FOURTH EDITION

Technical Report

Forms S and T

Walter H. MacGinitie
Ruth K. MacGinitie
Katherine Maria
Lois G. Dreyer

RIVERSIDE PUBLISHING
a subsidiary of Houghton Mifflin Harcourt

GATES-MACGINITIE READING TESTS (GMRT)
Technical Report

What you will find in the manuals

In the *Directions for Administration*
Information about the *Gates-MacGinitie Reading Tests* series
How to choose appropriate test levels
What is in the tests
What to do before testing
How to give the tests
What to do if answer sheets will be scored by the Riverside Scoring Service®

In the *Directions for Online Administration*
Information about the *Gates-MacGinitie Reading Tests* series
How to choose appropriate test levels
What is in the tests
How to prepare for online testing
How to monitor the testing

In the *Manual for Scoring and Interpretation*
How to hand score the tests
How to use the tables of norms
What the scores mean
Tables of 1999 and 2006 norms

In *Linking Testing to Teaching: A Classroom Resource for Reading Assessment and Instruction*
How to use the scores as part of a comprehensive assessment of reading
How to use the scores to guide instruction

In the *Technical Report*
How the tests were developed
How the tests were standardized
Statistical information about the tests

In the *Technical Report Supplement*
How the tests were renormed in 2005–2006
Statistical information about the renorming
How the online tests were developed
Statistical information about the online tests

7 8 9 10—RRD—12 11 10

TECHNICAL REPORT

Contents

TEST DEVELOPMENT

Test Design

Development of the Fourth Edition of the *Gates-MacGinitie Reading Tests® (GMRT®)* was guided by a detailed description, or blueprint, specifying the test content and desired difficulty for each level. The blueprint for the Fourth Edition is similar to the one for the Third Edition. However, some changes were made for the Fourth Edition, particularly in the lower test levels in which subtests were added or substituted. These changes and the reasons for them are discussed in the sections that follow. In these sections, also, a number of references are made to information obtained from the "field test." These are references to the extensive field testing of test materials prior to the selection of test questions for the final test forms. See the sections "Field Testing," beginning on page 32, and "Question Selection," beginning on page 38, for a description of this field testing and the ways in which the obtained data were used in test development.

Pilot Studies

Several tests and subtests were new for the Fourth Edition at Levels PR (Pre-Reading), BR (Beginning Reading), 1, and 2. The new tests and subtests involved either formats or types of content that had not been used in earlier editions. The authors needed to know if the new formats were easy for the students to follow, if the new content was appropriate in maturity and difficulty, and how much time was needed to administer the new tests. Pilot studies of the new tests for Levels PR, 1, and 2 were conducted in April of 1995. Classes from all parts of the country participated in the studies—22 classes at Kindergarten, 22 at Grade 1, and 21 at Grade 2. Teachers of these classes were generally pleased with the new tests, and the teachers' comments were very helpful in improving the tests and their administration. Test authors administered the pilot tests in four of the schools. The students' scores and data about the questions and the time needed to administer the tests in all the pilot schools were analyzed to provide guidance in developing the field-test forms.

A pilot testing on a smaller scale was carried out with the Basic Story Words subtest of Level BR in June of 1995 and again in May of 1996. Both testings involved a total of eight Grade 1 classes in three schools in Eastern states. These pilot tests were all administered by one or another of the test authors.

The pilot studies were used not only to find out how difficult the questions should be and how much time they required, but also to answer such fundamental and vitally important questions as

- Whether students had difficulty keeping the place;
- Whether the wording of the directions to the students was clear;
- Whether the directions were unnecessarily long;
- Whether the students understood where to look for the answer choices;
- When best to tell the students to put their pencils down.

These kinds of information were important to the authors because their aim was to develop a test in which the essential information was tested in a way that allowed the students to do their best. The authors' experiences had made clear to them that even the most knowledgeable adult may not know which wordings and page arrangements may cause the students unforeseen difficulties in understanding a task.

Level PR (Pre-Reading)

Changes from the Third Edition

Level PR (Pre-Reading) in the Fourth Edition replaces Level PRE of the Third Edition. In Level PR, Listening (Story) Comprehension is a new subtest for the Fourth Edition. The purpose of including this new subtest is to provide a progression of instruments to measure the development of comprehension

1. From listening to stories in Level PR;
2. Through reading Basic Story Words in context in Level BR;
3. To reading simple illustrated stories in Levels 1 and 2;
4. To reading progressively more advanced selections from published works in Levels 3 through 10/12.

Listening (Story) Comprehension is included in Level PR because the authors believe that students' experience in attending to important elements in a story, integrating information from different parts of a story, making inferences about story developments, and generally becoming engaged with oral text are critical components of the students' background for reading instruction.[1] The format of this new test is similar to that of the new format of the Comprehension tests of Levels 1 and 2. (See the section "Format of Levels 1 and 2," on page 16.) The teacher reads each story to the students. The story is read in five segments, and each segment is associated with a row of three pictures. The student's task is to choose the picture in each row that goes with the story. A small silhouette by each row of pictures helps the students attend to the proper row.

The format and sample stories of the Listening (Story) Comprehension subtest were evaluated in the pilot study described above. The results of the pilot study showed that the format the authors had developed was suitable for children at the end of Kindergarten and the beginning of Grade 1.

To permit the addition of the Listening (Story) Comprehension subtest without lengthening testing time, the types of background assessed by two Third Edition subtests—Literacy Concepts and Reading Instruction Relational Concepts—are assessed in the Fourth Edition in a single subtest: Literacy Concepts. By developing questions that focus on essential elements of these areas, the authors could provide a sensitive test of them with fewer total questions but higher reliability.

The other two subtests of Level PR—Oral Language Concepts (Phonological Awareness) and Letters and Letter-Sound Correspondences—are very similar to the corresponding subtests in the Third Edition. One change in the Oral Language Concepts subtest was that questions testing how well students can identify words that rhyme replaced questions that tested phoneme deletion. As students develop phonological awareness, the ability to recognize when two words rhyme generally precedes the ability to segment a word into phonemes.[2] The change from testing phoneme deletion to testing rhyme was made so that teachers might know when a student has difficulty with this early aspect of phonological awareness. Being alerted to such a difficulty allows the teacher to provide guidance and support that may help the student become aware of rhyme and also develop other aspects of phonological awareness.

