of language facility as well as proficiency in responding to open-ended items.[29]

Conclusions and Recommendations

The following five major conclusions are based on the results of the three analytical approaches just described.

1. **Judged in terms of actual student performance, many of the items selected as exemplars of the achievement levels are misleading.** In some instances, less than half the students performing within the range of a given achievement level correctly answered an exemplar item for that level. In other cases, more than 75% of the students performing at a given level correctly answered an item intended to be an exemplar for the next higher achievement level. Presenting such items as exemplars of a given level provides a

[29] In exploratory multivariate analyses of the characteristics of the differentiating test items that either discriminate among the levels or that account for differences in item p-values between adjacent levels, we included a measure of whether an item was open-ended or not as one of the possible explanatory variables. (Other factors considered were characteristics of the content measured by the test items (either the descriptors to which they were mapped by the judges or the TIMSS classifications assigned to the item) and linguistic features.) In several of the analyses at all three grades, open-ended items significantly contributed to the differences in p-values across achievement levels.

misleading impression of what students performing at a given level are actually able to do.

2. **The 1992 NAEP mathematics assessment did not measure some of the attributes included in the descriptions of the achievement levels and measured some other attributes only poorly.** That is, the 1992 item pool provided sparse coverage of some attributes and no coverage of others. This sparse coverage is especially problematic for the grade 4 basic and advanced levels and for the grade 12 advanced level. Thus, it is impossible to say with any confidence whether students scoring at the level in question can do what those aspects of the descriptions describe.

3. **Frequently, many—in some cases, a majority—of the students at a given level did not successfully answer items linked to certain aspects of the descriptions at that level.** Among students whose performance reached a given level, performance on items linked to that level (by the second of the approaches noted above) varied and was in many cases lower than many people would consider reasonable. For example, in some instances, the median percentage of students answering correctly was less than 50% on items associated with that level. Low percent correct values were especially frequent for items in the Basic range. This variation in performance is greatest for items corresponding to Basic level descriptions.

4. **The definitions of the levels overlap considerably and frequently differ only in terms of subtle nuances**. Consequently, the association of items with a given level was often found to be ambiguous. Experienced mathematics educators were generally unable to make such distinctions reliably without specific and detailed training. Thus, it is unlikely that general populations of mathematics specialists, professional educators or the lay public could be any

more successful at interpreting correctly the intended differences among levels.

5. **The characteristics of items that differentiate among achievement levels suggest descriptions of performance that differ substantially from the current achievement level descriptions.** Differentiating items were identified on the basis of statistical properties (i.e., high probability of correct response for students at that level and a relatively low probability of correct response for students scoring below that level), and judges ascertained the attributes of these items. Judging from this empirical evidence, the primary bases for differentiating the performance of students across levels appear to be the extensiveness and quality of curriculum mastery or exposure and potentially associated degrees of language facility.

In sum, then, our analyses do not support the validity of the published content descriptions as characterizations of what students within specified score ranges can do. Some of the attributes of the descriptions could not be mapped to the NAEP items; those that could be mapped to NAEP did not consistently show performance patterns that would support the validity of the descriptions; and the exemplars as a set do not accurately characterize the performance of groups in question.

To a certain extent, our findings are limited to shortcomings of the content descriptions and associated exemplar items at the time of the conduct of our study. It is conceivable, although by no means sure, that if either earlier versions of the content descriptions written in more specific content terms or new versions written with less ambiguity had been studied, our findings might have been more favorable. Likewise, within limits, it is possible to choose exemplar items that better illustrate the knowledge, skills, and understandings that students performing at a given level have achieved. To do so it would have been necessary to select items that, in addition to meeting the content criteria employed, also have a reasonably high probability of a correct response (e.g., .65) for students scoring at that level and a substantially lower probability of correct response for students scoring below that level. Unfortunately, the released item pool does not contain a sufficient number of items that satisfy both content and statistical criteria to adequately illustrate student performance associated with the levels.

The empirical evidence from our examination of items that successfully differentiated among the achievement levels suggest that the current pool of items is not particularly well suited to distinguish performance at some levels (especially at the Advanced level). Moreover, other than suggesting general curriculum exposure and accomplishment advantages as one moves up the levels, there is little solid evidence to go on in characterizing and describing what students were able to demonstrate. In fact, our data on the possible influence of linguistic features of the test items and of item format on student performance warrant more careful examination. Under the conditions present at the time of our study, we were able to analyze the performance evidence only for the entire NAEP population at each achievement level; achievement level data for specific subpopulations of students (defined, e.g., by social background, gender, or language background) were not available for analysis. Nor is our curriculum exposure characterization without possible challenge. The link we made was based empirically on the judgments of curriculum experts since we were unable to link students with their reported instructional experiences (as measured by the 1992 NAEP background questionnaires) in our analyses.

In our judgment, descriptions of the achievement levels are not informative unless they accurately portray what students at the various levels can do. Characterizations of the levels should align with the actual performance of students on the NAEP, and empirical evidence of that alignment should meet reasonable standards. The likelihood that these goals can be met depends not only on the processes used to set the levels and establish descriptions, but also on the characteristics of the NAEP itself. For example, the item pool must be rich at each of the levels, and it must represent adequately the skills and knowledge that are the basis for setting the levels and that are used to describe them. Neither of these criteria was consistently satisfied in the establishment of the 1992 achievement levels in mathematics.

The task in mathematics (and perhaps in other areas) is all the more difficult because the field is still in the early stages of major curriculum reform where there is considerable variability in the penetration and extent of reform at the classroom level. Under such circumstances, defining achievement levels based on what students can do now may differ markedly from what is deemed desirable that they be able to do if the reform takes hold. This creates the natural tension between building the assessment and associated achievement levels around the desired new curriculum frameworks to capture what we want students to be able to accomplish versus grounding them accurately in the current prevailing conditions based on assessment frameworks and associated item pools that no longer represent the full range of desired learning goals. **The flaws in the content descriptions identified in our work can be attributed in part to procedural problems—insufficient attention to align descriptions and exemplars with actual student performance.** Nevertheless, it may well be that some fundamental shortcomings of the current achievement level effort are inextricably tied to the mismatch between the natural desire to move beyond the current horizon with an assessment design and associated data that are not appropriate to the task.

The concerns identified above may need to be addressed in the context of the newly developed assessment framework in mathematics that will guide the 1995 NAEP mathematics assessment. One important step for NAGB to adopt in establishing achievement levels in mathematics would be to start the process anew by designing their level setting and characterization to align closely with the development of the new assessment frameworks, items, and associated data collection. We believe that linking level setting with assessment design from the outset may provide the only appropriate and fair means to determine whether it is possible to develop valid content descriptions of what students can do.

References

American College Testing. (1993). Description of mathematics achievement level setting process and proposed achievement levels descriptions. Washington, DC: National Assessment Governing Board.

Bourque, M.L. (1993). The NAEP achievement level setting process for the 1992 mathematics assessment. Paper presented at a joint symposium of the American Educational Research Association and the National Council for Measurement in Education annual meeting, Atlanta, GA.

General Accounting Office. (1993). Educational achievement standards: NAGB's approach yields misleading interpretations. Washington, DC: Author.

Koretz, D.M., & Deibert, E. (1993, forthcoming). Interpretations of NAEP anchor points and achievement levels by the media in 1991. Washington DC: The RAND Corporation.

Linn, R.L., Koretz, D.M., Baker, E.L., & Burstein, L. (1991). The validity and credibility of the achievement levels for the 1990 National Assessment of Educational Progress in mathematics (CSE Tech. Rep. No. 330). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Mullis, I.V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). NAEP 1992 mathematics report card for the nation and the states. Washington, DC: U.S. Department of Education.

National Assessment Governing Board. (1991). Response to the draft summative evaluation report on the National Governing Board's inaugural effort to set achievement levels on the National Assessment of Educational Progress. Washington, DC: Author.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

Novak, J., Burstein, L., & Sugrue, B. (forthcoming). Sources of variability in mathematics educators' mapping of achievement level descriptions to 1992 NAEP mathematics test items (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Novak, J., Burstein, L., & Larriva, C. (forthcoming). Characteristics of items that differentiate among students classified at different levels of achievement in the 1992 NAEP in mathematics (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Phillips, G.W., Mullis, I.V.S., Bourque, M.L., Williams, P.L., Hambleton, R.K., Owen, E.H., & Barton, P.E. (1993). Interpreting NAEP scales. Washington: U.S. Department of Education.

Reckase, M.D. (1993). The defensibility of domain descriptions for achievement levels and anchor points. Paper developed for the meeting of the Technical Advisory Committee on Standard Setting (TACSS) in support of the methodology devised by ACT for the National Assessment Governing Board, Washington, DC: February 9-10, 1993.

Spanos, G., Rhodes, N.C., Dale, T.C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R.R. Cocking & J.P. Mestre (Eds.), Linguistic and cultural influences on learning mathematics (pp. 221-240). Hillsdale, NJ: Erlbaum.

Stufflebeam, D.L., Jaeger, R.M., & Scriven, M. (1991). Summative evaluation of the National Governing Board's inaugural effort to set achievement levels on the National Assessment of Educational Progress. Kalamazoo: Western Michigan University.

Sugrue, B., Novak, J., Burstein, L., Lewis, E., Koretz, D., & Linn, R. (forthcoming). Matching test items to the 1992 NAEP mathematics achievement level descriptions: Mathematics educators' interpretations and their relationship to student performance (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Survey of Mathematics and Science Opportunities (1993) TIMSS curriculum analysis: A content analytic approach (Research Report Series No. 57, Third International Mathematics and Science Study). East Lansing: Michigan State University.

Appendix A

Parsed Versions of the NAEP Achievement Level Descriptions

NAEP Description of Mathematics Achievement Levels

for Basic, Advanced, and Proficient Fourth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra and functions. At the fourth-grade level, algebra and functions are treated in informal and exploratory ways, often through the study of patterns. Skills are cumulative across levels—from Basic to Proficient to Advanced.

GRADE 4

Basic 211

1. **Fourth-grade students performing at the basic level should show** *some evidence of understanding the mathematical concepts and procedures* **in the five NAEP content areas.**

2. Fourth graders performing at the level should be able to

   a. *estimate with whole numbers.*

   b. *use basic facts to perform simple computations with whole numbers;*

3. show

   a. *some understanding of fractions*

   b. *some understanding of decimals;*

4. and *solve some simple real-world problems* in all NAEP content areas.

5. Students at this level should be able to *use—though not always accurately—four-function calculators, rulers, and geometric shapes.*

6. Their *written responses* are often

a. *minimal*

b. and presented *without supporting information.*

Proficient 248

7. Fourth-grade students performing at the proficient level should *consistently apply integrated procedural knowledge and conceptual understanding to problem solving* in the five NAEP content areas.

8. Fourth graders performing at the proficient level should be able *to use whole numbers to*

a. estimate results,

b. *compute results,*

c. *determine whether results are reasonable.*

9. They should have a

a. *conceptual understanding of fractions*

b. *conceptual understanding of decimals;*

10. be able to *solve real-world problems* in all NAEP content areas;

11. *use four-function calculators, rulers, and geometric shapes appropriately.*

12. should *employ problem-solving strategies such as identifying and using appropriate information.*

13. Their *written solutions* should be

a. *organized*

b.*presented* both *with supporting information*

c. *presented with explanations of how they were achieved.*

Advanced 280

14. Fourth-grade students performing at the advanced level should *apply integrated procedural knowledge and conceptual understanding to complex and non-routine real-world problem solving* in the five NAEP content areas.

15. Fourth graders performing at the advanced level should be able to *solve complex and nonroutine real-world problems* in all NAEP content areas.

16. They should display *mastery in the use of four-function calculators, rulers, and geometric shapes.*

17. These students are expected to *draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved.*

18. They should

a. *go beyond the obvious in their interpretations*

b. and be able to *communicate their thoughts clearly*

c. and *communicate their thoughts concisely.*

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Eighth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

GRADE 8

Basic 256

1. Eighth-grade students performing at the basic level should exhibit *evidence of conceptual and procedural understanding* in the five NAEP content areas.

