Reading as an Access to Mathematics Problem Solving on Multiple-Choice Tests for Sixth-Grade Students

ROBERT HELWIG
MARICK A. ROZEK-TEDESCO
GERALD TINDAL
BILL HEATH
University of Oregon

PATRICIA J. ALMOND
Oregon Department of Education

ABSTRACT **The effect of providing middle school students with a video accommodation for a standardized mathematics test was examined. Two hundred forty-seven students were asked to solve 60 word problems. One half of the questions were presented in standard form, while the other half were read by an actor on a video monitor. Students were grouped according to mathematics and reading ability. A test accommodation effect was found for students possessing below-average mathematics skills. The problems were identified as having relatively high reading difficulty according to word count, number of verbs, and word familiarity. Students with above-average mathematics proficiency but low reading skill performed better when the questions were presented in video format. This accommodation may be useful on specific test items for students with certain reading deficiencies.**

Most states are currently conducting large-scale assessments for documenting school, teacher, and student accountability. At the same time, the Individuals with Disabilities Education Act (IDEA; 1997) is clear in its mandate to include all students with disabilities in state testing programs. In both efforts, the focus is shifting from the process to the outcomes of schooling. In the rhetoric of reform, schools no longer graduate students with minimum competencies; rather, they ensure that students meet the high standards reflective of skills needed to compete in a global economy.

This shift is causing the policy and practice of testing to change, with two important consequences. First, educators must accommodate students with disabilities to allow them equal access to our education programs, including both instruction and assessment. On a practical level, such accommodations usually appear in state policies in the form of administration and scoring prescriptions–admonitions or lists of allowable practices. Second, educators must base interpretation of student performance needs on a theoretical approach to ensure that accommodations do not violate the construct validity of measurement systems. Certain alterations in testing conditions, and decisions made from them, must be based on accurate and relevant information and not unduly influenced by extraneous factors. In the view of Messick (1989), the former issue relates to construct representation and the latter term to the presence of irrelevant factors.

The purpose of the current study was to approach the issue of accommodations from a practical perspective and to explain the outcomes of one specific accommodation from a theoretical perspective. Specifically, we compared middle school students' achievement on a traditional administration of a standardized mathematics test with an oral presentation of a comparable version. In addition, we developed criteria to identify the type of word problems that would present students with the most difficulty. Although the empirical relationships among different skills may help local and state education agencies develop policies for accommodating students with disabilities, an understanding of the construct being measured is more fundamental. In this study, we focused largely on reading and mathematics because of the major dissimilarity of the symbol system underlying them. In reading, the process of encoding and decoding relies on generalizations; in mathematics, the process is based on inviolate rules. Also in reading, literacy is shrouded amorphously in an asymmetric skill sequence of sounding out graphemes, knowing vocabulary word meaning, and placing information in context. In mathematics, "numeracy" is established by moving through a linear skill sequence that is hierarchically arranged with preskills and algorithms. By investigating accommodations in mathematics while controlling critical aspects of reading, we oriented the study to core issues regarding construct validity: Do multiple-choice test measures of mathematics reflect the necessary and

Address correspondence to Robert Helwig, Behavioral Research and Teaching, 5262 University of Oregon, Eugene, OR 97403–5262. (E-mail: bobh@darkwing.uoregon.edu)

essential features of mathematics proficiency without including irrelevant features? To address that question, we first examined the construct of reading and then investigated its relationship with mathematics problem solving.

Reading

Reading is often described as the interaction of two distinct processes: decoding and comprehension (Stothard & Hulme, 1996). For skilled readers, decoding is a highly automated task, thus releasing attention to be directed at comprehension (LaBerge & Samuels, 1974). Good comprehension is unlikely in the absence of reading fluency, although the reverse is not necessarily true (Yuill & Oakhill, 1991). Beginning in the intermediate grades, and occurring increasingly in junior and senior high school, the emphasis in general education classrooms switches from developmental reading to studying and learning from expository texts, which creates difficulties for low-ability readers. The purpose of reading changes to learning and remembering key concepts, integrating new learning with prior knowledge, and applying knowledge to problem solving (MacArthur & Haynes, 1995). A reader must initially understand material at a factual level, and then be able to interpret what is read before critical reading can occur. When readers can focus on the meaning of words rather than on the recognition of words, they can think and learn about the content itself rather than about reading the material (Richardson & Morgan, 1990). *Literacy* (reading ability) can be found only in the presence of both decoding and comprehension; both skills are necessary, but neither is sufficient (Gough, Hoover, & Peterson, 1996). To avoid confusion over our meaning, we used the term *reading* as a general label to refer to a global construct. When we reported the process by which students decode mathematics word problems, we used the terms *decoding* or *word identification*.

One important measure of literacy is *reading fluency* (correct words read per unit of time). Individuals who possess low reading fluency rates are plagued by more than just longer completion times. Several studies have revealed that slower readers tend to have lower comprehension and make fewer inferences from written materials than do students reading at higher rates (Marston, 1989; Yuill & Oakhill, 1991). Low reading rates also have been linked to problems with recall (Zabrucky & Ratner, 1992) as well as with miscues (Parker, Hasbrouck, & Tindal, 1992).

Another important component of literacy is the role played by memory. Because of competition for short-term memory space, students who process words slowly may be severely limited in their ability to integrate both the decoding and comprehension of written words. Slow processing results in less capacity for other cognitive processes necessary for understanding. Also, the rapid decay of short-term memory results in slowly processed information being less available when it is needed (Yuill & Oakhill, 1991). Some children read narrative so slowly and laboriously that before they come to the words at the end of the sentence, they have forgotten those at the beginning (Barney, 1972).

Theoretical Explanations of Reading and Mathematics Problem-Solving Relationships

Similar, yet distinct, models of reading comprehension illustrate the effect of passage complexity on cognitive processing (Miller & Kintsch, 1980; Turner, Britton, Andraessen, & McCutchen, 1996). According to those models, text comprehension is based on a two-phase series of inputs and reductions. As text is input into working memory, mental networks are constructed involving connections between overlapping propositions. Because of the finite capacity of working memory, a reduction phase follows involving the selective extraction of relevant information, images, and cognitive links. Those items become the starting point of the next cycle as new text is input. As text length increases, comprehension decreases for at least two reasons. First, memory capacity limits what can be stored. Second, because of this limitation, connected networks formed in earlier cycles are weakened as critical links are discarded to make room for new input. Thus, as more text is read, proportionally less information is available for processing; at the same time, relationships between relevant data are weakened or eliminated so that associations critical for comprehension are diminished or nonexistent.

For pure computation problems, reading is not an important access skill. Mathematics word problems, on the other hand, present low-ability readers with serious challenges. The primacy of reading in relationship to mathematics problem solving is evidenced by its placement at the top of various hierarchies of mathematics problem-solving steps (Clarkson, 1983). Empirical evidence of the problematic relationship between reading and mathematics is found in research indicating that inaccurate reading may be responsible for a significant number of errors found on mathematics tests. For example, Clements (1980) investigated the type of errors that seventh-grade low-achieving students made on word problems and found that 8% of 1,400 errors were *reading related* (defined as failures to decode a word resulting in an inability on the part of the student to comprehend the question). Others also have found that reading errors significantly contribute to lower performance on mathematics tests (Clarkson, 1983; Newman, 1977).