Design of Level PR

A blueprint for the final form of Level PR specified the number of questions to be included in the first three subtests—Literacy Concepts, Oral Language Concepts (Phonological Awareness), and Letters and Letter-Sound Correspondences. For the field test, a larger number of question *types* was included in each of these three subtests than was to be included in the final test. For example, in one type of question used in the field test, the student's task was to select the short string of printed letters that looked most like a real word. This task was based on research on students' developing intuitions about letter patterns. However, because the field test showed that several questions of that type were difficult and evidently confusing, that type of question was not used in the published test.

The design of question types and of individual questions was also guided in part by an analysis that had been done as part of the field-test data analysis for Level PRE of the Third Edition. That analysis is described in the section "Field-Test Data Analysis" on pages 15–16 of the *Technical Report* for the Third Edition.[3]

On some of the pages of Level PR, each question begins with a picture in a square. By naming these pictures, the teacher can help make sure the students are working at the right place. On pages without these pictures, a small silhouette with an easily recognized outline and a familiar name (e.g., "star," "tree") is placed next to each question. The silhouettes differ from question to question, and no silhouette is repeated within any two-page spread. By using the names of the

silhouettes, the teacher can guide the students' progress down the test page. The silhouettes were carefully chosen to avoid cueing any of the answer choices either by sight or by sound.

While estimates of the time needed to administer the questions for the new Listening (Story) Comprehension subtest were obtained from the pilot test, the number of questions to be included in this subtest was left to be determined by data from the larger, national student sample taking the field test. The 20-question length tentatively selected on the basis of the pilot testing—five questions in each of four stories—proved to be appropriate, since the field-test data showed that a 20-question test would

- Take about the right length of time to administer;
- Be suitably reliable.

Level BR (Beginning Reading)

Changes from the Third Edition

Level BR (Beginning Reading) in the Fourth Edition replaces Level R of the Third Edition. In Level BR, the Basic Story Words subtest replaces the Use of Sentence Context subtest of the Third Edition Level R. The Basic Story Words subtest was introduced as part of the developmental sequence for assessing the growth of comprehension, from the Listening (Story) Comprehension subtest of Level PR to the reading comprehension tests at higher levels. The format and representative stories of the Basic Story Words subtest were evaluated in the pilot study described above.

The Basic Story Words subtest is designed to measure how well the student can read a sample of essential words that appear very frequently in stories for children, and, in fact, in all text. It is generally agreed that knowledge of such words (many of which are often taught as "sight words") is a critical aspect of beginning reading development.[4] The tested words are embedded in the sentences of simple stories that were written to provide meaningful context for the "story words." The teacher reads each sentence, including the "story word." The student must then choose, from among four printed words, the "story word" that the teacher has read.

The words tested in the Basic Story Words subtest were selected from either the "Dolch List"[5] or the "Revised Dolch List."[6] The other words in the sentences of these stories were also selected from these lists, with the restriction that no tested word could be used elsewhere in any of the story sentences. This restriction was included so that students could not learn to recognize a right answer by having it read to them in the context of one of the other questions during the testing.

Design of Level BR

For Level BR, the blueprint for the final test specified the number of questions to be included in each of the four subtests:

- Letter-Sound Correspondences: Initial Consonants and Consonant Clusters;
- Letter-Sound Correspondences: Final Consonants and Consonant Clusters;
- Letter-Sound Correspondences: Vowels;
- Basic Story Words.

For the first three subtests, the number of questions using each of two different formats was also specified. The grouping of letter-sound correspondences questions into subtests followed the pattern of the Third Edition. An analysis of various possible groupings for the Third Edition suggested that the grouping listed above would be most helpful as a basis for providing further instruction. (See pages 17–18 of the *Technical Report* for the Third Edition.[7]) To ensure their appropriateness, the selected letter-sound correspondences were checked against a list of skills taught in twenty-one reading programs.[8]

Following the design of the Third Edition, the three tests of letter-sound correspondences are administered in only two testing sessions. The total time required for administering any test includes not only the time students spend working on the test, but also the time spent in organizing the room, passing out the test materials, giving directions, and collecting the test materials. Therefore, combining the three subtests in two testing sessions saves considerable total time in administering Level BR. This arrangement is possible, since

- The question formats of the three subtests are sufficiently similar that the transition from one subtest to another does not require additional instructions to the students;
- The good reliability of the subtests means that they can be short enough that part of the third subtest can be administered with each of the other two in a testing session of reasonable length.

The administration of three subtests in only two sessions makes the scoring of the subtests slightly less convenient, but the total time saved—especially the students' time—is considerable.

As in Level PR, small silhouettes with familiar shapes and names were used for helping the students keep the place. For the letter-sound correspondences questions of Level BR, particular care was necessary to ensure that the sounds in the name of a silhouette do not cue any of the answer choices for the corresponding question.

Level AR (Adult Reading)

Level AR is an entirely new test level for the Fourth Edition. The purpose of Level AR is to provide community colleges and training programs with a reading test that, in concert with other assessment, can help locate students in need of improved reading skills. If such students can be located, they can usually be helped to develop their reading skills, giving them a better opportunity to be successful in their regular classes. Norms for Level AR were therefore desired that would reflect the range of reading skill typical of students entering community college—or training programs at that level. For that reason, norms for Level AR were obtained only at one time of year—in the fall, when the majority of community college students first enter.

After informal consultation with several community college programs, it was determined that Level AR should include a wide range of question difficulty and that the average question difficulty should be between that of Level 7/9 and that of Level 10/12. Thus, Level AR is *not* a further step in the progression of tests for the regular school grades.

The general structure of the tests for Levels 3 through 10/12 seemed suitable for Level AR. Indeed, some community colleges had been using Level 7/9 of the Third Edition for screening entering students. These community colleges had evidently found the general structure of Level 7/9 suitable, but wished that the range of question difficulty was wider and that the content of the Comprehension passages was somewhat more mature. Thus, Level AR is designed with the same structure as Levels 3 through 10/12 but the selection of Comprehension passages is designed to be suitable for young adults.

Word Decoding Tests

The Word Decoding tests for Levels 1 and 2 in the Fourth Edition have the same format as the Vocabulary tests for these levels in the Third Edition. These Fourth Edition tests are called "Word Decoding" for two reasons:

- To distinguish them from the new Word Knowledge test of Level 2
- To emphasize that the test format and the tested words measure primarily decoding skills and word identification, rather than knowledge of word meanings.

Selection of Test Words

Test words for the Word Decoding tests for Levels 1 and 2 are words that

- Were judged by the authors to be words that nearly all students in the grade for which the test level was designed would be likely to know *in speech.*
- Are commonly used in *reading* materials for Grade 3 or lower.[9]

- Follow common orthographic rules as exemplified by the letter-sound correspondence skills commonly taught in reading programs.[10]
- Comprise common orthographic patterns that permit the test word to be contrasted with wrong answer choices that are similar in spelling, so that choosing the correct answers depends on using specific decoding skills.