2. This level of performance signifies *understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.*

3. Eighth graders performing at the basic level should *complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs.*

4. They should be able to *solve problems* in all NAEP content areas

a. *through the appropriate selection and use of strategies*

b. *the appropriate selection and use and technological tools—including calculators, computers, and geometric shapes.*

5. Students at this level should also be able to

a. *use fundamental algebraic concepts in problem solving.*

b. and *use informal geometric concepts in problem solving.*

6. As they approach the proficient level, students at the basic level should be able to

a. *determine which of available data are necessary and sufficient for correct solutions*

b. and *use them [data] in problem solving.*

7. However, these 8th graders show l*imited skill in communicating mathematically.*

Proficient 294

8. **Eighth-grade students performing at the proficient level should *apply mathematical concepts and procedures consistently to complex problems* in the five NAEP content areas.**

9. Eighth graders performing at the proficient level should be able to

a. *conjecture,*

b.*defend their ideas,*

c. and *give supporting examples.*

10.They should

a. *understand the connections between fractions, percents, decimals,*

b. and *[connections between] other mathematical topics such as algebra and functions.*

11. Students at this level are expected to have a *thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical solutions.*

12. *Quantity and spatial relationships in problem solving and reasoning* should be familiar to them,

13. and they should be able to *convey underlying reasoning skills beyond the level of arithmetic.*

14. They should be able to

a.*compare and contrast mathematical ideas* and

b. *generate their own examples.*

15. These students should *make inferences from data and graphs;*

16. *apply properties of informal geometry;*

17. and *accurately use the tools of technology.*

18. Students at this level should

a. *understand the process of gathering and organizing data*

b. and be able to *calculate and evaluate results within the domain of statistics and probability.*

c. and *communicate results within the domain of statistics and probability.*

Advanced 331

19. **Eighth-grade students performing at the advanced level should be able to**

a. reach beyond the recognition, identification, and application of mathematical rules in order to generalize

b. **and** ***synthesize concepts and principles*** **in the five NAEP content areas.**

20. Eighth graders performing at the advanced level should be able to *probe examples and counter-examples in order to shape generalizations from which they can develop models.*

21. Eighth graders performing at the advanced level should

a. *use number sense to consider the reasonableness of an answer.*

b. and *use geometric awareness to consider the reasonableness of an answer.*

22. They are expected to

a. *use abstract thinking to create unique problem-solving techniques*

b. and *explain the reasoning processes underlying their conclusions.*

Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Twelfth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

GRADE 12

Basic 287

1. Twelfth-grade students performing at the basic level should *demonstrate procedural and conceptual knowledge in solving problems* in the five NAEP content areas.

2. Twelfth-grade students performing at the basic level should be able to *use estimation* to

a. *verify solutions as applied to real-world problems*

b. and *determine the reasonableness of results as applied to real-world problems*.

3. They are expected to

a. *use algebraic reasoning strategies to solve problems.*

b. and *use geometric reasoning strategies to solve problems.*

4. Twelfth graders performing at the basic level should *recognize relationships presented in verbal, algebraic, tabular, and graphical forms;*

5. and *demonstrate knowledge of geometric relationships and corresponding measurement skills.*

6. They should be able to *apply statistical reasoning*

a. i*n the organization and display of data*

b. and *in reading tables and graphs.*

7. They should be able to

a. *generalize from patterns and examples in the area of algebra,*

b. *generalize from patterns and examples in the area of geometry,*

c. *generalize from patterns and examples in the area of statistics.*

8. At this level, they should

a. *use correct mathematical language and symbols to communicate mathematical relationships*

b. and *use correct mathematical language and symbols to communicate mathematical reasoning processes;*

9. *use calculators appropriately to solve problems.*

Proficient 334

10. Twelfth-grade students performing at the proficient level should *consistently integrate mathematical concepts and procedures to the solutions of more complex problems* in the five NAEP content areas.

11. Twelfth-grade students performing at the proficient level should

a. *demonstrate an understanding of algebraic reasoning.*

b. *demonstrate an understanding of statistical reasoning.*

c. *demonstrate an understanding of geometric and spatial reasoning.*

12. They should be able to *perform algebraic operations involving polynomials;*

13. *justify geometric relationships;*

14. and *judge and defend the reasonableness of answers as applied to real-world situations.*

15. These students should be able to *analyze and interpret data in tabular and graphical form;*

16. *understand the elements of the function concept in symbolic, graphical, and tabular form;*

17. and *use elements of the function concept in symbolic, graphical, and tabular form;*

18. and

a. *make conjectures,*

b. *defend ideas,*

c. and *give supporting examples.*

Advanced 366

19. Twelfth-grade students performing at the advanced level should a.*consistently demonstrate the integration of procedural and conceptual knowledge*

b.and *consistently demonstrate the synthesis of ideas* in the five NAEP content areas.

20. Twelfth-grade students performing at the advanced level should *understand the function concept;*

21. and be able to

a. *compare the numeric, algebraic, and graphical properties of functions.*

b. and *apply the numeric, algebraic, and graphical properties of functions.*

22. They should a*pply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous and discrete mathematics.*

23. They should be able to *formulate generalizations and create models through probing examples and counter examples.*

24. They should be able to *communicate their mathematical reasoning* through the *clear, concise, and correct use of mathematical symbolism and logical thinking,*

Appendix B

Final Versions of Descriptors Used to Map NAEP Assessment Items

Grade 4 Descriptors

Block ________ Item _______ Item ID __________________________

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ____ If the item involves whole numbers, check any of the following descriptions that apply:

   The item calls for:

   1(a) ____ using basic number facts to perform simple computations with whole numbers ?

   1(b) ____ estimating with whole numbers ?

   1(c) ____ using whole numbers to compute results ?

   1(d) ____ using whole numbers to estimate results ?

   1(e) ____ determining of the reasonableness of whole number results ?

2. ____ If the item involves fractions or decimals, indicate which one of the following descriptions best applies to the item:

   The item calls for:

   2(a) ____ some understanding of fractions or decimals ?

   or

   2(b) ____ conceptual understanding of fractions or decimals ?

3. ____ The item calls for understanding of mathematical concepts or mathematical procedures. ?

4. ____ The item calls for applying integrated procedural and conceptual understanding to problem solving ?

5. ____ The item calls for applying integrated procedural and conceptual understanding to complex and nonroutine real-world problem solving ?

6. ____ If the item calls for real-world problem-solving, check which one of the following best describes the item:

The item calls for:

6(a) ____ solving a simple real-world problem ?

or

6(b) ____ solving a [routine] real-world problem ?

or

6(c) ____ solving a complex and nonroutine real-world problem ?

7. ____ The item calls for employing problem-solving strategies such as identifying and using appropriate information ?

8. ____ If the item calls for a written response, check any of the following descriptions that apply:

The item calls for:

8(a) ____ giving supporting information ?

8(b) ____ explaining how the answer or solution process was achieved ?

8(c) ____ explaining why the answer or solution process was achieved ?

8(d) ____ clear or concise communication ?

Grade 8 Descriptors

Block ________ Item _______ Item ID ______________________

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ____ The item calls for an understanding of arithmetic operations—including estimation — on whole numbers, decimals, fractions or percents. ?

2. ____ The item calls for a thorough understanding of basic-level arithmetic operations — an understanding sufficient for problem solving in practical situations. ?

3. ____ The item calls for understanding the connections among any of the following: fractions, percents, decimals. ?

4. ____ The item calls for using fundamental algebraic concepts in problem solving. ?

5. ____ The item calls for understanding of the connection between algebra and functions. ?

6. ____ If the item involves geometric concepts, check any of the following descriptions that apply:

   The item calls for:

   6(a) ____ using informal geometric concepts in problem solving ?

   6(b) ____ applying the properties of informal geometry ?

   6(c) ____ using geometric awareness to consider the reasonableness of an answer ?

7. ____ The item calls for familiarity with quantity or spatial relationships in problem solving or reasoning. ?

8. ____ The item calls for completing problems with the help of structural prompts such as diagrams, charts, or graphs. ?

9. ____ The item calls for solving problems through the appropriate selection and use of strategies. ?

10. ____ The item calls for solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes. ?

11. ____ The item calls for using abstract thinking to create unique problem-solving techniques. ?

12. ____ The item calls for determining which of available data are necessary andsufficient for correct solutions. ?

13. ____ If the item involves working with data, check any of the following descriptions that apply:

    The item calls for:

    13(a) ____ making of inferences from data or graphs ?

    13(b)____ understanding of the process of gathering and organizing data ?

14. ____ If the item involves statistics or probability, check any of the following descriptions that apply:

    The item calls for:

    14(a) ____ calculating results within the domain of statistics or probability ?

    14(b)____ evaluating results within the domain of statistics or probability ?

    14(c) ____ communicating results within the domain of statistics or probability ?

15. ____ The item calls for conceptual understanding or procedural understanding ?

16. ____ The item calls for applying mathematical concepts and procedures to complex problems. ?

17. ____ The item calls for reaching beyond the recognition, identification, and application of mathematical rules to generalize and synthesize concepts and principles. ?

18. ____ The item calls for comparing and contrasting mathematical ideas. ?

19. ____ The item calls for generating one's own examples. ?

20. ____ The item calls for probing of examples and counter examples in order to shape generalizations from which the student can develop models. ?

21. ____ The item calls for the using number sense to consider the reasonableness of an answer. ?

22. ____ If the item requires a written response, check any of the following descriptions that apply:

The item calls for:

22(a) ____ making conjectures ?

22(b)____ defending ideas ?

22(c) ____ giving supporting examples ?

22(d)____ explaining the reasoning process underlying conclusions ?

22(e) ____ conveying underlying reasoning skills beyond the level of arithmetic ?

Grade 12 Descriptors

Block ________ Item _______ Item ID ______________________

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ____ If the item involves geometry, check any of the following descriptions that apply:

    The item calls for:

    1(a) ____ using geometric reasoning strategies to solve problems ?

    1(b) ____ knowledge of geometric relationships and corresponding measurement skills ?

    1(c) ____ an understanding of geometric reasoning ?

    1(d) ____ an understanding of spatial reasoning ?

    1(e) ____ justifying geometric relationships ?

    1(f) ____ generalizing from patterns or examples ?

2. ____ If the item involves algebra, check any of the following descriptions that apply:

    The item calls for:

    2(a) ____ using algebraic reasoning strategies to solve problems ?

    2(b) ____ an understanding of algebraic reasoning ?

    2(c) ____ performing algebraic operations involving polynomials ?

    2(d) ____ generalizing from patterns or examples ?

3. ____ If the item involves functions, check any of the following descriptions that apply:

    The item calls for:

3(a) ____ understanding of elements of the function concept in symbolic, graphical or tabular form ?

3(b) ____ understanding of the function concept ?

3(c) ____ using elements of the function concept in symbolic, graphical or tabular form ?

3(d) ____ comparing the numeric, algebraic, or graphical properties of functions ?

3(e) ____ applying the numeric, algebraic, or graphical properties of functions ?

4. ____ If the item involves data analysis or statistics, check any of the following descriptions that apply:

The item calls for:

4(a) ____ applying statistical reasoning in the organization and display of data ?

4(b) ____ applying statistical reasoning in reading tables or graphs ?

4(c) ____ an understanding of statistical reasoning ?

4(d) ____ analyzing and interpreting data in tabular or graphical form ?

4(e) ____ generalizing from patterns or examples ?

5. ____ The item calls for solution of problems in the more advanced area of continuous and discrete mathematics. ?

6. ____ The item calls for recognizing relationships presented in verbal, algebraic, tabular, or graphical forms. ?

7. ____ The item calls for formulating generalizations and creating models through probing examples and counterexamples. ?

8. ____ The item calls for using estimation to verify solutions to real-world problems. ?

9. ____ The item calls for using estimation to determine the reasonableness of results as applied to real-world problems. ?

10. ____ The item calls for judging or defending the reasonableness of answers as applied to real-world situations. ?