Problem solving also is difficult for low-ability readers because the vocabulary of mathematics includes words that are specific to the components and processes of mathematics as well as words used in the natural language (O'Mara, 1981). Some words found in mathematics texts have several meanings, are not part of a person's general vocabulary, may generally convey a process, or are defined in a specific way (Lees, 1976). That variety of usage and context magnifies the already difficult problem of an increase in the number of students possessing low basic reading skills (Dolgin, 1977).

Mayer (1987) identified four components of mathematics problem solving: translation, integration, solution planning, and execution. The first two components are heavily dependent on reading skill. *Problem translation* uses linguistic knowledge to interpret each statement for formulation into an internal representation. *Integration* involves combining and synthesizing each statement into a coherent depiction of the problem. That process includes distinguishing between relevant and irrelevant information as well as the appropriate choice of schema. That reliance on reading puts students possessing low comprehension and recall skills at a distinct disadvantage. The result is that substantial numbers of students become underachievers (Dolgin, 1977; Noonan, 1990). Teachers frequently contend that many students do not read well enough to succeed in mathematics (Earp & Tanner, 1980) even though significant numbers of those students possess average and above skills in mathematics problem solving (O'Mara, 1981).

Several factors appear to be related to the difficulty of written passages in general, and specifically those related to mathematics. Because readability concerns "the ease with which a passage . . . [can] be read" (Aaron & Joshi, 1992, p. 218), it is central to any discussion of factors that affect low-ability readers. Readability formulas designed to identify the reading level of a written passage tend to rely on a limited number of factors. For example, four commonly used formulas (Dale–Chall, Flesch Reading Ease, Farr–Jenkins–Patterson, and Fry Readability Graph) each use only two criteria, one of which concerns sentence length. The remaining criteria used are number of syllables per 100 words, frequency of one-syllable words, and word familiarity. Two formulas designed for short passages of the type often associated with mathematics word problems (Short Passage Readability Formula [Fry, 1990], the Homan–Hewitt Readability Formula) use some combination of words per sentence, clauses per sentence, word length, and word familiarity.

At least one study indicates that students may benefit when the readability level of word problems is reduced. Thompson (1967) administered two versions of the same mathematics problem-solving task to sixth-grade students. One version was written at a 2.7-grade level; the other version was written at an 8.7-grade level. Students performed significantly better on the lower level version. In addition to this work in readability, researchers have linked a variety of variables specifically to student success rates on mathematics word problems. In a study involving mathematics students from Grades 4 through college, Jerman and Mirman (1974) found that the total number of characters, syllables, words, and sentences were all related to problem difficulty levels, as were word and sentence length. Other researchers have confirmed the above findings for several of the variables, including word count (Loftus & Suppes, 1972); sentence length (Cohen & Stover, 1981; Cook, 1973); and total characters (Segalla, 1973).

Several studies also have indicated that knowledge of mathematics vocabulary is correlated with mathematics achievement (Aiken, 1972). Skrypa (1979) found that instruction in mathematics vocabulary was followed by improved problem solving. In addition, student success rate with word problems has increased when vocabulary was simplified (Linville, 1969). Much of this mathematics–vocabulary relationship may be unrelated to reading skill but rather is based on other factors such as concept knowledge.

To summarize, various features associated with written passages have been shown to affect student problem-solving achievement. For many students, that effect may be small or nonexistent (Paul, Nibbelink, & Hoover, 1986). On the other hand, because of problems associated with comprehension, recall, and miscues, low-ability readers may be at a serious disadvantage when asked to rely on their reading skills to demonstrate their proficiency with mathematics. If tests are to measure what they are intended to measure, a distinction must be drawn between a child's knowledge base and ability to respond to a test (Grise, Beattie, & Algozzine, 1982). From that perspective, the need for access to test accommodations by qualified students is clear.

Test Accommodations

The term accommodation has varied meanings and is used often interchangeably with other terms such as modification (Thurlow, Ysseldyke, & Silverstein, 1993). For this article, we used a broad definition of *accommodation* as an alteration in test presentation or response methods to provide students with equal access to demonstrate what they know. Accommodations generally have been targeted for students with disabilities. For example, the IDEA (1977) states that "children with disabilities . . . [be] included in general State and district-wide assessment programs, with appropriate accommodations, where necessary" (p. 41). In discussions of accommodations, most writers address the need for alterations in testing practices for students with physical, mental, or learning disability (Phillips, 1994; Shriner, Ysseldyke, Thurlow, & Honetschlager, 1994; Thurlow et al., 1993). Recently, however, discussions have been held regarding the inclusion of more students as qualified for accommodation. As of 1995, five states had implemented provisions for making accommodations for students other than those holding individualized education plans (Thurlow, Scott, & Ysseldyke, 1995).

A variety of accommodations that are presently in use include Braille, audiotapes, large print, oral responses, extended time, and interpretation of directions, among many others (Yell & Shriner, 1996). As early as the 1970s, several states had, or were developing, accommodation procedures (McCarthy, 1980). However, the amount of research conducted on accommodations has been minimal (Thurlow, Ysseldyke, & Silverstein, 1995).

One obvious choice of accommodation for low readers on tests of mathematics proficiency is to have questions read aloud. Although that testing format does not eliminate students' need to solve written mathematics problems in other situations, it allows for measurement in an important area of mathematics without the confounding factor of reading ability. In addition, it enables students to solve comparable problems. Beyond the point of word recognition, listening and reading require essentially the same processes. The lexicon, grammar, and linear format of the two processes are almost identical. In addition, the same background knowledge is brought to bear on both printed and spoken words (Gough et al., 1996). Thus, if one desires information on students' levels of mathematics achievement one can give them the opportunity to respond to verbal prompts instead of written ones, with only a limited likely loss of validity.

A literature search, however, revealed few studies that examined specifically the effects of oral presentation of mathematics problem-solving tasks as an accommodation for standardized mathematics tests. Data collection in this area has been limited primarily to investigating reliability and factor structures of standard and nonstandard large-scale test administrations to establish construct validity. For example, Bennett, Rock, and Kaplan (1987) compared data from students taking a standard version (N = 299) and an audiocassette accommodation of an SAT examination and found no broad classes of items that operated differently between the two groups. We found no studies that attempted to determine the degree to which an oral reading accommodation affected performance.

When students respond to standardized test items incorrectly, "it is unknown whether they do so because of lack of knowledge or because of inability to successfully comprehend the test and its items" (Homan, Hewitt, & Linder, 1994, p. 349). Mathematics tests that largely reflect written mathematics in their verbal context (as opposed to symbolic) are in part measuring reading ability rather than pure mathematics ability (Tack, 1995). It is, therefore, imperative that tests and testing conditions be constructed so that low readers are free to demonstrate their problem-solving capabilities.