None of the Word Decoding test words selected for field testing in the Fourth Edition had been used in the Third Edition.

Selection of Wrong Answer Choices

Wrong answer words were chosen so that each one was similar to the test word but differed from it in some significant way so that, when pronounced using common letter-sound correspondences, it would clearly be a different word from the test word. Thus, the spelling of the wrong answer and the test word might be just the same except for one different consonant or vowel letter or except for an added or omitted letter. For example, one of the practice questions for the Word Decoding test shows a picture of a hat, and the wrong answer choices are *hot, hit,* and *hut.* All answer choices are real words; no nonsense words are used.

A wrong answer word did not have to be as familiar as the test word, as long as its pronunciation would make it clearly wrong for a student who can read the test word by using the decoding skill that distinguishes the test word from the wrong answer. Homophones of test words were not used as wrong answers.

Pictures for Representing Test Words

Specifications were written for each of the pictures depicting the correct answer words. These specifications were used by the illustrators to guide their work. Several guidelines were established for preparing these specifications. The guidelines were intended to ensure that the specifications would lead the illustrators to draw pictures that

- Picture the most common version or style of the object or action (whatever would be most recognizable to students all across the country);
- Focus on the named object or action:
  - The view chosen should make the object or action evident,
  - The object or action should be as large as possible in the picture space,
  - Other objects or actions inherent in the picture should be de-emphasized;
- Show only what is necessary to communicate the object or action clearly (Unnecessary detail, background, or shading can reduce the clarity with which the object or action is depicted.);

- Ensure that nothing in the picture could reasonably be interpreted as an illustration of one of the wrong answer words;
- Picture whole objects (Pictures of parts of objects tend to be difficult to interpret and should be used only when showing an entire object is not possible or when providing context for an action.).

These guidelines were followed in creating specifications for each picture. The specifications were given to an illustrator, and the resulting pencil drawings were critiqued by the authors and often were then changed or redrawn. Once the pencil drawings were approved by the authors, they were inked in by the illustrator and then examined again by the authors.

Decoding Skills Analysis Forms and Reports

To increase the usefulness of the Word Decoding scores, a Decoding Skills Analysis Form and a Decoding Skills Analysis Report were developed for each of the three Word Decoding tests. Decoding Skills Analysis Forms are filled out by the teacher. Decoding Skills Analysis Reports are available as a separate service for tests that are scored by the Riverside Scoring Service®. Both the Decoding Skills Analysis Forms and the Decoding Skills Analysis Reports show, for each Word Decoding question, the decoding skill that a student who chose a wrong answer may not know.

The three wrong words for each Word Decoding question look and sound much like the correct word, so that selecting the correct word ordinarily requires knowing the sound that corresponds to a tested letter or letter sequence. The listed skills were determined by comparing each wrong answer with the correct answer and noting the crucial difference between the two similar words. When more than one skill is involved in choosing the correct over an incorrect answer, the skill listed is the one that was considered to be primary in importance—the one that, if not known, suggests the most serious problem in decoding. In general, if a wrong answer choice suggests that the student does not know a skill that is usually learned early and well by most beginning readers, that error was considered more serious than an error involving a skill that is usually learned later or less well. This ranking of seriousness was based partly on the authors' experience and partly on the sequence in which categories of skills are generally taught.[11]

Word Knowledge and Vocabulary Tests

For the Word Knowledge test at Level 2 and for the Vocabulary tests at Levels 3 and higher, the test words were selected to

- Represent an appropriate range of difficulty for the test level in which they are used;
- Be words of general usefulness, not obscure or specialized words;
- Represent an appropriate distribution of parts of speech.

None of the test words selected for field testing in the Fourth Edition had been used in the Third Edition. Nearly all test words for the Level 2 Word Knowledge test and the Level 3 Vocabulary test appear on the grade level lists given by Harris and Jacobson.[12] Nearly all the test words chosen for Levels 4, 5, and 6 appear in either the Harris-Jacobson lists or *The Living Word Vocabulary* and its *Supplement.*[13] Because grade level designations for words in *The Living Word Vocabulary* begin at Grade 4, the Harris-Jacobson grade level lists were more useful for selecting the easiest words for Levels 4, 5, and 6. For Levels 7/9, 10/12, and AR (Adult Reading), nearly all test words appear in *The Living Word Vocabulary. The American Heritage Word Frequency Book*[14] and *The Educator's Word Frequency Guide*[15] were consulted primarily for questions about the probable relative familiarity of words, for, although these two word frequency lists give separate entries for different inflections, they do not differentiate among meanings.

For all test forms, the grade level designations of the words found in the word lists were used to supplement the authors' judgments in developing a sample of test words that would be appropriate for the range of vocabulary knowledge characteristic of the grade level(s) for which a test level was designed. The word lists also helped in selecting the test words for the equivalent final test forms.

Test words selected for the Word Knowledge test at Level 2 were judged by the authors as likely to be known in speech and print by those Grade 2 students who possess good reading vocabularies, but much less likely to be known than the test words in the Word Decoding test. The average Harris-Jacobson grade level rating of words in both published forms of the Word Knowledge test is 4.0. The average ratings in the Word Decoding test are 2.8 for Form S and 2.7 for Form T.

The Word Knowledge and Word Decoding tests in Level 2 have the same format. A picture is presented along with four words, one of which is represented by the picture. However, not only are the test words in the Word Knowledge test less familiar than those in the Word Decoding test, the two tests also differ in the nature of the wrong answer choices.

- In the Word Decoding test, all the wrong answers are visually similar to the correct answer, often differing in only one letter or short letter cluster, so that selection of the right answer depends on knowledge of a particular letter-sound correspondence.
- In the Word Knowledge test, there is typically little visual similarity among the answer choices for a particular question. The answer choices differ in meaning, and only one fits the picture, although the wrong answer choices may be associated with the correct answer. Thus, a wrong answer might be a word that is often used in similar contexts or that shares some semantic features with the right answer. For example, the practice question for the Word Knowledge test shows a picture of a road, and the wrong answer choices are *bus, truck,* and *city.*

Parts of Speech

The parts of speech of the test words in a vocabulary test should roughly reflect the proportions of the parts of speech of words that students need to learn in order to read appropriate materials with understanding. Data on these proportions of parts of speech were obtained from two sources: the "Dale List" (a list of 3,000 words known by students in Grade 4), as updated for the new Dale-Chall Readability Formula,[16] and the *Frequency Analysis of English Usage: Lexicon and Grammar* by Francis and Kučera.[17]

Since the words on a list such as the Dale List are not presented in context, their part-of-speech membership cannot be determined from the list. In fact, many of the word forms on the list occur in text as, and are familiar as, more than one part of speech. Each word on the list was therefore considered according to the criteria given by Francis[18] to determine its part-of-speech memberships. The part-of-speech categories that were used were *noun, verb, adjective,* and *adverb and other.* A word was listed under each part of speech for which it met the criteria of membership. The words in each of these lists were counted, and each of these counts was divided by the total of the four counts. The resulting percentages of words on the Dale List that can be used as nouns, verbs, adjectives, and adverbs or other parts of speech are shown in Table 1.