11. ____ The item calls for procedural knowledge or conceptual knowledge in solving problems. ?

12. ____ The item calls for integrating mathematical concepts and procedures to the solution of more complex problems. ?

13. ____ The item calls for the integration of procedural and conceptual knowledge, and the synthesis of ideas. ?

14. ____ If the item requires a written response, check any of the following descriptions that apply ?

The item calls for:

14(a) ____ using mathematical language and symbols to communicate mathematical relationships ?

14(b) ____ using mathematical language and symbols to communicate reasoning processes. ?

14(c) ____ clear and concise use of mathematical symbolism and logical thinking to communicate mathematical reasoning. ?

14(d) ____ defending ideas ?

14(e) ____ making conjectures ?

14(f) ____ giving supporting examples ?

APPENDIX C

SUMMARY OF CHARACTERISTICS OF THE JUDGES

A background and teaching experience questionnaire given to the 18 raters revealed the following information:

Personal History: There were 13 females and five males. Ethnic representation included 9 Caucasians, 5 African-Americans, 3 Hispanics, and 1 Asian.
All but one were currently in a teaching position at the time of the judging (the one exception was working as a clinical consultant for secondary mathematics in the UCLA teacher training program.)

Education Level: Every judge held a bachelors degree, eleven of which were in the fields of math, science or engineering, three in education, and six in other fields. Four held masters degrees and two had doctorates.

Years Teaching Math: Judges' mathematics teaching experience ranged from 1 to 33 years experience at the Elementary level, 1 to 9 years at Middle/Jr. High level, and 1 to 16 years at the Sr. High level. The mean number of years of mathematics teaching experience was 12.2 (median 12).

Certification: Every judge held a current teaching credential. Ten of the eighteen held credentials in mathematics, six in high school education, 8 in middle school education and 9 in elementary education.

Exposure to Topics through University Courses or In-service:
All judges (100%) had exposure to the following topic areas: methods of teaching math, numeration, measurement, problem solving, manipulatives, psychology of learning and teaching students from various cultures. All but one had exposure in geometry and probability.
The majority (at least 83%) had received training in the use of calculators, in the understanding of students' thinking about mathematics and in estimation. The majority had also spent more than 35 hours during the last 3 years on in-service education in the teaching of math.

Familiarity with Mathematics Standards: The majority (83%) were familiar with NCTM Curriculum and Evaluation Standards for School Mathematics and every judge was familiar with the California Mathematics Framework.

Conference Workshop Participation: Seventeen of the eighteen judges participated or presented at national or district conferences and/or workshops during the past two years.

Appendix D

DESCRIPTORS: LINGUISTIC FEATURES

Block______ Item______ Item ID______________________

Analyze each test item according to the following categories (descriptors). When a category does not apply write a 0. If you are not sure of a decision write a question mark "?."

1a. ______ **NUMBER OF COMPARATIVES-** Count the number of comparatives:
greater than/less than
older than/younger than
n times as much as *as in* *Roberto earns twice as much as I do.*
as...as *as in* *Jenna is as old as Rita.*
List the comparatives______________________________

1b. ______ **LOGICAL CONNECTORS-** Count the number of logical connectors:
Logical connectors are words or phrases which carry out the function of marking a logical relationship between two or more basic linguistic structures (and) serve a semantic, cohesive function indicating the nature of the relationship between parts of a text.
if...then *as in* *If Wendy earns 12 dollars an hour, then how much does she earn in 8 hours?*
if and only if *as in* *a+b=c if and only if b+a=c.*
given that *as in* *Given that a=0, axb=0.*
other examples include *such that, that is, for example, but, consequently, and either...or.*
List the logical connectors__________________________

2a. ______ **MATHEMATICAL VOCABULARY-** Count the number of words which are specific to mathematics. For example: *divisor, denominator, triangle, equation, quotient, polynomial.*
List the words____________________________________

2b. ______ **NATURAL LANGUAGE VOCABULARY-** Count the number of everyday words that have a different or specialized meaning in mathematics. For example: *rational, expression, radical, face, line, quarter, column, table, figure, simplify.*
List the words____________________________________

2c. _____ **COMPLEX STRING OF WORDS OR PHRASES-** Count the number of phrases which represent mathematical concepts. For example:
Additive inverse, obtuse triangle, least common multiple, rational expression.
List the phases or string of words____________________

2d. _____ **WORDS WHICH SIGNAL OPERATIONS-** Count the number of words and/or phrases which signal operations. For example:
add, plus, add, less, in all, exceed, differ, more than, sum, total, combined.
List the words and or phrases_______________________

3. _____ **CONCEPTS REQUIRING EXPERIENCE OR KNOWLEDGE** - Count the number of words and or phrases which represent concepts or knowledge from fields of experience other than mathematics. For example: Market-place concepts such as *markup, wholesale, sales tax rates, balance, checks..*
List the words and/or phrases_______________________

4a. _____ **WORDS WHICH FUNCTION AS UNITS OR HAVE QUANTITATIVE ATTRIBUTES-** Count the number of words and/or phrases which function as units of measure or denote a quantitative attribute. For example:
5 yards, 10 apples, 2 dollars, 20 mph, long, tall, old, deep, wide, far, weight, cost.
List the words and/or phrases_______________________

4b. _____ **QUANTITIES EXPRESSED IN WRITTEN TEXT-** Count the number of words which express quantities. For example:
one-fourth, twenty-one, quarter, half.
List the words____________________________________

5. _____ **MULTI-STEPNESS-** Count the number of steps in the solution process. Here a distinction is made between procedures or algorithms which require multiple steps (this will be considered a one step problem), and a solution process which requires multiple steps. The translation of a real world problem into its mathematical equivalent counts as one step, and then the solution of the appropriate equation counts as another.

Appendix E

Sources of Variability in Mapping of Descriptors to Items[1]

A series of large-scale generalizability analyses of the mappings of descriptors to the items by the judges were carried out for particular clusters of descriptors at each grade. The purpose was to examine the variability (technically, the variance components) associated with judges, descriptors, assessment items (classified variously by content and by item format and type) and their interactions. It was not possible to carry out these analyses for the fully crossed design. As mentioned earlier, the resulting matrices of observations at each grade level are gigantic when judges, items, and descriptors are crossed. (20,000 to 50,000 data points). Moreover, some of the descriptors are simply not applicable to all types of items. For example, if an item doesn't require a written response, then the descriptors that apply only to such items are logically impossible. To treat these combinations as meaningful observations would be to introduce data taht artificially impacts the estimated variance components. Therefore, descriptors and items were partitioned into clusters that were realistically mappable and analyses run for clusters of descriptors. Further details in the formation of descriptor clusters are provided in Novak, Burstein, & Sugrue (1993).

[1]This analysis was conducted by John Novak with assistance from Leigh Burstein and Brenda Sugrue.

Tables E.1, E.2, and E.3 show the percentage of total variance accounted for by each variance component in the item x descriptors x judge design when run for each each cluster of descriptors. The percentage of total variance accounted for by each variance component differs from cluster to cluster, and from grade to grade, making it difficult to draw general conclusions. However, it seems that, in 4th and 8th grade, variance components accounting for the greatest percentages of the variance in mapping of descriptors to items are descriptors and the interaction of raters with descriptors. This indicates that there was considerable variability in judges' interpretation of some clusters of descriptors, in particular written response and estimation descriptors in 4th grade, estimation and number and operations descriptors in 8th grade, geometry, data analysis, and written response descriptors in 12th grade. For other clusters of descriptors, in general, item and the interaction of raters with items were the variance components that accounted for the largest percentage of total variability.

Table E.1: Percentage of total variance accounted for by each variance component for each cluster of descriptors, Grade 4

Cluster	I	R	D	IR	ID	RD	Error
Whole Numbers	6	4	17	4	17	11	42
Written Resp	0	11	12	0	4	25	48
"Prob Solv 4,5"	5	0	38	2	1	16	37
Estimation	4	11	0	2	2	22	59
Gray Area	2	0	57	1	2	10	28

Table E.2: Percentage of total variance accounted for by each variance component for each cluster of descriptors, Grade 8

Cluster	I	R	D	IR	ID	RD	Error
Geometry	5	10	7	11	6	6	55
Alg & Functions	10	3	16	14	13	1	42
Problem Solving	3	6	13	0	9	7	62
Data Analysis	0	6	5	3	5	10	71
Nums & Ops	1	8	17	0	14	16	45
Estimation	0	5	46	0	6	13	31
Gray Area	0	9	0	0	0	1	90
Other Vague	6	1	0	9	3	8	73
Written Resp	2	15	10	7	4	11	51

Table E.3: Percentage of total variance accounted for by each variance component for each cluster of descriptors, Grade 12

Cluster	I	R	D	IR	ID	RD	Error
Geometry	3	7	7	5	12	16	49
Alg & Functions	6	4	15	6	9	8	52
Algebra	6	5	15	5	6	11	52
Functions	20	5	0	21	2	3	49
Problem Solv	0	1	12	1	18	7	61
Prob Solv 9-10	15	4	1	24	5	2	50
Data Analysis	8	5	2	5	8	16	56
Estimation	1	3	5	8	0	4	77
Gray Area	0	5	22	8	5	8	51
Other Vague	0	2	0	3	6	4	85
Written Resp	3	1	1	15	2	14	64

Relative and absolute generalizability coefficients for each cluster of descriptors are presented in Tables E.4, E.5, and E.6. Generalizability varied from cluster to cluster of descriptors. G-coefficients were higher for clusters of descriptors that related to specific mathematics content than to processes such as problem solving, or to characteristics of written responses. The lowest coefficients occur in the case of the real world problem descriptor in 4th grade, the gray area descriptors in 8th grade, and the estimation descriptors in 12th grade. This indicates that the descriptors which were most inconsistently interpreted and mapped to items were those that did not reference specific mathematics content.

Table E.4: Relative and absolute generalizability coefficients for each cluster of descriptors, Grade 4

Cluster	Relative	Absolute
Whole Numbers	.88	.92
Written Resp	.42	.64
"Prob Solv 4,5"	.58	.66
Estimation	.39	.52
Gray Area	.64	.70

Table E.5: Relative and absolute generalizability coefficients for each cluster of descriptors, Grade 8

Cluster	Relative	Absolute
Geometry	.61	.69
Alg & Functions	.78	.80
Problem Solving	.82	.89
Data Analysis	.36	.42
Nums & Ops	.75	.85
Estimation	.63	.76
Gray Area	.03	.03
Other Vague	.37	.40
Written Resp	.52	.68

Table E.6: Relative and absolute generalizability coefficients for each cluster of descriptors, Grade 12

Cluster	Relative	Absolute
Geometry	.80	.88
Alg & Functions	.85	.88
Algebra	.74	.80
Functions	.79	.82
Problem Solv	.92	.93
Prob Solv 9-10	.69	.71
Data Analysis	.81	.87
Estimation	.19	.21
Gray Area	.48	.54
Other Vague	.41	.43
Written Resp	.49	.52

Appendix F

The Distribution of Differentiating Items

Across Blocks of Items

As part of our routine descriptive analyses, we generated the distributions of the items that differentiated among achievement levels across all blocks of items at all three grades. The results from these analyses are provided in Tables E.1-E.3. Two overall findings reflected in these tables are

(a) **The number of differentiating items varies across the blocks of test items**; and

(b) **Blocks vary in terms of the levels at which the items they contain differentiate.**

Essentially, this implies that students are being classified on the basis of extrapolation in that some students are administered blocks that contain few or no items that differentiate at the achievement levels appropriate for them.

The evidence on the first point is that across the three grades, there were anywhere from 0 to 9 items which differentiated the levels (range 1-5, median 3 at grade 4; range = 0-9, median 3 at grade 8; range = 1-9, median 4 at grade 12).