Method

Participants

This study took place in 15 sixth-grade mathematics classrooms within three school districts and four middle schools in western Oregon. Students in all schools were predominantly Caucasian; most came from low to middle socioeconomic backgrounds. The 15 classrooms were distributed among 10 teachers; 5 of the teachers contributed two classes each, and the remaining teachers contributed one class. Four of the classes were conducted in resource rooms for students who had been identified as needing assistance in mathematics. Those classes ranged in size from 7 to 12 students. The 11 general education classrooms contained from 18 to 34 students. The number of students involved in the study was 325. Two hundred forty-seven students completed all portions of the study. Approximately 85% of those were Caucasian. Twelve percent were identified as special education students. Slightly more than half of the participants were male (54%). All teachers were contacted before the study began and volunteered to participate. Total years of teaching experience ranged from 2 to 20 (M = 10.6).

Procedure

Our study was completed during a 1-month period. Four measures of student achievement were administered to each student.

Oral reading fluency (ORF). As a measure of reading proficiency, each student was administered individually a 1-min oral reading fluency (ORF) that was scored by counting the number of correct words read per minute. The reading passage was taken from a sixth-grade science book. As students read, a trained test administrator followed along and noted reading errors on a scoring sheet. Credit was given for self-corrected errors. Students were told at the start of the test that if they hesitated more than 2 s on any word, they would be told what the word was and to go on. ORFs have been used by teachers and specialists for a wide variety of purposes including screening and determining eligibility for special programs and instructional groupings, setting instructional goals, and monitoring academic progress (Hasbrouck & Tindal, 1992). All ORFs were administered by trained personnel.

Basic mathematics skill. A 21-question basic mathematics skill test was group-administered in each participating classroom. Nineteen items were computational problems involving addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals. The remaining two items were one-step word problems, each containing approximately 20 words. All questions were part of the selected districts' fourth- or fifth-grade general education mathematics curriculum.

Standardized Mathematics Achievement Test. As part of a larger research project, we obtained unsecured 1996 versions of standardized fifth- and sixth-grade mathematics tests developed and used by a southwestern state. We considered the tests to be representative of the type of questions that mathematics students around the country are required to answer as part of statewide assessment procedures. Forty-nine questions from the sixth-grade test and 11 questions from the fifth-grade version were combined to produce 60 questions used in the present study. The questions covered a variety of areas such as number concepts, mathematics relationships, geometry, estimation, statistics, and measurement. The test contained no straight computational problems.

Students were required to read a problem-posing situation and choose from four or five possible answers. Many problems contained some type of chart, table, or diagram (n = 16). The number of words (excluding numerals, symbols, and abbreviations) in the problem statements ranged from 5 to 61 (M = 24.3). The multiple-choice answers to some problems were numbers, whereas others were presented as sentences or phrases. Of the 11 problems in the latter category, the total number of words of all choices ranged from 4 to 65 (M = 31.5). The total number of words in both problem statement and answer choices ranged from 7 to 101 (M = 31.6).

All students answered each of the 60 items during two testing sessions of 30 questions each. The sessions occurred 1 to 4 days apart in the students' regular mathematics classroom. One half was presented in standard test booklet format. The other half was presented via a video monitor. The composition of the two halves varied by classroom so that the 30 questions answered in the standard administration by 1 student may have been different from those answered by a student in another classroom. The tests were counterbalanced so that roughly half of the students took the standard version first. That version was presented in a test booklet that the students completed at their own pace. Students read the questions to themselves and responded by circling the appropriate answer in the booklet. No calculators were allowed for either version. All necessary calculations were relatively simple. In many cases, no computations were required. Test duration was limited to 45 min, but all students were able to finish in the allotted time. Students worked at their own pace, completing questions in any order, and were free to make changes in their responses.

When students took the video version of the test, they were given a test booklet that was identical in format to the standard administration booklet with the exception that the questions were different. Students were told that the words of each test question would be displayed on a video monitor positioned in front of the classroom. As the words were displayed, they also were read aloud over the monitor's speakers. The narrator reading the questions was not visible on the monitor. Participants were told that they could watch the words on the screen as they were read aloud or follow along in their test booklets where the words were also printed. For questions on which the multiple-choice answers consisted only of numbers, only the problem statement was read. The narrator then instructed the students to "mark your answers now." For questions that required words to be read for the answers, the narrator read each choice aloud. The amount of time students had to work on each problem after it was read and before the next problem was started was predetermined. The time spans ranged from 15 to 60 s, depending on item difficulty. Only occasionally did a student not mark an answer in the allotted time. Students were instructed that they were free to go back at any time and finish, rework, or change any previous response. Some students were observed doing that. Because students had the written test booklets in front of them, they could work ahead of the video pacing or ignore the video. That was discouraged by the proctors, but some students were observed working on problems in advance of the video. Other students may have chosen to proceed at the pace of the video but read each question and response for themselves. An estimate of the number of students choosing that method could not be determined because there was no outward behavioral difference between that group and those who listened to the reader.

Independent Variables

We chose from our literature review those passage attributes that were applicable to our particular needs. Specifically, we selected factors that would differentially affect low-ability readers given that the reading passages we were concerned with contained relatively few words. For each of the 60 items on the test, values for the following variables were determined.

Words. We determined the total number of words (WORDS) by including all words in the problem statement as well as all answer choices. Numerals, symbols, and abbreviations were not counted.

Syllables. Syllables (SYLL) were counted in each word included in the above category.

Multisyllable words. Multisyllable words (MSYLL) included all words containing two or more syllables. In addition, words of eight letters or more were designated as long words (LONG). Proper nouns were not included in either of these categories.

Word familiarity. Words that may have been unfamiliar to significant numbers of students taking the test were designated as difficult words (DIFF). We determined difficulty levels by comparing each word on the test with a word familiarity list contained in The Living Word Vocabulary (Dale & O'Rourke, 1976). Over 43,000 words are contained in that volume. Each entry includes a grade level and percentage of students correctly identifying the meaning of the word. Scores were obtained from a national sample of students from Grades 4 through college. Word familiarity data from this source are used in various readability formulas (Fry, 1990; Homan et al., 1994). Difficult words were designated as those that were familiar to less than 90% of sixth-grade students.

Mathematics vocabulary. Determining the count of mathematics vocabulary words was difficult because of the unique properties of many of the words. Whereas some words have different meanings inside and outside of the mathematics classroom, others have similar meanings in both arenas. Clearly, the word *inch* is more easily defined by sixth-grade students than the word *mean* partly because inch is normally introduced at a much earlier age and also because it has no specialized meaning.

Therefore, we designated a word as mathematics vocabulary (M-VOCAB) if it met two criteria. First, it had to be used in a context that had a specialized mathematics or science meaning; second, it had to be unfamiliar to more than 10% of fourth-grade students. We purposely chose a lower grade level than the students we would be working with because we believed that this level would be more reflective of the abilities of the low readers who we were targeting.

Thus, *divide* did not qualify (meets criteria one, but fails criteria two), nor did *occupied* (fails criteria one, but meets criteria two). Words such as *probability, random, radius, polygon,* and *mean* qualified by meeting both requirements.

Passage complexity. The complexity of the passages was measured by two methods. The first method involved counting the number of Hunt's (1970) T-units. *T-units* (T-U) are complete expressions of thought that can stand alone in grammatically correct fashion. They frequently have been used as measures of syntactic complexity (Reed, Burton, & Vandett, 1988). Verbs also have been used to determine the complexity of sentences (Jerman & Mirman, 1974; Scarborough, 1990). We counted the total number of verbs (VERB) that occurred in each passage.