The *Frequency Analysis of English Usage: Lexicon and Grammar* by Francis and Kučera gives the percentages of different word classes in the 1,000,000 word "Standard Corpus of Present-Day American English,"[19] often referred to as "the Brown Corpus." These percentages are percentages of *running* words. To obtain rough estimates of the percentages of *different* words in the corpus, only those words used as nouns, verbs, adjectives, and adverbs were considered. Several closed classes—such as determiners, prepositions, pronouns, and conjunctions—always make up a very large percentage of *running* words in normal text (roughly

40% of the Brown Corpus) but represent only a small fraction of the number of *different* words a reader encounters. To exclude the closed classes from the calculations, the percentages of nouns, verbs, adjectives, and adverbs in the Brown Corpus were each divided by 57.3%, the sum of those four percentages. The resulting percentages of each of these parts of speech are shown in Table 1.

Table 1. Percentages of Parts of Speech

Part of Speech	Dale List	Francis & Kučera	*GMRT* Vocabulary Tests
Noun	50	47	50
Verb	35	32	35
Adjective	9	12	10
Adverb and Other	5	9[a]	5

[a] Adverbs only

In the field-test forms and in the final published forms, the percentages of the parts of speech used as test words in each level and form of the Word Knowledge and Vocabulary tests conform closely to the figures shown in the last column of Table 1. The part-of-speech classification of test words that could be used as more than one part of speech was determined by the answer choices that were offered by the picture (in the Level 2 Word Knowledge test) and by the context frame (in the Vocabulary tests of Levels 3 through 10/12 and AR). All answer choices for any one question were required to be the same part of speech.

Context

In Levels 3 through 10/12 and AR, each Vocabulary test word is presented in a brief context frame. For the purpose of measuring reading achievement, knowledge of word meanings should not be thought of as separate from comprehension but as an important component of comprehension. Various measures of word difficulty have been used in readability research with the result that, whatever measure is used, word difficulty turns out to be the primary contributor to the reading difficulty of texts.[20] It is not surprising, then, that knowledge of word meanings has repeatedly been shown to be the preeminent factor in reading comprehension.[21] Measuring vocabulary knowledge as a component of reading comprehension, however, requires distinguishing it as much as possible from other components. The more information a context frame provides, the more the question resembles a general comprehension task and the less it measures the distinct contribution of word knowledge. For this reason, the context frames in a reading vocabulary test should provide very little information about the meanings of the words.

One basic kind of information about a word that a reader may learn from context, even without knowing the meaning of the word, is its part of speech. In fact, except for words in sentence initial position, the part of speech of most words is predictable from their context.[22] Thus, it seems desirable for test words to be presented in contexts that suggest their parts of speech, but that do not give other clues to the meanings of the words.

Each Vocabulary test word was therefore embedded in one of about twenty defined grammatical frames that suggest the part of speech of the word. Not all of the frames uniquely determine the part of speech of the test word; they do, however, limit the possibilities, and they provide natural, commonly used contexts. They are also brief and simple enough that they do not require much reading time or effort. In keeping with the function of the context frame, it was required that all answer choices for a particular test question be the same part of speech as the test word and fit the context semantically and grammatically.

The type of context provided in the Vocabulary questions of the *GMRT* permits considerable independence for the Vocabulary scores from the Comprehension scores, as is shown in the section "Correlations among Tests," beginning on page 60. This relative independence makes Vocabulary scores on the *GMRT* useful in exploring reasons for a student's low reading achievement.

Independence of Questions

To ensure the independence of Vocabulary questions and a broad sampling of words, rules were developed to guide the reuse of words. For the test questions to be independent, knowledge of a given word should not be tested more than once, even at different test levels. Also, if the same words are offered too often as answer choices, students are likely to assume that there is some right/wrong pattern to the use of the repeated words in various questions. In trying to discern that pattern, even if none exists, students may be distracted from the task of matching the meaning of the test word with that of the correct answer choice. The following rules were therefore observed in writing Vocabulary questions:

- A test word may not be reused as a test word at any level.
- A test word may not be used as a right answer, wrong answer, or content word in a context frame ("context word") in the same level as the test word, but it may be so used at other levels.
- A right answer may not be reused as a right answer in the same level, but it may be used in the same level twice as a wrong answer and twice as a context word, or only once as a wrong answer and three times as a context word.
- A wrong answer may be used in the same level a total of
  - One time as a wrong answer and five times as a context word;
  - Two times as a wrong answer and four times as a context word;
  - Three times as a wrong answer and three times as a context word;
  - No more than three times as a wrong answer.
- A context word may be used within a given test level without restriction, if the word is not also a right answer or a wrong answer in the same test level.
- A word that is a right answer, wrong answer, or context word at a given level may be used in any of these roles at other levels.

In applying these rules, the Level 2 Word Decoding and Word Knowledge tests were treated as if they were separate levels. The rules concerning context words do not apply to Levels 1 and 2.

Types of Vocabulary Answer Choices

When students do not know the answer to a vocabulary question, they may adopt strategies for choosing an answer that are not very relevant to the assessment of their knowledge of word meanings. For example, a student who does not know the answer to a vocabulary question may choose the longest answer, the third answer, an answer that begins with the same letter as the test word, an answer that has a similar word part, an answer that has a similar general appearance, an answer that is suggested by as much of the test word as the student can decode, an answer that seems related in some way to a vague categorization of the meaning of the test word, or an answer based on some other strategy.

In writing vocabulary questions it is important, therefore, to present an array of answer choices that will not produce a high score for a student who consistently uses such a strategy when he or she does not know the meanings of the test words. For this reason, the *GMRT* Vocabulary tests include many questions in which a wrong answer is the longest answer or begins with the same letter as the test word. The correct answers are essentially randomly distributed in the various answer positions, and no one position contains a preponderance of correct answers.