Moreover, some blocks have a considerable number of items from a single level but no, or very few, items that differentiate at other levels. At grade 4, Block 11 has 4 Proficient items and none at the other levels; all 3 of Block 4's differentiating items are also at the Proficient level. Block 5 has 8 Basic level differentiating items at Grade 8 while

Block 7 has 3 Basic and no others and Block 13 has 3 Proficient and no others. At grade 12, Basic level differentiators were most frequent for Blocks 6(6) and 4(4) while Proficient level items were most prevalent for Block 7(4). Only two blocks (Block 11 at Grade 8 and Block 3 at Grade 12) have at least 2 differentiating items at each level.

Of course, students were administered booklets of 3 blocks of items so the mixtures of items from the achievement levels may have evened out. However, given the intent of the use of achievement levels, adequate measurement at all levels in the assessment exercises presented to each student should not be left to chance. This is particularly important given that the levels are intended to represent, not just points on a continuous unidimensional scale, but mastery of specific types of knowledge and skills.

Table F.1

Distribution of differentiating items across blocks for grade 4.

Block	B	P	A	Total
3	3	0	1	4
4	0	3	0	3
5	2	0	2	4
6	3	0	1	4
7	0	0	2	2
8	1	0	0	1
9	2	1	2	5
10	1	0	0	1
11	0	4	0	4
12	2	0	0	2
13	1	2	0	3
14	0	1	1	2
15	0	1	0	1
16	0	1	1	2

Table F.2

Distribution of differentiating items across blocks for grade 8.

Block	B	P	A	Total
3	1	2	1	4
4	3	1	2	6
5	8	1	0	9
6	3	1	0	4
7	3	0	0	3
8	1	3	3	7
9	0	1	1	2
10	0	0	0	0
11	3	2	2	7
12	1	1	0	2
13	0	3	0	3
14	0	0	0	0
15	1	0	2	3
16	0	2	0	2

Table F.3

Distribution of differentiating items across blocks for grade 12

Block	B	P	A	Total
3	2	2	3	7
4	4	1	0	5
5	3	2	2	7
6	6	2	1	9
7	2	4	2	8
8	3	3	1	7
9	2	1	0	3
10	0	2	0	2
11	2	2	0	4
12	1	2	0	3
13	1	0	0	1
14	1	3	0	4
15	1	1	2	4
16	1	1	1	3

Appendix G

Description of the Coding of Differentiating Items According to Item Signatures

The math content coding of the items that statistically differentiated among the achievement levels took place in two phases. In the first phase, the items that had been identified by all three statistical criteria were examined. A later phase examined the two criteria items. The coding of the three criteria items began with the eighth grade items. A panel of five experts (former mathematics teachers and current graduate students) examined each item satisfying the three criteria, and listed the relevant content attributes of that item. If an attribute was already on the list, that item was added to the list of items possessing that attribute. Any attribute not already on the attribute list was added. All decisions about attributes were made by consensus, and the existing list was carried over to each successive set of items. The eighth grade, three criteria items at the Basic, Proficient, and Advanced levels (in that order) were examined on the first day. On the next day, the fourth grade items were coded (again, Basic, Proficient, then Advanced) by a panel of three experts, and finally the twelfth grade items were coded by a panel of three experts. Panel membership overlapped across the sessions.

Approximately three months separated the two phases of coding. The same basic procedure was used in the second phase, which focused on the items that met only the first two statistical criteria. One major difference was that at the start of the second phase, the list of attributes was nearly complete; only two additional attributes were added during this phase. A second difference was that the experts were also asked to code the items on their linguistic attributes as well as their math content attributes. It took two days to complete the content examination of the two criteria items and the linguistic analysis of the two and three criteria items.

Appendix H

Classification of Differentiating Items According to the TIMSS Curriculum Framework

In an effort to provide a more parsimonious characterization, the descriptors were also assigned to appropriate categories in the content aspect of the TIMSS curriculum framework in mathematics. When there was no appropriate content category from TIMSS that applied to a descriptor, new categories were constructed. These categories largely reflected either the type of problem or exercise (real world problems, story problems, written response, calculator, multi-step, complex problem solving, recall of definition and recall of rule/formula/property) or the type of representation in the item or the required response (diagram, alternative symbol systems, spatial reasoning, visualization). Table H.1 contains both the original categories of descriptors and their correspondence to the TIMSS content categories as supplemented.

When the content attributes are collapsed into the supplemented TIMSS content categories[2], the resulting data (Tables H.2-H.4) further focus attention on the major loci of items at each level and grade. Apparently, items entailing arithmetic operations and number sense are prevalent at all levels and grades. Real world or story problems are also frequent differentiators everywhere except at the Advanced level

[2]The item entries in the TIMSS content category tables are unduplicated counts. Any items associated with two or more content attributes from the same TIMSS cell were counted only once.

at grade 8 and the Proficient level at Grade 12; likewise measurement content differentiates performance for at least two levels at each grade. On the other hand, while there are a few differentiating items earlier, a substantial number of such items covering either geometry or algebra content were identified only at grade 12. Data Analysis, probability, and statistics items were prominent at grade 8 in particular. There were what we thought, at first, to be a surprising number of recall items differentiating at both grades 8 and (especially) 12. Apparently, however, these codes highlight the increasing need for students to have solid command of mathematical terminology, formulae and algorithms to solve problems as they progress to more sophisticated coursework, which in turn is tapped by the distinctions among the levels in the middle and upper grades.

Using the TIMSS classifications, there is also a general shift from the Numbers category to the Recall category. An examination of the items indicates that what differentiates between levels on these categories for students at lower grade/achievement levels is computational ability, whereas at higher grade/achievement levels, specific content knowledge plays an increasing role.

Table H.1

TIMSS based categories	Original math coding categories
Numbers	algebra of integers decimals fractions money number line number sense percentage place value square root arithmetic operations conversions conversions/%-decimal-fraction
Measurement	estimation measurement metric units use of rulers/tools
Geometry	coordinate geometry geometric properties geometry
Proportionality	proportional reasoning
Functions, relations, and equations	algebraic operations number sentences pattern recognition algebraic reasoning substitution
Data representation, probability, and statistics	permutations/combinations probability statistics tables/graphs/charts sampling
Validation and structure	explain reasoning logical organization logical reasoning
Real World / Story Problems	real world story problems written response
Diagrams / Alternative Symbol Systems	alternative symbol systems diagram
Calculator	calculator
Complex / Multi-step problems	multi-step complex problem solving
Recall	recall of definition recall of rule/formula/property
Visualization / Spatial	spatial reasoning visualization

Table H.2
Mapping of Differentiating Items to Third International Mathematics and Science Study (TIMSS) Framework, Grade 4

TIMSS category	Level			
Frequency Col pct	Basic	Proficient	Advanced	Total
numbers	8 22.86	10 23.81	9 28.13	27
measurement	3 8.57	4 9.52	2 6.25	9
geometry1	1 2.86	1 2.38	1 3.13	3
proportionality	0 0.00	1 2.38	0 0.00	1
func, rel, equat	1 2.86	3 7.14	0 0.00	4
data, prob, stat	4 11.43	2 4.76	2 6.25	8
validation struc	1 2.86	0 0.00	1 3.13	2
real world, story problems	5 14.29	8 19.05	6 18.75	19
diagram alt symbol	6 17.14	4 9.52	3 9.38	13
calculator	1 2.86	0 0.00	0 0.00	1
complex multi-step	3 8.57	7 16.67	5 15.63	15
recall	0 0.00	1 2.38	2 6.25	3
visual / spatial	2 5.71	1 2.38	1 3.13	4
Total	35	42	32	109

Table H.3
Mapping of Differentiating Items to Third International Mathematics and Science Study (TIMSS) Framework, Grade 8

TIMSS category	Level			
Frequency Col pct	Basic	Proficient	Advanced	Total
numbers	11 18.33	14 40.00	4 10.00	29
measurement	5 8.33	1 2.86	4 10.00	10
geometry1	6 10.00	1 2.86	3 7.50	10
proportionality	3 5.00	1 2.86	0 0.00	4
func, rel, equat	3 5.00	1 2.86	2 5.00	6
data, prob, stat	7 11.67	5 14.29	5 12.50	17
validation structure	1 1.67	0 0.00	2 5.00	3
real world / story problems	7 11.67	6 17.14	1 2.50	14
diagram / alt symbol	8 13.33	1 2.86	4 10.00	13
calculator	1 1.67	2 5.71	1 2.50	4
complex / multi-step	1 1.67	0 0.00	5 12.50	6
recall	6 10.00	2 5.71	5 12.50	13
visual / spatial	1 1.67	1 2.86	4 10.00	6
Total	60	35	40	135

Table H.4
Mapping of Differentiating Items to Third International Mathematics and Science Study (TIMSS) Framework, Grade 12

TIMSS category Frequency Col pct	Level Basic	 Proficient	 Advanced	Total
Numbers	**15** 20.00	**10** 12.20	**6** 9.84	31
Measurement	**4** 5.33	1 1.22	**3** 4.92	8
Geometry 1	**10** 13.33	**13** 15.85	**8** 13.11	31
Proportionality	1 1.33	1 1.22	**3** 4.92	5
Func, Rel, Equations	**8** 10.67	**14** 17.07	**5** 8.20	27
Data, Prob, Stat	**5** 6.67	1 1.22	**3** 4.92	9
Validation Structure	1 1.33	0 0.00	0 0.00	1
Real World / Story probs	**7** 9.33	2 2.44	**5** 8.20	14
Diagram / Alt Symbol	8 10.67	9 10.98	7 11.48	24
Calculator	**3** 4.00	**5** 6.10	**3** 4.92	11
Complex / Multi-Step	**3** 4.00	**7** 8.54	**8** 13.11	18
Recall	**8** 10.67	**15** 18.29	**9** 14.75	32
Visual / Spatial	2 2.67	**4** 4.88	1 1.64	7
Total	75	82	61	218

Appendix I

NAEP Public Release Items that Differentiate Among Achievement Levels

Grade 4

```
                          1111112223456666788888
                           abcde ab     abc  abcd

M021901 1 B  5  1 6306020005516240410100
M022802 1 B  5 11 3111221105404310410100
M041501 1 B 12  3 6306030005305050410100
M042001 1 B 12  8 5215110005205140410100

                          1111112223456666788888
                           abcde ab     abc  abcd
M044001 1 P 14  5 5303030006303210310100
M048801 1 P 15  8 1000010006312101330101

                          1111112223456666788888
                           abcde ab     abc  abcd
M023101 1 A  5 14 0000000005612101310100
M023401 1 A  5 17 6406020006505050410100
M045001 1 A  7  6 2201026156506051410100
M045101 1 A  7  7 5305031016516141640102
M044301 1 A 14  9 2202015326406150510100
```

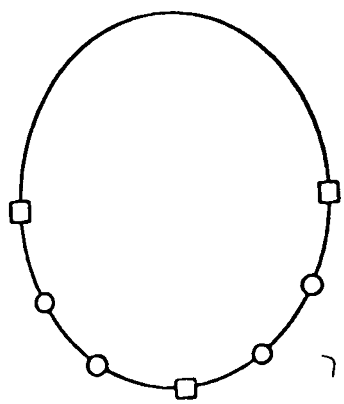
DESCRIPTORS

```
                        1111112223456666788888
                         abcde ab    abc  abcd

M021901 1 B  5  1 630602000551624041010O
```

BASIC GR 4 Released

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M021901	*	.62	.38	.73	.91	.97



Each □ costs 6¢
Each ○ costs 4¢

1. If the string does not cost anything, how much does the necklace above cost?

Ⓐ 10¢

Ⓑ 24c

Ⓒ 28c

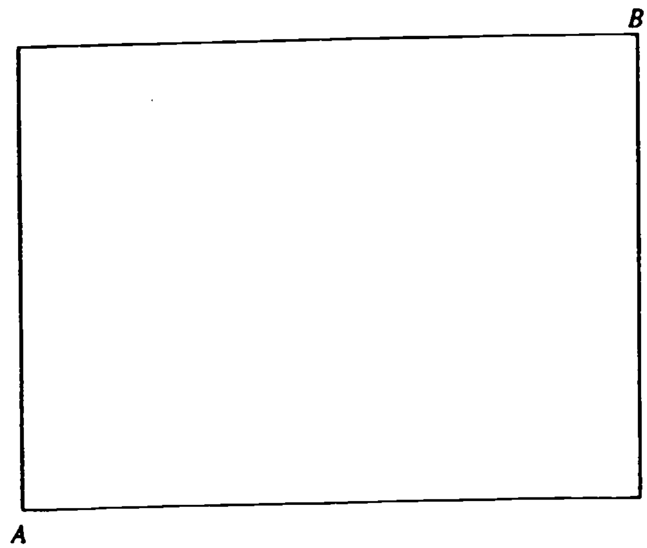
Ⓓ 34¢

M021901

DESCRIPTORS
11111122234566666788888
abcde ab abc abcd

M022802 1 B 5 11 3111221105404310410100

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M022802	*	.60	.31	.75	.92	.87



Use your centimeter ruler to make the following measurements to the nearest centimeter.