Although each of the above variables is associated with readability, we believed that because of the relatively short lengths of the passages we were working with, many of the variables might not have a linear relationship with student success. For example, even slow readers may be able to navigate through passages of 5, 10, or 20 words and still have time to solve a problem. Difficulties may not arise until much longer passages are encountered. Similarly, one or two unfamiliar words may pose few problems as students use context to derive the meaning of a problem. Comprehension may be affected only when students do not recognize larger numbers of words. More likely, those factors will work in combination with each other.

Because one purpose of our research was to ascertain under what circumstances accommodations are effective, we developed standards to identify which test questions would present the greatest challenge for low readers. We used three criteria for that purpose. First, the test items had to come from those containing 40 or more words (n = 13). Second, the passage had to contain five or more verbs (n = 16). Finally, each had to contain at least three difficult words (familiar to less than 90% of sixth-grade students). Of the total of 60 test items, six (10%) met all three criteria. Those items are presented in the Appendix. We hypothesized that on those questions, low readers would respond positively to a video accommodation.

Subject Groupings

All portions of the study were completed by 247 students. Means for the ORF and mathematics ability test were 99.2 (SD = 29.5) and 49.3 (SD = 12.8), respectively. Because we were interested in the interaction between reading and mathematics skills and their combined effects on mathematics achievement, we subdivided the study population. Using the Basic Mathematics Skill test (M = 49.3, SD = 12.8) as a measure of computation and basic skill, we first divided the students into high mathematics and low mathematics groups. Students scoring at or above the mean (n = 149) were placed in the former group. Those scoring below the mean (n = 98) were placed in the latter group.

The groups were divided further on the basis of their reading skill. We have been involved in ORF research for several years (see Hasbrouck & Tindal, 1992). On the basis of that work, we are certain that sixth-grade students with ORF scores below 90 need special reading attention. We labeled those students as low readers. Of the 149 high mathematics students, 35 students were classified in the low reader category. For comparison, from the remaining 114 students, we identified 35 medium readers (ORF = 101–113) and 35 high readers (ORF above 124). Because a disproportionate number of students in the low mathematics group were also low readers (n = 59), the other two groups (medium readers and high readers) contained insufficient numbers of students to do meaningful statistical analysis. Therefore, data analysis was performed on only four subgroups: low reader/high mathematics; medium reader/high mathematics; high reader/high mathematics; and low reader/low mathematics.

Data Analysis

In our attempt to investigate the effects of word identification on mathematics achievement, we conducted three separate analyses. First, to examine the overall effectiveness of the accommodation, we used paired t tests to compare the performance of each student on the 30-problem standard administration and the 30-problem video presentation of the mathematics achievement test. Means for both absolute and adjusted scores were analyzed. Absolute scores consisted of each student's total number of correctly answered problems on a particular version. A score adjustment was calculated because the assignment of an item to the standard or video version of the test differed by classroom. Instead of assigning 1 point for each problem solved, with the score adjustment, a point value was assigned to each problem according to difficulty. Difficult problems were assigned higher values. The easiest problem was given a base value of 1. The point value for any given problem was calculated by determining the relative numbers of students correctly solving that problem. For example, if Problem A were solved correctly by half as many student as Problem B, it would be worth twice as many points. Each student's adjusted score was calculated by summing the point value of each correctly solved problem.

We performed a second analysis to determine the relationship between each of the independent variables and the testing accommodation. The differences in success rates between the standard and video versions of the test on each item were calculated. The 60 values were then correlated with the eight passage attributes.

Finally, we investigated the effect of the video accommodation specifically on the six complex test items by comparing the success rates of those students receiving a standard administration to those receiving a video presentation. The success rates were compared for each item using a two-proportion significance test (Moore & McCabe, 1993). Each of the above analyses was conducted for the entire study population as well as for each appropriate subgroup.

Results

Results from the three data analyses are presented below. Our first result involves the examination of the global effect of the accommodation. Table 1 contains the difference in mean scores for the standard and video versions of the mathematics achievement test. The data compare students' performance on the two administrations of the test, including both absolute and adjusted scores. Students taking the video version of the test scored slightly higher (absolute score) than those taking the standard version (18.16 vs. 17.75), although that difference was not statistically significant. Of the subgroups examined, only the low mathematics group showed a preference that reached significance (14.69 vs. 13.85 in favor of the video). When adjusted scores were examined, the low mathematics students again were the only group whose difference reached significance, although both the total population and the low mathematics/low ORF groups approached that level.

All differences were relatively small. The largest absolute difference in mean scores for the two versions for any group was less than 1 point (.88, low mathematics/low ORF). That result accounted for an approximately 7% increase in performance from standard to video. The total population difference of .41 (not significant) accounted for only about a 2% increase. Adjusted scores differences were of similar proportion.

Table 2 illustrates the relationship between the eight independent variables and the testing accommodation. There appears to be little or no association between how many words, syllables, long words, or other language variables are present in a given test item and the difference in success rate on the standard or video version of the test. That lack of relationship held across subgroups. Only analyses involving verbs resulted in any significant correlations. As the number of verbs present in a passage increased, the difference in success rate in favor of the video accommodation tended to increase for both the low ORF group in general and the low ORF/high mathematics group. The low ORF/low mathematics group did not experience the same relationship. Although the correlations appeared to be noticeably higher for the low ORF/high mathematics group, none of the other correlations reached statistical significance. The overwhelming predominance of positive correlations, although nonsignificant, indicated that a small effect was likely present for some of the passage variables.

When viewed individually in isolation, language variables appear to possess limited value for predicting differential performance; therefore, we performed a final analysis to determine if an additive effect of those variables might produce different results. The six complex test items were used as a pool from which to examine differential performance on the standard and video presentations. Differences in percentage correct between the two testing modes are presented in Table 3. As hypothesized, problems that contained an accumulation of challenging language factors tended to be more easily solved when read aloud as opposed to requiring students to read the problems. An examination of the data for all participants indicated that five of the six problems were solved correctly a higher proportion of the time by those students having the problem read aloud. Three differences were significant. Two of the items favored a video presentation, whereas the remaining item favored a

Table 1.—Means, Standard Deviations, and Differences of Student Scores on Standard and Video Versions of Mathematics Achievement Test

		Absolute score						Adjusted score					
		Video		Standard				Video		Standard			
Group	*n*	*M*	*SD*	*M*	*SD*	Diff.	*p*	*M*	*SD*	*M*	*SD*	Diff.	*p*
All	247	18.16	6.03	17.75	6.07	.40	.12	27.56	9.45	26.84	9.68	.72	.08
High math	149	20.44	5.07	20.32	5.07	.11	.74	31.10	8.16	30.76	8.47	.34	.52
Low math	98	14.69	5.74	13.85	5.35	.85	.03	22.18	8.74	20.87	8.27	1.31	.05
Low ORF	94	14.72	5.92	14.34	6.05	.38	.39	22.27	9.03	21.57	9.31	.70	.32
High ORF	45	22.51	4.75	22.31	4.76	.20	.67	34.51	7.62	34.53	8.41	-.02	.98
Low math/ ORF > 100	33	16.67	4.85	16.33	4.58	.33	.65	24.94	7.36	24.54	7.05	.40	.73
Low math/ low ORF	59	13.46	6.01	12.58	5.44	.88	.06	20.48	9.23	19.07	8.49	1.41	.10
High math/ low ORF	35	16.86	5.15	17.31	5.94	-.46	.61	25.31	7.92	25.78	9.22	-.48	.71
High math/ medium ORF	35	19.29	3.99	19.06	3.70	.23	.71	29.50	6.76	28.36	5.86	1.14	.22
High math/ high ORF	35	23.91	3.79	23.83	3.43	.09	.87	36.75	6.32	37.07	6.55	-.33	.70

Note. ORF = oral reading fluency.