Levels 3 through 10/12 and AR also include wrong answers that are related to more complex strategies that students may apply when they do not know the meaning of the test word. At these levels, many Vocabulary questions include one or more of three different types of wrong answers: *visual similarity, miscue,* and *association.* It was assumed that these three wrong-answer types appeal to three different strategies that represent three levels of understanding the meaning of a word in print.[23]

- ***Visual Similarity Wrong Answers.*** A visual similarity wrong answer looks like the corresponding test word in some way. Typically, it begins with the same two or three letters or ends with the same rime or suffix. For example, a visual similarity wrong answer for the test word *carpet* might be *carton;* a visual similarity wrong answer for *timid* might be *solid.* Care was taken to make the visual similarity wrong answers resemble the test words in a variety of ways. Also, a few correct answers resemble the test word in similar ways.

  It was assumed that a student who chooses a visual similarity wrong answer is likely to have had difficulty sounding out or recognizing the test word. If the student does not identify the test word correctly, he or she cannot compare the meaning of the test word with the meanings of the answer choices and therefore cannot see the similarity in *meaning* between the test word and the correct answer choice. Visual similarity may be the only relationship the student sees. Lacking other relevant knowledge, he or she may rely too much on visual similarity.

- ***Miscue Wrong Answers.*** A miscue wrong answer may be particularly appealing to a student who has misread the test word or who has read only part of it. A miscue wrong answer typically is associated with an incorrect reading of a test word but is not a synonym of that incorrect reading. (Synonyms of incorrect readings are likely to be amusing and distracting to some students.) For example, a student who misreads *dozen* as *dozing* may think that *nap* is a good right answer; a student who sees *ear* in *earth* but cannot read the whole word may be attracted to *sound* as an answer; a student who sees *list* in *listen* but does not recognize the word *listen* may think *write* is a correct answer. A miscue wrong answer is likely to appeal to a student who does not decode well but who can partially decode the test word, arriving at a misreading of it. The student would then look for an answer choice that is related to his or her misreading of the test word.

- ***Association Wrong Answers.*** An association wrong answer is typically a word that names some object, activity, feature, or feeling that frequently occurs in the same physical setting or the same verbal context as the test word. For example, *want* is a feeling often associated with *grab;* a *crime* is often the occasion for an *alarm; hope* can be used in many of the same verbal contexts as *pleasure.* An association wrong answer typically shares some semantic feature(s) with the test word but is a poor synonym for it. It is likely to appeal to a student who recognizes the test word but has only a vague understanding of its meaning.

Visual similarity, miscue, and association wrong answers were included in the Vocabulary questions for the express purpose of providing a variety of wrong answers that would be consistent with various strategies a student might use when he or she does not know the right answer. One or more of these three types of wrong answers was included in a question when doing so would make the question more effective. Many of the Vocabulary questions include a visual similarity and an association wrong answer, but relatively few questions include a miscue wrong answer.

Since visual similarity, miscue, and association wrong answers were not systematically included in the Vocabulary questions, a listing by question of these wrong answer types is not provided in this *Technical Report.* A more important reason is that an extensive study of these answer types undertaken for the Third Edition showed that scores based on the types of wrong answers students chose were not very reliable. Many students, at least, are evidently not very consistent in which of these three wrong answer types they choose. It is not surprising that this is the case. For example, when a student has only a vague idea of the meaning of a test word, he or she might choose an association wrong answer. When that *same* student either cannot decode the test word or has *no* idea of its meaning, he or she, having nothing else to go on, might choose a visual similarity wrong answer.

Tallying which of these types of wrong answers a student chooses is thus not justified as a means of diagnosing the basis of a student's low Vocabulary score. The authors' clinical experience, however, indicates that these wrong answer types can be useful to keep in mind when working with a student individually and probing for the student's reasoning in choosing answers.[24] Interviews conducted by the authors with students at all the test levels showed that some students did tend to rely on these strategies. Often, however, students gave idiosyncratic reasons for choosing an answer or indicated that they just guessed.

Familiarity of Answer Choices

For the following reasons, many of the visual similarity wrong answers were intentionally chosen to be relatively unfamiliar.

- If a student does not know the meaning of a test word, he or she may be drawn to a visually similar wrong answer. If the student also does not know the meaning of the wrong answer, the appeal of the visually similar wrong answer word will probably not be diminished and might even be enhanced by sharing with the test word an unknown meaning.
- If the student *does* know the meaning of a wrong answer that resembles the test word, that wrong answer might possibly seem incorrect in the context of the other answer choices.
- By not restricting too much the familiarity of visual similarity wrong answers, wrong answers that look more like the test word can be used.

Association and miscue wrong answers, however, will not be appealing unless the meanings of the wrong answers are at least somewhat familiar. Therefore, correct answers, association wrong answers, and miscue wrong answers were all chosen to be relatively familiar compared to the test word. To try to ensure that all answer choices, except for visual similarity wrong answers, would be at least as familiar as the test word, it was required that the grade rating of the answer choices, as given by Harris and Jacobson[25] or by *The Living Word Vocabulary,*[26] be the same as or lower than that of the test word.

Comprehension Tests

Test development procedures were designed to provide a set of reading materials and test questions that, within the limits of a multiple-choice test, would be broadly representative of the materials students are expected to learn to read and of the types of questions about those materials that students should be able to answer from their reading. Passages for the Listening (Story) Comprehension subtest in Level PR, the Basic Story Words subtest in Level BR, and the Comprehension tests in Levels 1 and 2 were written specifically for the test. In Levels 3 through AR, passages for the Comprehension tests were chosen from published materials appropriate for students at the grade levels for which the test level was developed. The selected passages were all new for the Fourth Edition. None of the passages in any field-test or final form had been used in the Third Edition.

Format of Levels 1 and 2

A new format was devised for the Comprehension test for Levels 1 and 2. The new format permits using actual stories or informational texts but retains the simple and uniform answer format of the Third Edition. Each passage contains four segments (except for the last passage in each test form, which contains only three). Each segment is associated with a row of three pictures. The student's task is to choose the one picture that illustrates the segment or that answers a question about the segment. In many cases, in order to choose the correct picture, the student must consider information from earlier segments of the text as well as the current segment.

This format avoids burdening young children with

- The additional heavy reading load of reading a series of answer words or phrases;
- The complications of relating these phrases to a question *and* to the passage;
- The complications of choosing among many written alternatives.

The task is thus relatively easy to understand, and the pictures add interest and familiarity.