4-B

11. What is the length in centimeters of the diagonal from *A* to *B* ?

Answer: ______________________ M022802

DESCRIPTORS
1111112223456666788888
abcde ab   abc abcd

M041501 1 B 12   3 6306030005305050410100

BASIC GR 4 Released

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M041501	*	.55	.26	.66	.91	.92

3. A store sells 168 tapes each week. How many tapes does it sell in 24 weeks?

Ⓐ 7

Ⓑ 192

Ⓒ 4,032

Ⓓ 4,172

Did you use the calculator on this question?

◯ Yes ◯ No

M000638

DESCRIPTORS
11111122234566667888888
abcde ab   abc abcd

M042001 1 B 12 8 52151100052051404101OO

BASIC GR 4 Released

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M042001	*	.66	.45	.75	.89	.92

▶ **Questions 8-10** refer to the following table.

POINTS EARNED FROM SCHOOL EVENTS

Class	Mathathon	Readathon
Mr. Lopez	425	411
Ms. Chen	328	456
Mrs. Green	447	342

M000653

8. Which class earned the most points from the two events?

Ⓐ Mr. Lopez' class

Ⓑ Ms. Chen's class

Ⓒ Mrs. Green's class

Ⓓ All classes earned the same amount.

Did you use the calculator on this question?

○ Yes ○ No

M000655

DESCRIPTORS
11111222345666678888
abcde ab   abc abcd

M044001 1 P 14 5 53030300063032103101 00

PROFICIENT GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M044001	*	.37	.18	.41	.72	.87

5. Marlene made 6 batches of muffins. There were 24 muffins in each batch. Which of the following number sentences could be used to find the number of muffins she made?

Ⓐ $6 \times \square = 24$

Ⓑ $6 \div 24 = \square$

Ⓒ $6 + \square = 24$

Ⓓ $6 \times 24 = \square$

M000577

Did you use the calculator on this question?

○ Yes ○ No

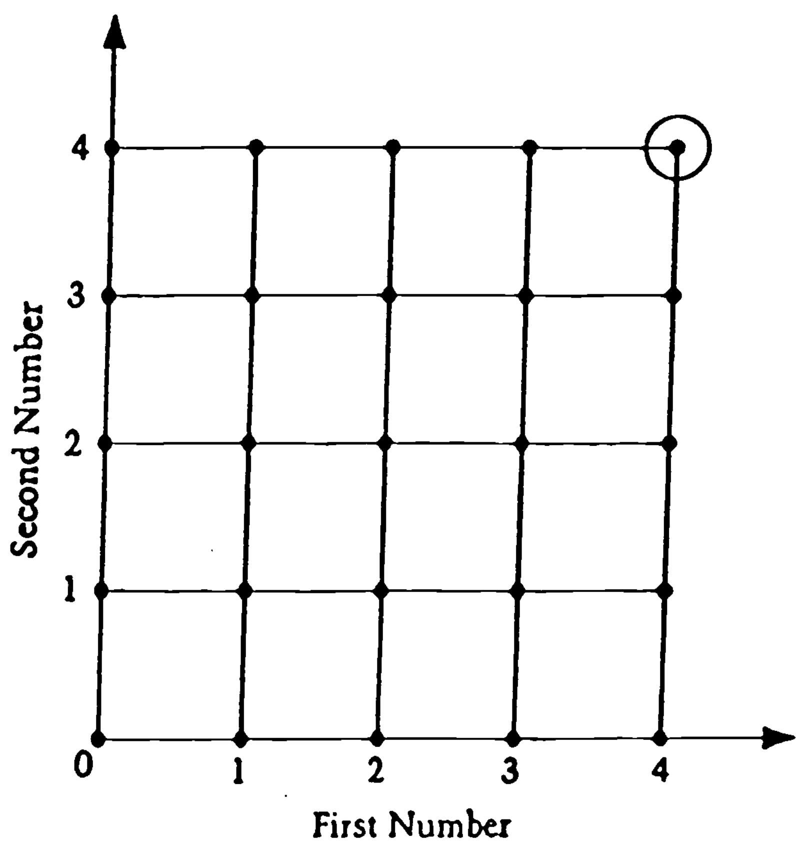
DESCRIPTORS
111111222345666678888
abcde ab   abc   abcd

M048801 1 P 15 8 100001000631210133010l

PROFICIENT GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M048801	*	.38	.10	.44	.79	.95

8. On the grid below, the dot at (4, 4) is circled. Circle two other dots where the first number is equal to the second number.



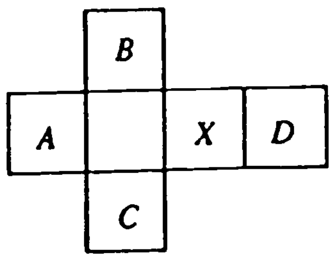
Y002408

DESCRIPTORS
11111122234566666788888
abcde ab   abc abcd

M023101 1 A 5 14 00000000005612101310100

ADVANCED GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023101	*	.22	.13	.18	.48	.90



14. The squares in the figure above represent the faces of a cube which has been cut along some edges and flattened. When the original cube was resting on face *X*, which face was on top?

- Ⓐ *A*
- Ⓑ *B*
- Ⓒ *C*
- Ⓓ *D*

M023101

DESCRIPTORS
1111112223456666788888
abcde ab   abc abcd

M023401 1 A 5 17 6406020006505050410100

ADVANCED GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023401	*	.19	.14	.16	.36	.67

17. A rectangular carpet is 9 feet long and 6 feet wide. What is the area of the carpet in square feet?

Ⓐ 15

Ⓑ 27

Ⓒ 30

Ⓓ 54

M023401

DESCRIPTORS
11111122234566666788888
abcde ab   abc abcd

M045001 1 A 7 6 22010261565060051410100

ADVANCED GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M045001	*	.21	.13	.19	.37	.73

B 6. If $1\frac{1}{3}$ cups of flour are needed for a batch of cookies, how many cups of flour will be needed for 3 batches?

Ⓐ $4\frac{1}{3}$

Ⓑ 4

Ⓒ 3

Ⓓ $2\frac{2}{3}$

M000561

DESCRIPTORS
11111122234566666788888
abcde ab   abc abcd

M045101 1 A 7 7 5305031016516141640102

ADVANCED GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M045101	*	.22	.06	.23	.50	.85

7. Jill needs to earn $45.00 for a class trip. She earns $2.00 each day on Mondays, Tuesdays, and Wednesdays, and $3.00 each day on Thursdays, Fridays, and Saturdays. She does not work on Sundays. How many weeks will it take her to earn $45.00 ?

Answer: ______________________________ M000558

DESCRIPTORS
11111122234566667888888
abcde ab   abc abcd

M044301 1 A 14 9 22020153264061505101000

ADVANCED GR 4 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M044301	*	.08	.04	.05	.17	.70

9. A package of birdseed costs $2.58 for 2 pounds. A package of sunflower seeds costs $3.72 for 3 pounds. What is the difference in the cost <u>per pound</u>?

Ⓐ $0.05

Ⓑ $1.14

Ⓒ $1.24

Ⓓ $1.29

M000581

Did you use the calculator on this question?

○ Yes   ○ No

Grade 8

```
                              11111111111111122222222
                  12345666678901233344445678901222222
                        abc       ab abc        abcde

M022201 1 B  5  4 01000645265552000000006100100100100

M022501 1 B  5  7 30200525152451000000006100101200100

M022801 1 B  5 10 21100663143140100000005000001200100

M022901 1 B  5 12 44000000010300000000005100004000000

M023001 1 B  5 13 45120000021211000000004100004000000

M023201 1 B  5 15 34121000025411144111114110002000000

M023301 1 B  5 16 32120000052401111153425100013000000

M023601 1 B  5 19 21111000034400122000004010013000000

M044601 1 B  7  2 23110656245331011100005100112200010

M045001 1 B  7  6 54310000021211011111015200014000000

M045301 1 B  7  9 11120000025411321255256110021100000

M053501 1 B 12  1 32142000020222100000006101222000000

M048701 1 B 15  7 35211000031322111100004001222525451

M049601 1 B 15 15 00000000035604222210002121110000000
```

```
                              11111111111111122222222
                  12345666678901233344445678901222222
                        abc       ab abc        abcde

M023501 1 P  5 18 25120000011511222254325210003000000

M053901 1 P 12  5 55620000010361100000003300003000000
```

```
                              11111111111111122222222
                  12345666678901233344445678901222222
                        abc       ab abc        abcde

M049101 1 A 15 10 51100000030120100000005101115000000

M049401 1 A 15 13 45023000014511133100004111203000000
```

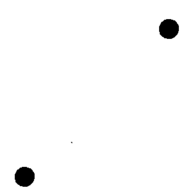
BASIC GRADE 8 RELEASE

```
                                   DESCRIPTORS
                              11111111111111122222222
                     12345666678901233344445678901222222
                            abc        ab abc      abcde

M022201 1 B  5  4 0100064[illegible]265552000000006100100100100
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M022201	*	.67	.46	.76	.86	.93

4. In the space below, use your ruler to draw a square with two of its corners at the points shown.



M022201

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                                 1111111111111122222222
                         12345666678901233344445678901222222
                              abc        ab abc        abcde

M022501 1 B  5  7 302005251624510000000006100101200100
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M022501	*	.58	.27	.71	.89	.95

7. In the space below, draw a rectangle 2 inches wide and $3\frac{1}{2}$ inches long.

M022501

BASIC GRADE 8 RELEASE

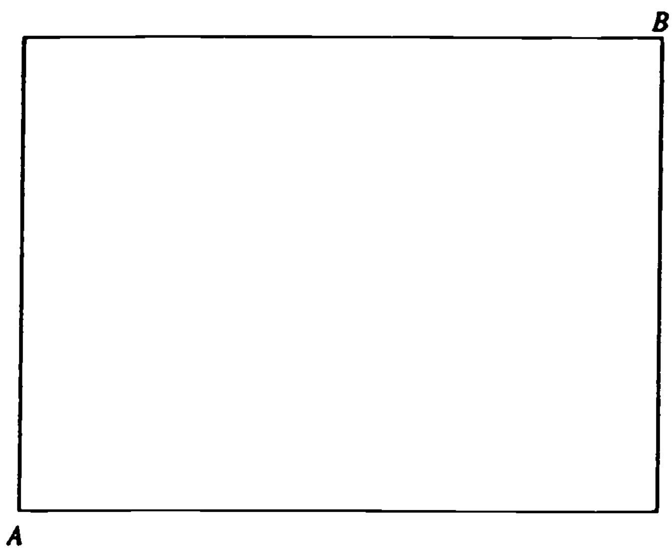
DESCRIPTORS

```
                         1111111111111122222222
             12345666678901233344445678901222222
                   abc          ab abc        abcde

 M022801 1 B  5 10 211006631431401000000005000001200100
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M022801	*	.71	.43	.84	.95	1.00

▶ **Questions 10–11** refer to the rectangle below.



Use your centimeter ruler to make the following measurements to the nearest centimeter.