Table 2.—Correlations Between Difference in Video and Standard Administration of 60 Items and Independent Variables

Variable	All (n = 247)	High math (n = 149)	Low math (n = 98)	Low ORF (n = 94)	High ORF (n = 45)	ORF > 100 Low math (n = 33)	Low ORF Low math (n = 59)	Low ORF High math (n = 35)	Med. ORF High math (n = 35)	High ORF High math (n = 35)
Words	.12	.10	.06	.17	.08	.08	.02	.23	-.09	.10
Syllable	.15	.13	.09	.22	.05	.07	.07	.24	-.03	.07
M-syllable	.17	.14	.15	.18	.12	.06	.14	.13	-.01	.11
Long	.11	.12	.09	.11	.01	.08	.08	.12	.06	.05
Diff-words	.10	.16	-.06	.13	-.03	-.06	.02	.23	-.04	.01
M-Vocab	.04	.06	-.03	.00	-.09	-.02	.05	.04	.19	-.10
T-U	.06	.10	.00	.13	.03	.02	.05	.17	.03	.09
Verb	.17	.15	.07	.26*	.01	.00	.09	.29*	.00	.06

Note. Positive correlations indicate difference in favor of video presentation. ORF = oral reading fluency. Words = all words in the problem statement and all answer choices; syllable = all syllables in all words, as above; M-syllable = multisyllable words; Long = words of eight or more letters; Diff-words = difficult words; M-Vocab = mathematics vocabulary; T-U = t units — complete expressions of thought.
*$p < .05$.

Table 3.—Differences in Percentage Correct Between Standard and Video Versions of Mathematics Test for Complex Items

Item	All (n = 247)	High math (n = 149)	Low math (n = 98)	Low ORF Low math (n = 59)	Low ORF High math (n = 35)	Med. ORF High math (n = 35)	High ORF High math (n = 35)
1	7	3	8	11	25	2	4
5	4	–4	10	9	25	–26*	0
15	–12*	–11	–6	–2	–25	–6	–6
35	19**	22**	13	3	42**	12	7
47	11*	17*	3	6	30*	20	9
57	5	9	–1	–6	28*	–16	–5

Note. Positive numbers indicate differences in favor of video presentation. ORF = oral reading fluency.
*$p < .05$. **$p < .01$.

standard presentation. Examination of the high and low mathematics groups separately revealed that the majority of the differences could be attributed to the high mathematics group for which two of the items remain significant. No item produced a significant difference for the low mathematics group.

As expected, within the high mathematics group the differences were generally much higher for low ORF students than for more skilled decoders. The smallest difference within that group was 25%. Three of the differences were significant in favor of the video presentation. The other three items approached significance ($p < .07$). Two of the differences favored a read-aloud presentation. Item 5 within the medium ORF group produced a 26% difference in favor of the standard presentation ($p < .05$). All differences within the high ORF group were nonsignificant.

In summary, the greatest effect of the test accommodation was found for selected test items and certain population subgroups. Comparisons between the 30-item standard and video versions of the mathematics achievement test generally revealed relatively small differences. Those differences reached statistical significance only in the case of the low mathematics group. The correlations between language variables and differences in success rates between the standard and video presentation were generally positive, with the exception of VERBS, nonsignificant. On the other hand, items that contained large numbers of words, verbs, and unfamiliar vocabulary resulted in more impressive differential performance. Students who combined low reading fluency with above-average performance on the mathematics skills test experienced notable improvements in performance when the selected items were read aloud.

Discussion

According to the standards of the National Council of Teachers of Mathematics (NCTM; 1993), "Assessment should promote equity by giving each student optimal opportunities to demonstrate mathematics power. . . ." (p. 27). The aggregation of students for test-reporting purposes can give misleading information in relationship to that desire, particularly when little regard is given to test-item differences. Our results show that comparing only score totals from the standard and accommodated administrations of the achievement test provides an incomplete description of the interaction between test taker and item. In general, the main effect of testing mode demonstrated limited consequence. Examining low mathematics students separately, however, revealed a significant preference for a video presentation. In addition, among above-average mathematics students, only those experiencing difficulties in word identification performed better when the six complex reading items were read aloud.

The results are consistent with our knowledge of the interaction between reading and content area performance. "An inadequate grasp of the language of instruction is a major source of underachievement in school" (Cuevas, 1984, p. 134). Students who cannot efficiently decode word problems are at a distinct disadvantage in comprehending and solving those problems. That fact has been acknowledged by mathematics authors and textbook publishers who for many years have used a variety of strategies in an attempt to keep readability levels low (Noonan, 1990).

As our results indicate, however, accommodations are unnecessary for the majority of students. Only low mathematics students exhibited a preference for having problems read aloud. In addition, the majority of students did not favor a video presentation on any of the six identified complex test items. The last finding is consistent with Phillips's (1994) list of considerations for appropriate testing accommodations. She recommended that measurement specialists ask themselves whether nontargeted examinees would benefit if allowed the same accommodation. An answer of no, which is the case in our study, is one piece of evidence supporting construct validity. Students able to decode at average and above rates evidently gain no advantage from having test items read aloud.

That finding does not appear to be the case for many low fluency readers. The problems faced by those students when taking written mathematics tests are comparable to those encountered by some students taking tests in a second language. In both instances, students have difficulty understanding the media through which testing information is presented. Low-fluency readers are generally at a disadvantage because of time and comprehension concerns; second-language students are asked to navigate in a language in which many have limited experience. A video accommodation allows low-fluency students to avoid nonrelevant information input limitations in a manner similar to second-language students taking tests in their native language with the opportunity to avoid comprehension problems during text translation. Thus, with an appropriate accommodation, students are more likely to perform at levels comparable with their mathematics abilities. Studies by Adetula (1989) and Cuevas (1984) support that conclusion for second-language students just as our data supports it for students with poor decoding skills.

Yet, our study and those of second-language investigators differ in substantial ways. Foremost is the importance of the interaction between participant group and item type. The accommodation was not effective for all low-fluency participants, nor for all test items. The logic, supported by empirical data, linking low fluency with low comprehension holds for all students with low ORF scores; yet only those low ORF students with above-average performance on the mathematics skills test were aided by the accommodation. Those in the low mathematics skills groups showed no difference in performance on the standard and video presentation for any of the six selected test items. We conclude that although reading is a critical access skill to mathematics problem solving, the low mathematics skills group lacked the mathematics proficiencies necessary to take advantage of the accommodation. In other words, an oral presentation of the test items may have allowed those students to decode the written text, but their low mathematics skills prevented them from solving the problems correctly.