The various stories and informational texts were written by the authors and by five professional writers.[27] These passages had to conform to the same requirements, when applicable, as those described in the section "Passage Characteristics in Levels 3 through 10/12 and AR" on page 17, with the additional requirement that each segment could be a source for three pictures—one that illustrated the segment or answered a question about it, and two related but incorrect pictures.

Picture Choices at Levels 1 and 2

Specifications were written for the three pictures to be drawn for each segment of each passage. Guidelines for preparing these specifications were similar to those for the Word Decoding tests as described in the section "Pictures for Representing Test Words" on page 7 and were used in the test development process in a similar way.

A further central requirement for the picture choices was that neither of the two wrong answer pictures for each segment could reasonably be interpreted as an illustration of the passage. In order to be effective wrong answers, the wrong answer pictures had to relate to the general context of the passage—for example, might depict the same characters. However, some obvious aspect of the picture had to be wrong or not fit the story—for example, a wrong answer picture might show the story characters doing something different from what was described in the segment. The picture panels for each narrative passage were examined to make sure that the story could not be educed from the sequence of panels.

Passage Characteristics in Levels 3 through 10/12 and AR

For Levels 3 through 10/12 and AR, the authors, with the help of several teachers and former teachers, located appropriate Comprehension passages in a variety of published sources. The teachers and former teachers also suggested questions that might be asked about the passages. In selecting the passages, the authors and the teachers and former teachers followed a number of guidelines designed to ensure that the passages would be appropriate for the tests. The passages were to

- Be like passages that the students would be expected to read in school or be likely to read for their own enjoyment or information;
- Be complete in themselves—not seem out of context. Understanding the passage and answering the questions should not depend on inferring events or information given earlier in text not included in the passage. For a few passages, the authors provided a brief introduction to make sure the students would have an easy entry into the passage;
- Not be from books or other materials that were currently very popular, or used in many classrooms, or likely to have been read by many students. The aim was to avoid material that would already be familiar to many students;
- Be of varied authorship. There should be no more than one passage from any book and no more than two passages from the same author (and those not to be used at the same level);
- Have conceptual content roughly at the same level as their reading difficulty;
- Not have content that would be offensive or disturbing to students at the grade level for which the passage was intended. Since the students would not have an opportunity to discuss the passages with their teacher before taking the test, the passages should not contain content that would distract the students from doing their best to think about the passage and answer the questions.

Additional guidelines were that

- Expository passages should not be about a very familiar topic. The students should not be able to answer the questions on the basis of prior knowledge;
- Narrative passages should have a story line that rings true and that is not a variant of a familiar plot.

Number of Passages

It has long been recognized that prior knowledge plays an important role in reading comprehension. Indeed, much current research and practice in reading concerns helping students learn to apply their prior knowledge.[28] Thus, prior knowledge is an inextricable influence in the score a student gets on a comprehension test.[29] For a reading achievement test such as the *GMRT,* the aim must be to try to assess how well the student can apply reading skills *and* prior knowledge to construct an understanding of text.

Texts used in a reading achievement test should be what students might be expected to read in school or might read for their own enjoyment or information. Such texts will include a wide range of content and genre. A single passage could not represent that range. If the passage were expository, a student might happen to know a great deal about the topic of the passage—or very little about it. If the passage were a narrative, a student might be very familiar with the situations and events being described—or not at all familiar with them. Thus, one or two expository passages and one or two narrative passages would not be an adequate sample for measuring a student's ability to construct meaning from text. In addition, most stories include a significant amount of text that is not actually narrative, but *setting*—descriptions of scenes or situations. Since setting text is such a familiar part of most narratives, some representation of this type of text is also desirable in a survey test of reading comprehension.

In the *GMRT,* there are 10 passages in each Comprehension test in Levels 1 and 2 and 11 passages in each Comprehension test in Levels 3 through 10/12 and AR. It is not known if that is an optimal sample, but it is surely better than a sample of four or five passages would be.

A practical limitation on the number of passages is that increasing the number of passages means that, for the testing time to be of reasonable length, the passages must be shorter. In the authors' experience, the passage length that results from using 10 or 11 passages and a 35-minute testing time is fully adequate to provide the basis for testing the students' ability to construct meaning from expository, narrative, and setting passages.

Even with 10 or 11 passages, however, there are many types of reading tasks that cannot be adequately sampled. The Fourth Edition of the *GMRT* does not contain any samples of poetry, for example. Response to poetry is such a different type of reading that including one or two examples of poetry would have reduced the opportunity to measure understanding of prose without giving an adequate sample of response to poetry. The *GMRT* also does not contain timetables, assembly

directions, and advertisements. Including such specialized genres would dilute the validity of the score as an indication of ability to understand expository, narrative, and setting text, while sampling the specialized genres inadequately.

Comprehension Questions at Levels 3 through 10/12 and AR

Several guidelines governed the construction of questions about the passages in Levels 3 through 10/12 and AR.

- The questions should assess understanding of significant concepts and relationships necessary for constructing meaning that is complete, consistent, and rational.
- There should, of course, be only one right answer to any question.
- The reading level of the question and the answer choices should be as easy as possible and, in any case, less difficult than the passage. (The question should test understanding of the passage, not the ability to read the question.)
- Similarly, the *amount* of reading required by the questions and answer choices should be relatively small compared to the passage; they should not unnecessarily add reading effort and reading time.
- The questions should be clear, and the answer choices, both right and wrong, should clearly relate to them—should seem, to a student who has not read the passage, like plausible answers. (Exceptions would be some text-phrase wrong answers that simply repeat appealing phrases from the passage. See the section "Types of Comprehension Answer Choices" on page 26.)
- In keeping with the need for the questions to be clear, short, and easy to read, the question stem might be either an actual question or an incomplete statement. The question form was generally preferred, but if using that form added wordiness and detracted from the clarity of the question, the incomplete statement form was used.
- The stem should state a clear question. This is often not possible when the stem is an incomplete statement. In that case, the question should be clear from reading the stem and any one of the answer choices.
- The correct answer to a question should not be based entirely on prior knowledge. It should not be possible to answer the question by reading only the stem and the answer choices.
- The questions should be independent of each other. Knowing the answer to one question should not help in answering another.
- It should also not be possible for a student, without reading the passage, to construct the theme or main idea of a passage or the outline of a story from the set of questions about it.