10. What is the length in centimeters of one of the longer sides of the rectangle?

Answer: ______________________

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                         1111111111111122222222
               123456666789012333444456789012222222
                  abc        ab abc         abcde

M022901 1 B  5 12 44000000010300000000000510000400000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M022901	*	.72	.49	.80	.97	1.00

12. By how much would 217 be increased if the digit 1 were replaced by a digit 5 ?

Ⓐ 4

Ⓑ 40

Ⓒ 44

Ⓓ 400

M022901

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                         1111111111111122222222
              12345666678901233344445678901222222
                   abc        ab abc        abcde

 M023001 1 B  5 13 451200000212110000000004100004000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023001	*	.69	.43	.78	.95	1.00

13. Christy has 88 photographs to put in her album. If 9 photographs will fit on each page, how many pages will she need?

Ⓐ 8

Ⓑ 9

Ⓒ 10

Ⓓ 11

M023001

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                                   1111111111111122222222
                        12345666678901233344445678901222222
                             abc        ab abc         abcde

M023201 1 B  5 15 34121000025411144111114110002000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023201	*	.66	.37	.78	.95	.99

15.

Puppy's Age	Puppy's Weight
1 month	10 lbs.
2 months	15 lbs.
3 months	19 lbs.
4 months	22 lbs.
5 months	?

John records the weight of his puppy every month in a chart like the one shown above. If the pattern of the puppy's weight gain continues, how many pounds will the puppy weigh at 5 months?

Ⓐ 30

Ⓑ 27

Ⓒ 25

Ⓓ 24

M023201

BASIC GRADE 8 RELEASE

```
                                            DESCRIPTORS

                                         1111111111111111122222222
                              12345666678901233344445678901222222
                                    abc          ab abc          abcde

 M023301 1 B  5 16 32120000052401111153425100013000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023301 *		.73	.41	.88	.99	.99

16. In a bag of marbles, $\frac{1}{2}$ are red, $\frac{1}{4}$ are blue, $\frac{1}{6}$ are green, and $\frac{1}{12}$ are yellow. If a marble is taken from the bag without looking, it is most likely to be

Ⓐ red

Ⓑ blue

Ⓒ green

Ⓓ yellow

M023301

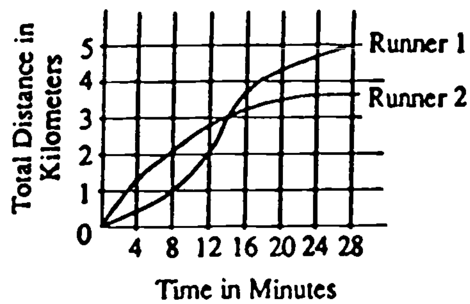
BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                                 1111111111111122222222
                      12345666678901233344445678901222222
                           abc          ab abc        abcde

M023601 1 B  5 19 2111100003440012200000401001300000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023601	*	.65	.35	.76	.91	.94



19. The total distances covered by two runners during the first 28 minutes of a race are shown in the graph above. How long after the start of the race did one runner pass the other?

Ⓐ 3 minutes

Ⓑ 8 minutes

Ⓒ 12 minutes

Ⓓ 14 minutes

Ⓔ 28 minutes

M023601

BASIC GRADE 8 RELEASE

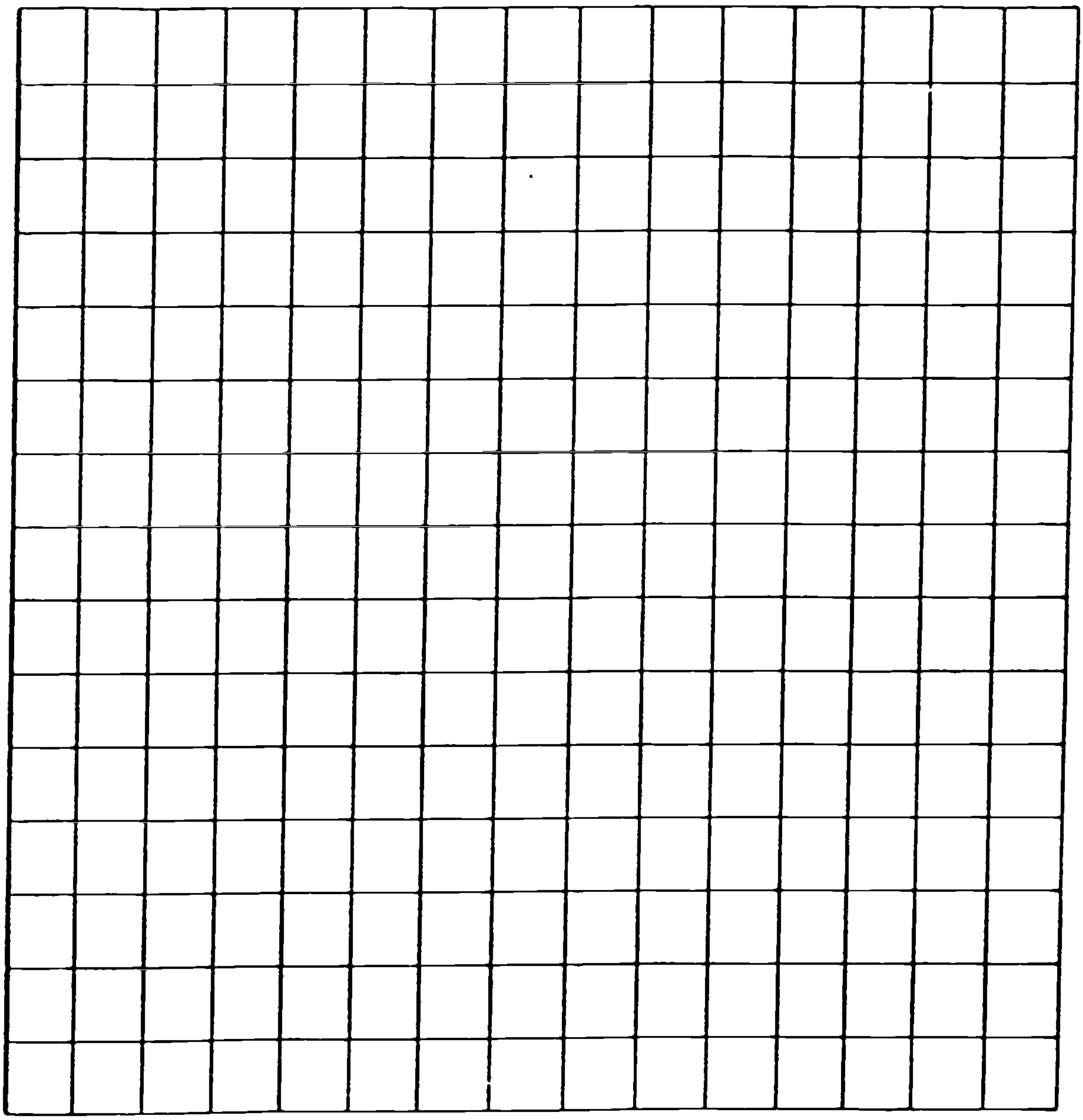
DESCRIPTORS

```
                                      11111111111111122222222
                          12345666678901233344445678901222222
                               abc        ab abc        abcde

 M044601 1 B  7  2 23110656245331011100005100112200010
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M044601	*	.66	.40	.76	.90	.98

2. On the grid below, draw a rectangle with an area of 12 square units.



BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                                  1111111111111122222222
                        123456666789012333444456789012222222
                             abc        ab abc         abcde

M045001 1 B  7  6 $4310000002121101111101520001400000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M045001	*	.64	.31	.73	.97	1.00

6. If $1\frac{1}{3}$ cups of flour are needed for a batch of cookies, how many cups of flour will be needed for 3 batches?

Ⓐ $4\frac{1}{3}$

Ⓑ 4

Ⓒ 3

Ⓓ $2\frac{2}{3}$

M000561

BASIC GRADE 8 RELEASE

```
                                        DESCRIPTORS

                                     11111111111111122222222
                         12345666678901233344445678901222222
                               abc      ab abc         abcde

M045301 1 B  7  9 11120000025411321255256110021100000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M045301	*	.59	.27	.68	.91	.96

9. Steve was asked to pick two marbles from a bag of yellow marbles and blue marbles. One possible result was one yellow marble first and one blue marble second. He wrote this result in the table below. List all of the other possible results that Steve could get.

y stands for one yellow marble.

b stands for one blue marble.

First Marble	Second Marble
y	b

M000553

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                               1111111111111122222222
                     1234566667890123334444567890122222
                         abc       ab abc       abcde

 M053501 1 B 12  1 3214200002022210000000061012220000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M053501	*	.72	.39	.84	.99	1.00

1. If k can be replaced by any number, how many different values can the expression $k + 6$ have?

   Ⓐ None

   Ⓑ One

   Ⓒ Six

   Ⓓ Seven

   Ⓔ Infinitely many

   Did you use the calculator on this question?

   ○ Yes ○ No

W000645

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                                   1111111111111122222222
                        123456666789012333444445678901222222
                             abc         ab abc          abcde

 M048701 1 B 15  7 352110000313221111000040012225254 51
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M048701	*	.61	.38	.69	.84	.82

7. Lynn had only quarters, dimes, and nickels to buy her lunch. She spent all of the money and received no change. Could she have spent $1.98 ?

O Yes O No

Give a reason for your answer.

____

____

____ Q000?C3

BASIC GRADE 8 RELEASE

DESCRIPTORS

```
                              1111111111111122222222
                  1234566667890123334444567890122222
                       abc        ab abc        abcde

M049601 1 B 15 15 0000000003560422221000212111000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M049601	*	.62	.37	.67	.90	.97

15. Harriet, Jim, Roberto, Maria, and Willie are in the same eighth-grade class. One of them is this year's class president. Based on the following information, who is the class president?

    1. The class president was last year's class vice president and lives on Vine Street.
    2. Willie is this year's class vice president.
    3. Jim and Maria live on Cypress Street.
    4. Roberto was not last year's class vice president.

    Ⓐ Harriet

    Ⓑ Jim

    Ⓒ Roberto

    Ⓓ Maria

    Ⓔ Willie

N262701

DESCRIPTORS

```
                         1111111111111122222222
             1234566667890123334444567890122222
                   abc        ab abc        abcde

 M023501 1 P  5 18 2512000001151122225432521000300000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M023501	*	.36	.14	.32	.68	.90

18. From a shipment of 500 batteries, a sample of 25 was selected at random and tested. If 2 batteries in the sample were found to be dead, how many dead batteries would be expected in the entire shipment?

Ⓐ 10

Ⓑ 20

Ⓒ 30

Ⓓ 40

Ⓔ 50

M023501

PROFICIENT GRADE 8 RELEASED

DESCRIPTORS

```
                              1111111111111122222222
                  12345666678901233344445678901222222
                       abc       ab abc        abcde

M053901 1 P 12  5 55620000010361100000003300003000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M053901	*	.40	.19	.37	.69	.85

5. Ken bought a used car for $5,375. He had to pay an additional 15 percent of the purchase price to cover both sales tax and extra fees. Of the following, which is closest to the total amount Ken paid?

Ⓐ $806

Ⓑ $5,510

Ⓒ $5,760

Ⓓ $5,940

Ⓔ $6,180

Did you use the calculator on this question?

○ Yes ○ No

L001230

ADVANCED GRADE 8 RELEASE

DESCRIPTORS

```
                             1111111111111122222222
                 1234566667890123334444567890122222
                      abc      ab abc        abcde

M049101 1 A 15 10 5110000003012010000000510111500000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M049101	*	.22	.07	.16	.47	.83

10. A certain reference file contains approximately one billion facts. About how many millions is that?

Ⓐ 1,000,000

Ⓑ 100,000

Ⓒ 10,000

Ⓓ 1,000

Ⓔ 100

W000638

ADVANCED GRADE 8 RELEASE

DESCRIPTORS

```
                    1111111111111122222222
            123456666789012333444456789012222222
                 abc       ab abc        abcde

M049401 1 A 15 13 450230000145111331000041112030000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M049401	*	.25	.09	.21	.48	.79

A	*B*
2	5
4	9
6	13
8	17
14	?

13. If the pattern shown in the table were continued, what number would appear in the box at the bottom of column *B* next to 14 ?