It is also noteworthy that we did not find a significant accommodation effect for low ORF students when we examined all 60 items. That is undoubtedly due in part to the fact that several problems contained relatively few words. Given sufficient time to reread and solve a problem, there is no reason to assume that low-fluency students would be at a significant disadvantage by having to decode a short passage. Beyond problems containing one or two sentences, however, previous investigations of the correlation between various passage variables and item difficulty (Cohen & Stover, 1981; Cook, 1973; Jerman & Mirman, 1974; Loftus & Suppes, 1972; Segalla, 1973) have suggested that word count, syllables, difficult vocabulary, and so forth would be important to our investigation. Viewed individually, we found little evidence of that recommendation. Our examination of those variables, however, was from a somewhat different perspective. Specifically, we investigated how the factors differentially affect performance on video and standard testing situations. As the text comprehension models discussed earlier (Miller & Kintsch, 1980; Turner, et al., 1996) illustrate, lengthy word problems may be difficult to solve for reasons other than the difficulty in decoding long strings of words. Longer passages, whether written or spoken, are simply more difficult to process.

Similarly, some sixth-grade students may experience difficulty when reading the words *dimensions*, *symmetrical*, or *reflections*. However, the problem may lie in their unfamiliarity with the words rather than in difficulties with decoding (Devine, 1989). Neither the process of longer sentences' becoming more complex nor students' unfamiliarity with certain words would be significantly affected by the mode of test presentation. Students who have difficulty following the complex logic of a multisentence word problem in a test booklet may encounter the same problems when hearing the item read aloud. Therefore, the correlations of previous studies between language variables and item difficulty may have measured problem complexity as well as readability.

Only when several of those factors are combined, as they were in the six complex test items, is performance affected. Even a very long passage, containing only relatively easy words and a simple sentence structure, may pose little problem for low fluency students. When vocabulary becomes more difficult and sentences become longer and more complex, however, some students may experience difficulty with cognitive processing. The models of text comprehension discussed earlier suggest that complex problems will tax memory capacity limits to a greater extent than simpler problems. In addition, networks of information and relationships necessary for adequate comprehension will be diminished to a greater extent during the input of complex test items.

It is likely that students plagued by slow word identification are disproportionately affected by complex word problems. First, because of their slower fluency rate, they have less short-term memory capacity than more rapid decoders (Yuill & Oakhill, 1991). Second, because of this ineffective use of memory space, more information has to be deleted during the reduction phase of the cycle, thus weakening cognitive networks. The accommodation implemented during the current study would clearly aid low-fluency students in the effective use of memory and provide increased cognitive networking efficiency.

Aggravating the issues of comprehension are the sheer number and difficulty of the words in the complex problems that make decoding and word substitution errors more likely. In some instances, students simply do not have the skill to decode a word. In others, students mistakenly replace words such as *spaces* for *shapes*, *meaning* for *missing*, and *volume* for *value*, thus changing the meaning of sentences (Clarkson, 1983). When replacements such as these are made, students may take up valuable short-term memory space with incorrect information at the expense of accurate data and form faulty network links with other information. In extreme instances, a student may end up answering a completely different question than that intended by the test makers. Clarkson found that those students who believed they were obtaining correct answers, but in fact were not, had a tendency to make errors in reading. That finding suggests that reading errors are in large part responsible for what Clarkson termed misplaced confidence.

When questions are posed verbally, every word can be decoded. The more complex items chosen for closer inspection all contained several sentences and many total words. Each offered multiple opportunities for low ORF students to encounter words beyond their ability to decode. An oral presentation via a video monitor, on the other hand, allowed those students the same problem-solving opportunities as more able decoders. It is likely, therefore, that some of the students were able to solve problems on the basis of their mathematics ability, which otherwise would have been inaccessible to them because of their limitations with reading. We have thus taken the work of Thompson (1967) one step further. Instead of merely reducing the readability level of the problems, we controlled critical aspects of reading during the accommodation in order to test mathematics proficiency.

An additional reason that we believe specific students performed better during the video presentation for the six complex items is that long and hard-to-read passages tend to ensure that low ORF students have insufficient opportunity to integrate the decoding and comprehension of passages as well as to solve the problems in the allotted time (Barney, 1972; Yuill & Oakhill, 1991). Questions were read on the video at approximately the pace of the high ORF group. Thus, on the standard administration, slow decoders may have taken over twice as long to decode some passages. Although all students completed the standard format test within the allowable time, some may have felt rushed or responded to peer pressure to hurry to complete it.

It is reasonable to assume that a video presentation would essentially eliminate decoding, word substitution, and time problems, as well as bolster low-fluency readers' comprehension through more effective use of memory and processing. From the perspective of Mayer's (1987) four factors of problem solving, the accommodation aided students in translation and integration, which allowed them to concentrate cognitive processes on solution planning and execution.

High-fluency readers, on the other hand, would experience few of the above-mentioned benefits of an accommodation, even on complex problems. They would gain little in terms of short-term memory capacity by having items read to them. Some advantage might be gained by eliminating incorrect (as opposed to slow) decoding, but the results of this study indicate that for fluent readers that effect is minimal. Because the majority of our students possessed adequate fluency rates, the global effect of the accommodation was limited.

Explaining why low mathematics students were the only subgroup performing significantly better on the complete 60-item test when problems were read aloud is problematic. A preliminary analysis showed an interaction between mathematics and word identification skills. Dividing low mathematics students into poor and competent decoding groups reveals that only the low ORF group approached significance. That data agrees with our previous discussion; that is, students who can decode with sufficient skill do not benefit from an accommodation. Although that reasoning was supported by the data from the 6 complex items, it is an incomplete explanation when all 60-items are considered. Previously, when we examined only the 6 complex items, we argued that low ORF-low mathematics students did not benefit from the accommodation because they did not possess the mathematics skill necessary to take advantage of the intervention. Apparently, however, they did possess the mathematics skill to benefit from the accommodation over the entire 60 problems. Other than the complex reading required for the 6 problems, we know of no other differences between those items and the remaining 54 items. An item analysis of correct solution rates revealed that the 6 items were not particularly difficult. In addition, the 6 items require a representative sample of skills including algebra, statistics, geometry, and estimation. It appears, however, that some factors of the larger test combined to allow lower skill mathematics students to take advantage of a reading accommodation. Further research is required to determine what those factors are.

One surprising result of our study was the 25-point difference in success rate for Item 15 in favor of a standard administration for the low ORF-high mathematics group. That difference approached significance ($p < .07$). Several factors may have accounted for the anomaly—one being chance. Another explanation is the placement of the correct answer choice. The majority of the total number of words in that problem are contained in the four responses. The first of those responses happened to be the correct choice. Students marking their booklets after reading only this choice would have avoided reading almost half of the total words. Another explanation involves the use of a diagram with this problem. In studies with sixth graders, Cohen and Stover (1981) found that the presence of a diagram significantly increased student success rate on word problems. It is possible that the diagram helped to clarify information that poor decoders might have had trouble deriving only from the printed page, thus negating some of the effect of the video. Although a diagram or answer placement may have mitigated against the effect of the video, they do not explain why students would perform better during a standard format. More data are needed to answer that question, although saying words is only one of the processes involved in solving word problems. Other processes include comprehension, transformation of written words into mathematics models, use of processing skills, encoding of answers, and motivation (Clarkson, 1983). All of those areas might cause problems for students of any reading ability and may interact in different ways, depending on presentation format.