- A comprehension question should not be a vocabulary question in disguise. The answer should not depend on knowing the meaning of an uncommon word in the passage, unless its meaning can unambiguously be deduced from context. Some questions about words, however, are good comprehension questions. Questions requiring the student to choose among possible meanings of a word by using the context of the passage, and questions that ask the student to identify or infer what a particular word in the passage refers to (the referent of the word) can be good comprehension questions.
- Correct answers should not repeat salient words or phrases from the passage, except to balance text-phrase wrong answers. (See the section "Types of Comprehension Answer Choices" on page 26.)
- All answer choices must grammatically and logically fit the question stem. It should not be possible to eliminate any of the answer choices because they do not fit the stem grammatically or, except for some text-phrase wrong answers, because they do not make sense in relation to the stem.
- The same answer choice, whether right or wrong, should not be used more than once for any one passage. Students are likely to develop hypotheses about the correctness of an answer just from the fact of its being repeated.

Passage Content

Students need to learn to read a wide range of prose, including both literature and the content areas. To ensure that the tests present a good balance of content, the passages used in the field-test and final test forms were required to conform to the blueprint shown in Table 2. The content classifications of the passages in the published forms of Levels 1 through 10/12 and AR are shown in Appendix A.

Table 2. Numbers of Comprehension Test Passages in Each Content Category in Each Test Form

Content Category	Test Level			
	1	2	3–4	5–AR
Fiction	5	4	5	4
Social Science	2	3	3	3
Natural Science	3	3	3	3
Humanities[a]				1
Total	10	10	11	11

[a] Art, music, the study of literature

Narrative and Expository Text

Students are expected to learn to read narrative text, and they are also expected to learn to read expository text for information. Since reading and learning from expository text generally seem to require somewhat different strategies and background than reading narrative text,[30] and since the ability of students to read expository text is a matter of practical concern,[31] the amount of narrative and non-narrative content in the test forms was controlled.

The distinction between narrative and non-narrative text is more complicated than it appears on the surface.

- A clear, functional, and generally accepted definition of what constitutes a narrative passage is not readily available; some passages that seem like a story may not actually be narratives according to some definitions.
- Some materials written expressly to convey subject matter content may be written as narratives.
- Readers tend to use multiple criteria to classify passages as narrative or non-narrative.[32]

Nevertheless, it is possible to classify text passages fairly consistently.[33] The test passages were classified as narrative or non-narrative. Non-narrative passages were further divided into passages whose intent seemed primarily to instruct (*expository* passages) and passages that seemed characteristic of those sections of stories that do not actually move the account forward in time (*setting* passages). Only passages that contained "accomplishment" or "achievement" sentences[34] were classified as narrative. The blueprint shown in Table 3 was developed to provide a good balance of narrative, expository, and setting passages. The field-test and published test forms conform to this blueprint, with the exception that the published Level 3 forms contain five expository and no setting passages. No setting passages with acceptable question statistics survived the field test at that level. The narrative/exposition/setting classifications of the passages in the published forms of Levels 1 through 10/12 and AR are shown in Appendix A.

Table 3. Blueprint for Numbers of Narrative, Expository, and Setting Passages in Each Test Form

Passage Type	Test Level				
	1	2	3–4	5–7/9	10/12–AR
Narrative	6	5	6	5	4
Expository	3	4	4	5	6
Setting	1	1	1	1	1
Total	10	10	11	11	11

Age Appropriateness

The reading material used for testing comprehension should be suitable in content and tone for the students taking the test. The content should reflect the kinds of topics and story situations that the students are learning to read about in school and in leisure reading. Topics that appear too childish will sap motivation. Topics that are too mature may be puzzling or troubling. For similar reasons, the tone of the writing should also be appropriate to the age of the students.

In choosing passages for the tests, care was taken to avoid topics in expository passages that would be overly familiar. This was done to ensure that the students could not answer the questions on the basis of prior knowledge alone. Nonetheless, it is also important that the topics should be comprehensible, given the students' backgrounds. Based on the authors' experience working with students and teachers at many different grade levels, the majority of the passages selected for the field test were rated for the range of grade levels at which the content and tone of the passage would be appropriate. These ratings were used as guides in assigning the passages to Levels 3 through AR. Other passages that were not rated in this way were assigned to particular test levels on the basis of library lists and other guides.

Readability

Students vary in their knowledge about, and interest in, particular topics or situations; students also vary in their familiarity with particular types of prose. As a result, readability formulas are only crude predictors of the reading difficulty of a passage for an individual student. Readability formulas are better at predicting the average reading difficulty of a passage for a group of students. But there are still many passages with difficult concepts expressed in simple language, or simple concepts expressed in difficult and unfamiliar language, for which predictions made by any readability formula are grossly in error. In addition, different formulas will often give very different difficulty estimates. Thus, it is unwise to pay too much attention to the computed readability of a particular passage, especially when that readability figure is obviously distorted by some interaction between the characteristics of the passage and the formula. For example, a formula that measures vocabulary difficulty by the number of syllables per word may give a distorted estimate of reading difficulty for a passage that has a few long but familiar words repeated throughout.

When several passages are being evaluated, however, the average readability estimate for those passages may be a useful indication of the average difficulty of those passages for a typical class. The average readability of passages in a reading test usually will conform roughly to the grade level(s) for which the test is designed. This does not mean that a test for Grade 3, for example, will include only passages with Grade 3 readability estimates. Many students in Grade 3 typically read well only materials that are easier than those rated Grade 3, and many good readers in Grade 3 regularly read materials that are considerably harder than those rated Grade 3.

Passages for the field test of the Fourth Edition were selected to provide an appropriate range of readability at each test level. Readability of the passages was assessed with three readability formulas: Dale-Chall,[35] Fry,[36] and Spache.[37] Readability estimates from each of the three formulas were averaged separately to obtain readability estimates for each test level and form.[38]

Since the Spache formula is intended for use only with materials below Grade 4,[39] average Spache readability estimates are not reported for Level 5 and higher; at those levels, many of the Spache estimates were not valid, making averages of Spache estimates for those levels meaningless. All of the Spache estimates for passages in Level 4 and lower were valid—none were above 3.9. Even at Levels 3 and 4, however, the average Spache estimates were much lower than either the Fry or Dale-Chall estimates. The Spache estimates for those levels should probably be disregarded. The average readability estimates for each published level and form are shown in Table 4. That even the *averages* of the different formulas are sometimes quite different for a particular level and form makes clear that measures of reading difficulty based on formulas are very imprecise.