Ⓐ 19

Ⓑ 21

Ⓒ 23

Ⓓ 25

Ⓔ 29

M000516

Grade 12

```
                                                  11111111111
                        1111111222223333334444445678901234444444
                         abcdef abcd abcde abcde          abcdef

M024101 1 B  5  5 6242260000000000000000000100000000000000

M024801 1 B  5 12 4121110542201000011000010200003000000000

M025001 1 B  5 14 0000000653211001011000010100003000000000

M057501 1 B  7  4 0000000663200000000000000100004002100000

M057601 1 B  7  5 5343410211000000000000000100023030000000

M053901 1 B 12  5 0000000221000000000100000020321400000000

M055201 1 B 14  6 0000000101101000000000000020100401320000

M060901 1 B 15  2 6651110110001100000000000000003010000000
```

```
                                                  11111111111
                        1111111222223333334444445678901234444444
                         abcdef abcd abcde abcde          abcdef

M024401 1 P  5  8 2111110663012112121000100200005000000000

M024701 1 P  5 11 5342311442011111012100221300003234320000

M057701 1 P  7  6 0000000664222011010000002100004200000000

M057801 1 P  7  7 4141210543001010100000000200004100000000

M058001 1 P  7  9 0000000653406533150000001100003212100000

M058101 1 P  7 10 5353410544111000010000000400003110000000

M054001 1 P 12  6 6451121220000000000000000200004024300100

M054401 1 P 12  7 0000000533100000000000000200104123200000

M055101 1 P 14  5 6543010210100000000000000100003010000000

M055301 1 P 14  7 6652120111000000000000000100003010000000

M055601 1 P 14  9 6462132541000000000000000000003011100000

M061701 1 P 15 10 5231011110010010000000001000002010000000
```

```
                                     11111111111
                        1111111222222333334444445678901234444444
                         abcdef abcd abcde abcde         abcdef

M025101 1 A  5 15 0000000332313110022001011301212010000000
M025201 1 A  5 16 6564311331000000001000010300003110000000
M058201 1 A  7 11 6452110330000000000000000100103002100000
M058301 1 A  7 12 1000101100010000006262530200012003110010
M061201 1 A 15  5 0000000221100000000100000030001401000000
M061601 1 A 15  9 6534021221100000000000000200003111010000
```

DESCRIPTORS

```
                                      11111111111
               1111111222223333334444445678901234444444
                abcdef abcd abcde abcde          abcdef

M024101 1 B  5  5 624226000000000000000000000010000000000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M024101	*	.71	.41	.84	.98	1.00

5. Which of the following is NOT a property of every rectangle?

Ⓐ The opposite sides are equal in length.

Ⓑ The opposite sides are parallel.

Ⓒ All angles are equal in measure.

Ⓓ All sides are equal in length.

Ⓔ The diagonals are equal in length.

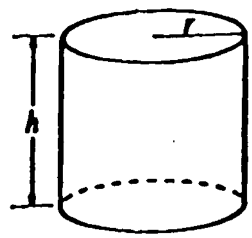
M024101

DESCRIPTORS

```
                   11111111111
111111122222333334444445678901234444444
 abcdef abcd abcde abcde          abcdef
M024801 1 B  5 12 412111054220100001100001020000030000000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M024801	*	.68	.36	.83	.98	.99



12. The volume V of a right circular cylinder like the one in the figure above is given by the formula $V = \pi r^2 h$. In terms of π, what is the volume of a cylinder with radius $r = 4$ and height $h = 10$?

Ⓐ 18π

Ⓑ 26π

Ⓒ 80π

Ⓓ 160π

Ⓔ $1,600\pi$

M024801

DESCRIPTORS

```
                             11111111111
111111122222333334444445678901234444444
 abcdef abcd abcde abcde          abcdef

M025001 1 B  5 14 000000065321100101100001010000300000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M025001	*	.75	.42	.92	1.00	1.00

14. If $x = -4$, the value of $-4x$ is

Ⓐ −16

Ⓑ −8

Ⓒ 8

Ⓓ 16

M025001

DESCRIPTORS

```
                                                11111111111
111111122222333334444445678901234444444
 abcdef abcd abcde abcde        abcdef

M057501 1 B  7  4 000000066320000000000000001000040021000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M057501	*	.67	.35	.82	.96	.99

12B

4. If $n \times n = 729$, what does n equal?

Answer: ________________________ N285321

Did you use the calculator on this question?

○ Yes ○ No

DESCRIPTORS

```
                                    11111111111
1111111222223333334444445678901234444444
 abcdef abcd abcde abcde         abcdef
```

M057601 1 B 7 5 5343410211000000000000000001000230300000000

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M057601	*	.53	.19	.66	.89	.96

5. It takes 64 identical cubes to half fill a rectangular box. If each cube has a volume of 8 cubic centimeters, what is the volume of the box in cubic centimeters?

Ⓐ 1,024

Ⓑ 512

Ⓒ 128

Ⓓ 16

Ⓔ 8

M000566

Did you use the calculator on this question?

○ Yes ○ No

```
                              DESCRIPTORS
                                         11111111111
              111111122222333333444444567890123444444
               abcdef abcd abcde abcde          abcdef

M053901 1 B 12  5 000000022100000000010000002032140000000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M053901	*	.69	.43	.80	.98	1.00

5. Ken bought a used car for $5,375. He had to pay an additional 15 percent of the purchase price to cover both sales tax and extra fees. Of the following, which is closest to the total amount Ken paid?

Ⓐ $806

Ⓑ $5,510

Ⓒ $5,760

Ⓓ $5,940

Ⓔ $6,180

Did you use the calculator on this question?

○ Yes ○ No

L001230

DESCRIPTORS

```
                                                  11111111111
                   1111111222223333334444445678901234444444
                    abcdef abcd abcde abcde          abcdef

M055201 1 B 14 6 00000001011010000000000000201004013200000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M055201	*	.72	.49	.82	.96	1.00

6. Raymond must buy enough paper to print 28 copies of a report that contains 64 sheets of paper. Paper is only available in packages of 500 sheets. How many whole packages of paper will he need to buy to do the printing?

Answer: ______________________________

Did you use the calculator on this question?

○ Yes ○ No

M000476

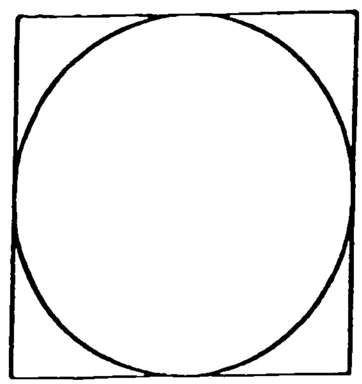
DESCRIPTORS

```
                                   1111111111
1111112222233333444444567890123444444
 abcdef abcd abcde abcde         abcdef

M060901 1 B 15  2 665111011000110000000000000000030100000000
```

BASIC GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M060901	*	.70	.46	.79	.92	.99



2. The length of a side of the square above is 6. What is the length of the radius of the circle?

Ⓐ 2

Ⓑ 3

Ⓒ 4

Ⓓ 6

Ⓔ 8

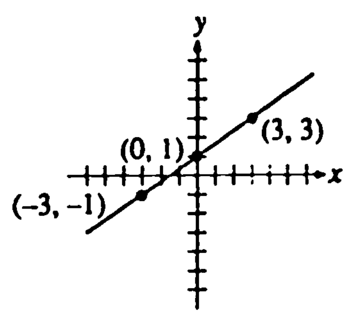
N251701

```
                              DESCRIPTORS
                                             11111111111
                  1111111222223333334444445678901234444444
                   abcdef abcd abcde abcde          abcdef

M024401 1 P  5  8 2111110663012112121000100200005000000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M024401	*	.41	.20	.43	.83	.94



8. What is the slope of the line shown in the graph above?

Ⓐ $\frac{1}{3}$

Ⓑ $\frac{2}{3}$

Ⓒ 1

Ⓓ $\frac{3}{2}$

Ⓔ 3

M024401

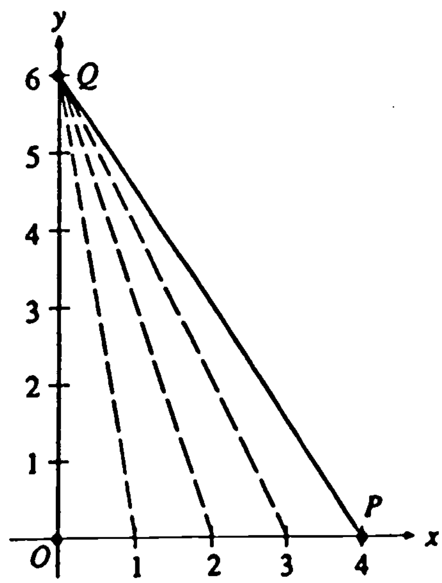
DESCRIPTORS

```
                                                  11111111111
                  1111111222223333334444445678901234444444
                   abcdef abcd abcde abcde          abcdef

M024701 1 P  5 11 5342311442011111012100221300003234320000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M024701	*	.29	.01	.30	.89	.98



11. In the figure above, point Q is fixed and point P starts at 4 and moves left along the x-axis. As P moves left along the x-axis toward O, the area of $\triangle POQ$ changes.

Use the information given to complete the table below to show how the area of $\triangle POQ$ changes as P goes from the position shown to the origin O.

x - coordinate of P	Area of ΔPOQ
4	
3	
2	
1	
0	

M024701

DESCRIPTORS

```
                                                   11111111111
                   1111111222223333334444445678901234444444
                    abcdef abcd abcde abcde          abcdef

M057701 1 P  7  6 00000006642220110100000002100004200000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M057701	*	.34	.09	.36	.82	.84

6. For what value of x is $8^{12} = 16^{x}$?

- Ⓐ 3
- Ⓑ 4
- Ⓒ 8
- Ⓓ 9
- Ⓔ 12

M000484

Did you use the calculator on this question?

○ Yes ○ No

DESCRIPTORS

```
                                         11111111111
             1111111222223333334444445678901234444444
              abcdef abcd abcde abcde          abcdef

M057801 1 P  7  7 41412105430010101000000002000041000000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M057801	*	.32	.20	.27	.72	.97

7. What is the distance between the points (2, 10) and (−4, 2) in the xy-plane?

Ⓐ 6

Ⓑ 8

Ⓒ 10

Ⓓ 14

Ⓔ 18

Y002308

Did you use the calculator on this question?

◯ Yes ◯ No

DESCRIPTORS

```
                                 11111111111
111111122222333333444444567890123444444
 abcdef abcd abcde abcde           abcdef

M058001 1 P  7  9 000000065340653315000000110000321210000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M058001	*	.39	.07	.46	.86	.94

10. In the xy-plane, a line parallel to the x-axis intersects the y-axis at the point (0, 4). This line also intersects a circle in two points. The circle has a radius of 5 and its center is at the origin. What are the coordinates of the two points of intersection?

Ⓐ (1, 2) and (2, 1)

Ⓑ (2, 1) and (2, −1)

Ⓒ (3, 4) and (3, −4)

Ⓓ (3, 4) and (−3, 4)

Ⓔ (5, 0) and (−5, 0)

Y002313

Did you use the calculator on this question?

○ Yes ○ No

DESCRIPTORS

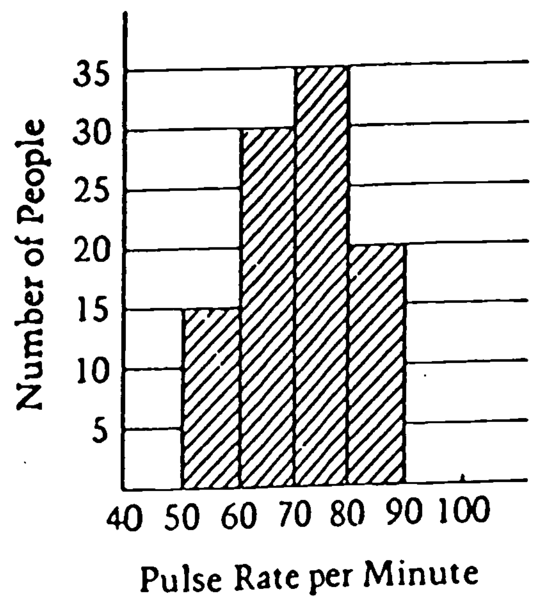
```
                                                11111111111
       1111111222223333334444445678901234444444
       abcdef abcd abcde abcde            abcdef

M058101 1 P  7 10 5353410544111000010000000400003110000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M058101	*	.32	.14	.28	.80	.95

PULSE RATE FOR 100 PEOPLE



12. The pulse rate for a group of 100 people is shown in the graph above. What is the average pulse rate per minute for these 100 people? (Note: Use the midpoint of each interval to represent the pulse rate for the entire interval. For example, 55 would be used for the pulse rate of the 15 people in the 50-60 group.)