Conclusion

Our research indicates that in some instances the label *mathematics achievement test* may be a misnomer. We have shown that low ORF–high mathematics students tend to be more successful when solving complex items via a video presentation. High ORF students did not respond to that accommodation. That lack of response is strong evidence that part of what is being tested on written tests are students' word identification skills. The significance of our results expanded because some students, by working ahead on their own, did not realize the full effect of the video presentation. The result of combining a significant amount of reading with mathematics problem solving is "a test that is not a valid indicator of mathematics skills" (Tack, 1995, p. 4). Much of our interest in testing lies in our desires to increase the validity of standardized achievement tests. Some researchers wonder "whether it is possible to make testing modifications that remove irrelevant sources of difficulty but still measure the same construct" (Thurlow et al., 1995, p. 264). We believe we have taken an important step in answering that question.

Much work remains to be done in the area of mathematics testing accommodations, however. We have identified only a few items in which certain students improved their performance through a change in test presentation format. More sensitive item identification procedures need to be developed. Possibly of more importance, however, are issues regarding construct validity. We believe we have addressed Phillips's (1994) concern that only the targeted population benefit from an accommodation. Still to be addressed, however, are issues of task and skill comparability. We have based our preliminary conclusions regarding those issues on the work of Gough et al. (1996) who, while admitting that reading and listening involve slightly different processes, concluded that "these differences pale in comparison to the similarities of processing in the two modalities" (p. 2). However, this is a complicated topic that we recommend as the subject for future research.

Inseparable from issues of construct validity is the role that reading should play in mathematics problem solving. Clearly, language is a central element in all school curriculum, including mathematics. What role written communication should play is uncertain. Problem solving is unquestionably one of the primary goals of mathematics education. Most students, outside of a school setting, face mathematics-related problems in written form. It is therefore reasonable that students should be taught and evaluated on their ability to solve written word problems. It is misleading, however, to integrate reading and mathematics into one assessment instrument while at the same time labeling it only as mathematics achievement with no acknowledgment of the reading that must occur. Because one of the goals of testing is to provide teachers with information about student progress, tests that offer only limited information in this area are of questionable use. Reading achievement tests are numerous and relatively easy to administer. Combining the reading measures with test accommodations such as the one described in this study will provide educators with information about both mathematics and reading achievement, instead of requiring teachers to guess the area in which difficulties exist.

Another conclusion we can draw from our work is that blanket accommodations may not be practical or necessary, even for selected groups of students. In the present study, we showed that a video accommodation had little or no effect on student performance for large numbers of the items on the test. Therefore, it may not be practical to make entire tests available for video presentation when effects will be shown for only a relatively few items. School districts and departments of education are more qualified than we are to make that determination.

Finally, it is imperative that researchers conduct follow-up studies to our research to help identify (a) types of word problems most suitable to video accommodations, (b) similar types of accommodations that are most effective, and (c) students who would benefit from changes in test format. Answers to those questions can give educators a deeper understanding of test validity so that "the manner in which behavioral instances are combined to produce a score [will] rest on knowledge of how the processes underlying those behaviors combine dynamically to produce effects" (Messick, 1995, p. 7).

REFERENCES

Aaron, P. G., & Joshi, R. M. (1992). *Reading problems: Consultation and remediation*. New York: Guilford.

Adetula, L. O. (1989). Solutions of simple word problems by Nigerian children: Language and schooling factors. *Journal for Research in Mathematics Education, 20,* 489–497.

Aiken, L. R. (1972). Language factors in learning mathematics. *Review of Educational Research, 42,* 359–385.

Barney, L. (1972). Problems associated with the reading of arithmetic. *The Arithmetic Teacher, 19,* 131–133.

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1987). SAT differential item performance for nine handicapped groups. *Journal of Educational Measurement, 24,* 41–55.

Clarkson, P. (1983). Types of errors made by Papua New Guinean students. *Educational Studies in Mathematics, 14,* 355–367.

Clements, M. A. (1980). Analyzing children's errors on written mathematics tasks. *Educational Studies in Mathematics, 11,* 1–21.

Cohen, S. A., & Stover, G. (1981). Effects of teaching sixth-grade students to modify format variables of mathematics word problems. *Reading Research Quarterly, 16,* 175–199.

Cook, B. (1973, February). *An analysis of arithmetic, linguistic, and algebraic structural variables that contribute to problem solving difficulty in algebra word problems.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 076 433)

Cuevas, G. J. (1984). Mathematics learning in English as a second language. *Journal for Research in Mathematics Education, 15,* 134–44.

Dale, E., & O'Rourke, J. (1976). *The living word dictionary*. Elgin, IL: Field Enterprises Educational Corporation.

Devine, T. G. (1989). *Teaching reading in the elementary school: From theory to practice*. Boston: Allyn & Bacon.

Dolgin, A. B. (1977). Improvement of mathematics learning through reading instruction. *The High School Journal, 61,* 59–69.

Earp, N. W., & Tanner, F. W. (1980). Mathematics and language. *Arithmetic Teacher, 28*(4), 32–34.

Fry, E. (1990). A readability formula for short passages. *Journal of Reading, 33,* 594–597.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Carnoldi & J. Oakhill. (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahwah, NJ: Erlbaum.

Grise, P., Beattie, S., & Algozzine, B. (1982). Assessment of minimum competency in fifth grade learning disabled students: Test modifications make a difference. *The Journal of Educational Research, 76,* 35–40.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24*(3), 41–44.

Homan, S., Hewitt, M., & Linder, J. (1994). The development and validation of a formula for measuring single-sentence test item readability. *Journal of Educational Measurement, 31,* 349–358.

Hunt, K. (1970). Syntactic maturity in school children and adults. *Monograph of the Society for Research in Child Development, 35* (1, Serial No. 134).

Individuals with Disabilities Education Act of 1997, 20 U.S.C. § 1400 et. seq.

Jerman, M. E., & Mirman, S. (1974). Linguistic and computational variables in problem solving in elementary mathematics. *Educational Studies in Mathematics, 5,* 317–362.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Lees, F. (1976). Mathematics and reading. *Journal of Reading, 19,* 621–626.

Linville, W. J. (1969). The effects of syntax and vocabulary upon the difficulty of verbal arithmetic problems with fourth grade students (Doctoral dissertation, Indiana University, 1969). *Dissertation Abstracts International, 30,* 4310A.

Loftus, E. F., & Suppes, P. (1972). Structural variables that determine problem-solving difficulty in computer-assisted instruction. *Journal of Educational Psychology, 6,* 531–542.

MacArthur, C. A., & Haynes, J. B. (1995). Student assistant for learning from text (SALT): A hypermedia learning aid. *Journal of Learning Disabilities, 28,* 150–159.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special education* (pp. 18–78). New York: Guilford.

Mayer, R. E. (1987). *Educational psychology: A cognitive approach.* Boston: Little, Brown.

McCarthy, M. M. (1980). Minimum competency testing and handicapped students. *Exceptional Children, 47,* 166–173.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–105). New York: Macmillan.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practices, 14*(4), 5–8.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 335–354.

Moore, D. S., & McCabe, G. P. (1993). *Introduction to the practice of statistics*. New York: Freeman.

National Council of Teachers of Mathematics (1993). *Assessment standards for school mathematics*. Reston, VA: Author.

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematics tasks. In M. A. Clements & J. Foyster (Eds.), *Research in mathematics education in Australia* (Vol. 2, pp. 239–258). Melbourne: Swineburne.

Noonan, J. (1990). Readability problems presented by mathematics text. *Early Child Development and Care, 54,* 57–81.

O'Mara, D. (1981). The process of reading mathematics. *Journal of Reading, 25,* 22–30.

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). Greater validity for oral reading fluency: Can miscues help? *Journal of Special Education, 25,* 492–503.

Paul, D. J., Nibbelink, W. H., & Hoover, H. D. (1986). The effects of adjusting readability on the difficulty of mathematics story problems. *Journal for Research in Mathematics Education, 17,* 163–171.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7*(2), 93–120.

Reed, W. M., Burton, J. K., & Vandett, N. M. (1988). Daly and Miller's writing apprehension test and Hunt's T-unit analysis: Two measurement precautions in writing research. *Journal of Research and Development in Education, 21*(2), 1–8.

Richardson, J. S., & Morgan, R. F. (1990). *Reading to learn in the content areas.* Belmont, CA: Wadsworth.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11,* 1–22.

Segalla, A. (1973, February). *Using structural variables to predict word-problem solving.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 076 431)

Shriner, J. G., Ysseldyke, J. E., Thurlow, M. L., & Honetschlager, D. (1994). "All" means "all": Including students with disabilities. *Educational Leadership, 51*(6), 38–42.

Skrypa, A. (1979). *Effects of mathematics vocabulary training on problem solving abilities of third and fourth graders.* Unpublished doctoral dissertation, Rutgers University, The State University of New Jersey. (ERIC Document Reproduction Service No. ED 172 169)

Stothard, S. E., & Hulme, C. (1996). A comparison of reading comprehension and decoding difficulties in children. In C. Carnoldi & J. Oakhill, (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 93–112). Mahwah, NJ: Erlbaum.

Tack, K. D. (1995). *The influence of reading skills on the Ohio Ninth Grade Proficiency Test of Mathematics.* (ERIC Document Reproduction Service No. ED 386 703)

Thompson, E. N. (1967). Readability and accessory remarks: Factors in problem solving in arithmetic. (Doctoral dissertation, Stanford University, 1967). *Dissertation Abstracts International, 28,* 2464A.

Thurlow, M. L., Scott, D. C., & Ysseldyke, J. E. (1995). *A compilation of states' guidelines for accommodations in assessments for students with disabilities* (Synthesis Report 18). Minneapolis, MN: National Center on Educational Outcomes. (ERIC Document Reproduction Service No. ED 385 058)

Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature* (Synthesis Report 4). Minneapolis, MN: National Center on Educational Outcomes. (ERIC Document Reproduction Service No. ED 358 656)

Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1995). Testing accom-

modations for students with disabilities. *Remedial and Special Education, 16*, 260–270.
Turner, A., Britton, B. K., Andraessen, P., & McCutchen, D. (1996). A predication semantics model of test comprehension and recall. In B. K. Britton & A. C. Graesser. (Eds.), *Models of understanding test* (pp. 33–71). Mahwah, NJ: Erlbaum.
Yell, M. L., & Shriner, J. G. (1996). Inclusive education: Legal and policy implications. *Preventing School Failure, 40*, 101–108.
Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, England: Cambridge University Press.
Zabrucky, K., & Ratner, H. H. (1992). Effects of passage type on comprehension monitoring and recall in good and poor readers. *Journal of Reading Behavior, 24*, 373–391.

APPENDIX
Test Questions Presenting Greatest Challenge for Low Readers

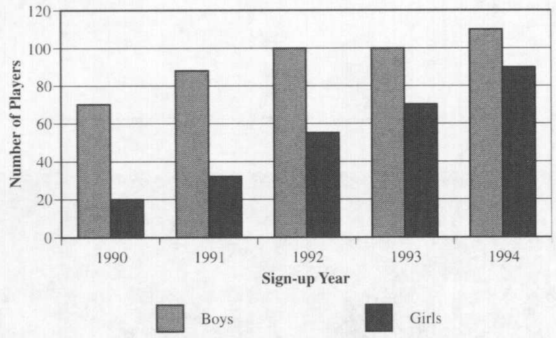
Complex problems meeting three selection criteria:

a. Word count greater than or equal to 10
b. Verb count greater than or equal to 5
c. Difficult words greater than or equal to 3

1. Alex has asked 1 of his friends to feed his 3 dogs while he is away on vacation. Each dog eats 11/2 bags of dog food per week. What else do you need to know to find how many bags of food Alex should leave for his dogs to eat while he is away on vacation?

a. The size of each of Alex's dogs
b. The number of pounds of dog food in one bag
c. The number of weeks Alex will be away on vacation
d. The amount of exercise the dogs need each day
e. The price of each bag of dog food

5. The following graph shows the number of boys and girls who signed up to play in the Ridgeview Little League:



Which is a reasonable conclusion that can be drawn from the information on the graph?

a. More girls than boys signed up for Little League in 1993.
b. The ratio of girls to boys in Little league increased every year after 1990.
c. The number of boys who joined Little League decreased every year.
d. Girls did not want to play Little League baseball before 1990.
e. There were more Little League sign-ups in 1992 than any other year.

15. The table shows the results of a cost comparison for three types of jeans at different stores.

Jean Type	Western Store	Department Store	Discount Store
X	$17.95	$25.95	$15.95
Y	$19.95	$18.98	$17.50
Z	$22.50	$20.95	$21.98

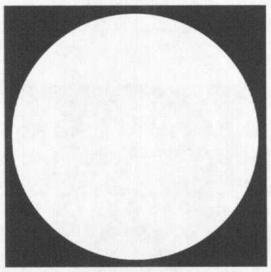
Which of the following statements is true?

a. Type X jeans cost the most at the department store.
b. The discount store has the best price for any of the three types of jeans.
c. Type Y jeans cost the most at the department store.
d. The least expensive place to get Type Z jeans is at the discount store.

35. Mr. Reston asked his mathematics students if they believe that every student in the school could stand in his classroom at the same time. His 22 students crowded into a corner of the room and found that they occupied 18 square feet of the 520 square feet available in the room. What other piece of information would allow the students to answer the questions?

a. The average weight of students in the school
b. The number of classrooms in the school
c. The average height of the students in the school
d. The number of students in the school
e. The number of teachers in the school

47. The figure shows a circle in the interior of a square.



Which method would you use to find the area of the shaded region?

a. Find the circumference of the circle and subtract the area of the square.
b. Find the perimeter of the square and subtract the circumference of the circle.
c. Find the area of the square and subtract the area of the circle.
d. Find the area of the square and add the circumference of the circle.
e. Find the area of the circle and add the area of the square.

57. Harold went to the mall on Monday. He spent 1/2 hr shopping for jeans, 40 min eating lunch, 1/4 hr looking at sweaters, and 10 min playing with the puppies at the pet shop. Which is a reasonable estimate of how much time he spent at the mall?

a. Less than 1 hr
b. Between 1 hr and 2 hr
c. Between 2 hr and 3 hr
d. Between 3 hr and 4 hr