Answer: ______________________ Y002434

Did you use the calculator on this question?

○ Yes ○ No

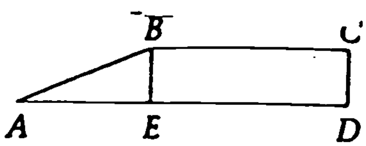
DESCRIPTORS

```
                                      11111111111
          1111111222223333334444445678901234444444
           abcdef abcd abcde abcde          abcdef

M054001 1 P 12  6 6451121220000000000000000200004024300100
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M054001	*	.23	.03	.21	.75	.95



6. The area of rectangle *BCDE* shown above is 60 square inches. If the length of *AE* is 10 inches and the length of *ED* is 15 inches, what is the area of trapezoid *ABCD*, in square inches ?

Answer: ______________________________

Did you use the calculator on this question?

○ Yes ○ No

W000646

```
                                   DESCRIPTORS
                                                          11111111111
                    1111111222223333334444445678901234444444
                     abcdef abcd abcde abcde           abcdef

M054401 1 P 12  7 00000005331000000000000000200104123200000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M054401	*	.26	.02	.27	.75	.86

$$\sqrt{8}\,N = 3^5$$

7. In the equation above, what is the value of N, rounded to the nearest tenth?

Answer: ______________________________

Did you use the calculator on this question?

○ Yes   ○ No

W000648

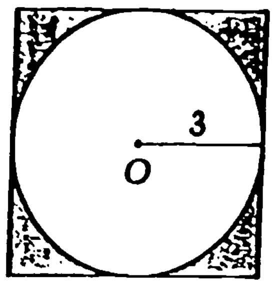
DESCRIPTORS

```
                                  11111111111
           1111111222223333334444445678901234444444
            abcdef abcd abcde abcde          abcdef

M055101 1 P 14  5 6543010210100000000000000100003010000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M055101	*	.37	.21	.36	.75	.93



5. In the figure above, a circle with center O and radius of length 3 is inscribed in a square. What is the area of the shaded region?

- Ⓐ 3.86
- Ⓑ 7.73
- Ⓒ 28.27
- Ⓓ 32.86
- Ⓔ 36.00

Did you use the calculator on this question?

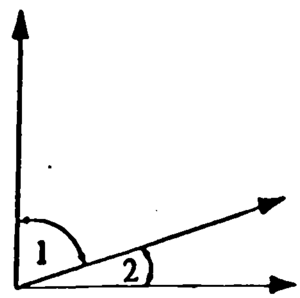
○ Yes ○ No

M000483

```
                                   DESCRIPTORS
                                              11111111111
                        1111111222223333334444445678901234444444
                         abcdef abcd abcde abcde          abcdef

M055301 1 P 14   7 6652120111000000000000000100003010000000
```

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M055301	*	.44	.30	.42	.80	.89



7. The sum of the measures of angles 1 and 2 in the figure above is 90°. What is the measure of the angle formed by the bisectors of these two angles?

Ⓐ 60°

Ⓑ 45°

Ⓒ 30°

Ⓓ 20°

Ⓔ 15°

Did you use the calculator on this question?

○ Yes ○ No

W000641

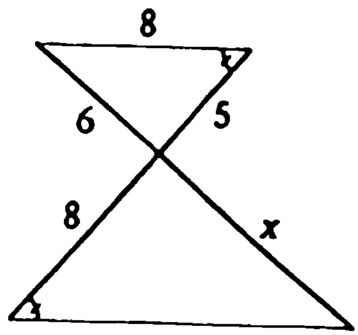
DESCRIPTORS

```
                                  11111111111
1111111222223333334444445678901234444444
 abcdef abcd abcde abcde          abcdef
```

M055601 1 P 14 9 6462132541000000000000000000000003011100000

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M055601	*	.24	.03	.21	.73	.94



9. In the figure above, the two triangles are similar. What is the value of x ?

Answer: ______________________

Did you use the calculator on this question?

O Yes   O No

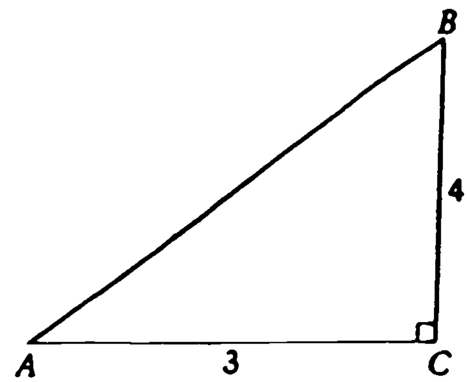
W000650

DESCRIPTORS

```
                                       11111111111
        1111111222223333334444445678901234444444
         abcdef abcd abcde abcde          abcdef
M061701 1 P 15 10 5231011111001001000000001000002010000000
```

PROFICIENT GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M061701	*	.30	.09	.31	.72	.93



10. In right triangle ABC above, $\cos A =$

Ⓐ $\frac{3}{5}$

Ⓑ $\frac{3}{4}$

Ⓒ $\frac{4}{5}$

Ⓓ $\frac{4}{3}$

Ⓔ $\frac{5}{3}$

M000497

DESCRIPTORS

```
                                              11111111111
1111111222223333334444445678901234444444
 abcdef abcd abcde abcde          abcdef
```

M025101 1 A 5 15 0000000332313110022001011301212010000000

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M025101	*	.31	.25	.29	.44	.77

15. It takes 28 minutes for a certain bacteria population to double. If there are 5,241,763 bacteria in this population at 1:00 p.m., which of the following is closest to the number of bacteria in millions at 2:30 p.m. on the same day?

Ⓐ 80

Ⓑ 40

Ⓒ 20

Ⓓ 15

Ⓔ 10

M025101

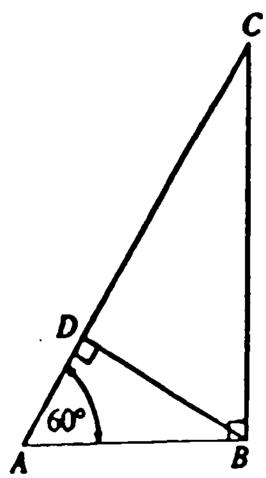
DESCRIPTORS

```
                                                  11111111111
          111111122222333333444444567890123444444
           abcdef abcd abcde abcde            abcdef

M025201 1 A  5 16 6564311331000000000100001030000311000000000
```

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M025201	*	.22	.15	.19	.43	.79



16. In $\triangle ABC$ shown above, $AC = 12$. What is the length of segment BD?

Ⓐ $3\sqrt{2}$

Ⓑ $3\sqrt{3}$

Ⓒ 6

Ⓓ $6\sqrt{2}$

Ⓔ $6\sqrt{3}$

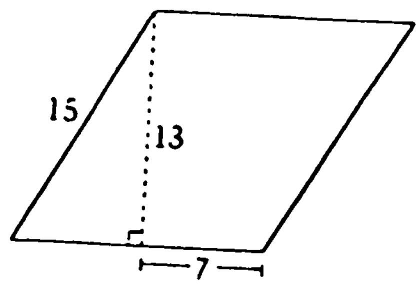
M025201

DESCRIPTORS

```
                                                    11111111111
              11111112222233333344444456789012344444444
               abcdef abcd abcde abcde           abcdef
M058201 1 A  7 11 6452110330000000000000000100103002100000
```

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M058201	*	.08	.00	.02	.33	.74



11. To the nearest whole number, what is the area of the parallelogram above?

Answer: ______________________

M0004??

Did you use the calculator on this question?

○ Yes ○ No

DESCRIPTORS

```
                                   11111111111
          1111111222223333334444445678901234444444
           abcdef abcd abcde abcde          abcdef

M058301 1 A  7 12 1000101100010000006262530200012003110010
```

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M058301	*	.09	.00	.04	.35	.67

9. If $f(x) = 4x^2 - 7x + 5.7$, what is the value of $f(3.5)$?

L000834

Answer: ______________________

Did you use the calculator on this question?

○ Yes ○ No

Do not use.

9. ⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

DESCRIPTORS

```
                                            11111111111
         111111122222333333444444567890123444444
          abcdef abcd abcde abcde           abcdef

M061201 1 A 15  5 000000022110000000100000030001401000000 0
```

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M061201	*	.25	.20	.21	.44	.83

5. In a group of 1,200 adults, there are 300 vegetarians. What is the ratio of nonvegetarians to vegetarians in the group?

Ⓐ 1 to 3

Ⓑ 1 to 4

Ⓒ 3 to 1

Ⓓ 4 to 1

Ⓔ 4 to 3

M000668

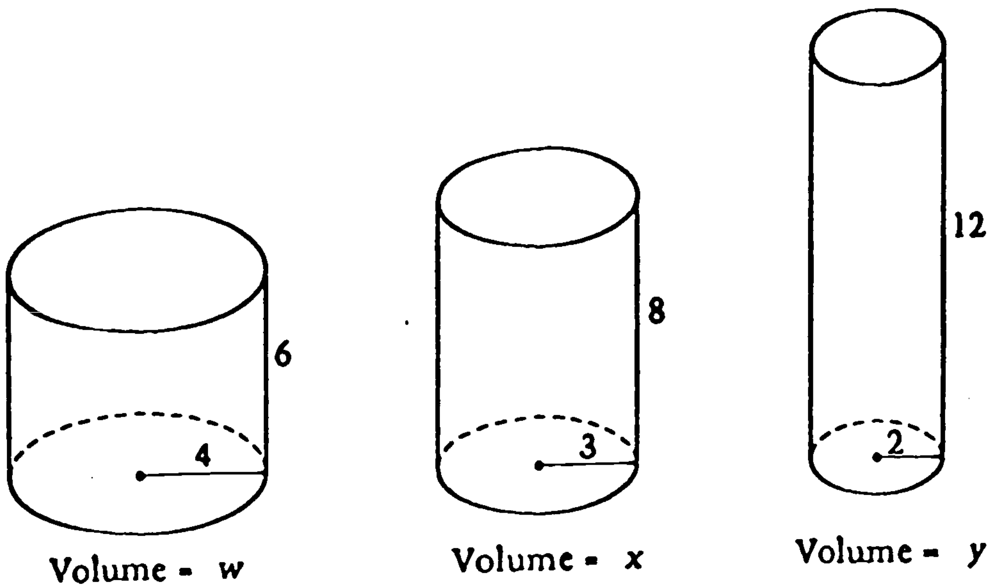
DESCRIPTORS

```
                                                11111111111
          1111111222223333334444445678901234444444
           abcdef abcd abcde abcde          abcdef

M061601 1 A 15  9 65340212211000000000000000200003111010000
```

ADVANCED GRADE 12 RELEASED

NAEPID	RELEASE	PPLUS	PPLUS1	PPLUS2	PPLUS3	PPLUS4
M061601	*	.30	.33	.25	.34	.78



9. In the figures above, the radius and height of each right circular cylinder are given. If w, x, and y represent the respective volumes of the cylinders, which of the following statements is true?

Ⓐ $y = w = x$

Ⓑ $y < x < w$

Ⓒ $y < w < x$

Ⓓ $w < y < x$

Ⓔ $w < x < y$

M000495

NOTICE

REPRODUCTION BASIS

- [ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [x] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").