Table 4. Average Readability by Level and Form

Average of	Form S	Form T
Level 1		
Fry	2.1	
Dale-Chall	2.2	
Spache	2.1	
Fry & D-C & Sp	2.1	
Fry & D-C	2.2	
Level 2		
Fry	2.7	3.0
Dale-Chall	3.1	2.5
Spache	2.3	2.4
Fry & D-C & Sp	2.7	2.6
Fry & D-C	2.9	2.8
Level 3		
Fry	3.9	3.7
Dale-Chall	3.0	3.3
Spache	2.8	2.5
Fry & D-C & Sp	3.2	3.2
Fry & D-C	3.5	3.5
Level 4		
Fry	5.4	4.8
Dale-Chall	3.9	3.9
Spache	3.3	3.1
Fry & D-C & Sp	4.2	3.9
Fry & D-C	4.6	4.3
Level 4 without *The Other Way to Listen*		
Fry	5.4	5.1
Dale-Chall	3.9	4.0
Spache	3.3	3.1
Fry & D-C & Sp	4.2	4.1
Fry & D-C	4.6	4.5

Average of	Form S	Form T
Level 5		
Fry	5.9	5.9
Dale-Chall	4.5	5.1
Fry & D-C	5.2	5.5
Level 5 without *The Legend of Food Mountain*		
Fry	6.2	5.9
Dale-Chall	4.7	5.1
Fry & D-C	5.4	5.5
Level 6		
Fry	6.7	7.3
Dale-Chall	5.6	5.5
Fry & D-C	6.2	6.4
Level 7/9		
Fry	8.5	8.2
Dale-Chall	7.4	7.6
Fry & D-C	7.9	7.9
Level 10/12		
Fry	10.5	9.5
Dale-Chall	10.0	10.5
Fry & D-C	10.2	10.0
Level AR		
Fry	9.1	8.5
Dale-Chall	8.0	7.7
Fry & D-C	8.5	8.1
Level AR without "Frank Lloyd Wright"		
Fry	8.8	8.5
Dale-Chall	7.4	7.7
Fry & D-C	8.1	8.1

One passage in Level 4, Form T, one in Level 5, Form S, and one in Level AR, Form S epitomize the difficulties of characterizing passage difficulty with formulas that simply count gross characteristics such as word frequency or sentence length. In Level 4, Form T, the passage from *The Other Way to Listen*[40] (questions 18–21) uses only short, familiar words, but the concepts and the required inferences are difficult. In Level 5, Form S, the passage from *The Legend of Food Mountain*[41] (questions 4–6) has relatively short sentences, partly because it includes conversation. Although the words are generally not difficult, the situation requires applying imagination to an unfamiliar scene, and the passage is not as easy to understand as the short sentences might suggest.

A passage about Frank Lloyd Wright in Level AR, Form S (questions 37–40) illustrates a passage that would be much easier to read than the formulas estimate. It includes a number of relatively long words that are not included in the Dale List of easy words but that would be quite familiar to most mature individuals for whom the test is designed—*expensive, gutters, visible, downspouts, horizon.* Such words also add to the average word length in syllables, used by the Fry formula as a measure of word difficulty. Other words, though familiar and on the Dale List, also add to the number of syllables—*designing, unnecessary, old-fashioned, basements, furniture, plastering, painting, amazing, buildings.* The extra sections in Table 4 for Levels 4, 5, and AR show the readability estimates when these three passages are not included in the averages.

Reading Difficulty

Before readability analyses had been run, the test authors rated the reading difficulties of the majority of the passages that were being considered for inclusion in the tests at Levels 3 through AR. These qualitative judgments were based on characteristics such as sentence structure, semantic structure, clarity of wording, organization, vocabulary, and signaling. The judgments also took into account the difficulty of the content—familiarity, abstractness, conceptual load, difficulty of inferences, and need for relating ideas. From a consideration of these factors, a rating of reading difficulty was assigned, expressed as a grade level.

Since the authors' ratings of reading difficulty were based on several factors in addition to vocabulary difficulty and sentence length, they provided information about passage difficulty that was not given by the readability estimates. The authors' subjective ratings of reading difficulty do, of course, correlate fairly highly with the readability estimates, since the authors did give considerable weight to vocabulary difficulty and syntactic complexity. Of the 154 passages used in the two published forms of Levels 3 through 10/12 and AR, 83 had been given grade-level ratings of passage difficulty. The correlations of these ratings with the Fry and Dale-Chall readability estimates are shown in Table 5.[42]

Table 5. Correlation of Authors' Ratings of Reading Difficulty with Fry and Dale-Chall Readability Estimates (*N* = 83)

	Correlation			
	Fry	D-C	Mean	*SD*
Reading Difficulty	0.76	0.79	6.80	2.81
Fry		0.89	7.12	2.60
Dale-Chall			6.01	2.99

Although the correlation between the Fry and Dale-Chall estimates for these 83 passages was quite high, the average grade level predicted by the two formulas was rather different. As can be seen in both Tables 4 and 5, the Fry readability grade level estimates were generally higher than the Dale-Chall estimates. Above Level 2, the Fry grade levels shown in Table 4 average roughly nine-tenths of a grade level higher than the Dale-Chall grade levels. For the 83 passages with authors' ratings of reading difficulty (Table 5), the Fry averages eight-tenths of a grade level higher than the Dale-Chall. Table 5 shows that the authors' grade level ratings of reading difficulty were intermediate, on average, between the Fry and Dale-Chall grade level estimates.

The authors' ratings of reading difficulty were a major determiner of the test levels to which the passages were assigned for the field test of Levels 3 through AR. For those passages that were not rated by the authors, library lists, book reviews, and similar guides were used in assigning the passages to test levels. The readability estimates described in the preceding section were also a major factor in assigning passages to test levels.

Literal and Inferential Questions

Information gained from reading a passage can be classified as being stated explicitly in the passage or as resulting from the reader operating on explicit statements by applying knowledge of the world and of prose conventions to produce information that is only implicitly included in the text. This popular distinction between explicit and implicit information, or literal and inferential information, divides what is actually a continuum.

Even seemingly literal statements invariably require some level of inference or interpretation, though the inference may be automatic and low level—such as inferring that the message is English, not a code, or that the writer intended to communicate and to be consistent in topic from one phrase to the next. At the other end of the continuum, inferences can be made that use the text only as a starting point for flights of fancy. Assessments of reading comprehension typically encompass a considerable range of the literal-inferential continuum, but appropriately limit the inferences required for choosing correct answers to those inferences that most well-educated readers would agree upon.

In developing the Fourth Edition, questions in Levels 3 and higher were classified as literal if the student could answer by choosing a restatement of something stated explicitly in the passage. Questions that could not be answered by choosing

a restatement were classed as inferential. Although the inferences become progressively more difficult at the higher levels, the aim was to have approximately half of the questions be inferential in all Levels, 3 through 10/12, and a little more than half in Level AR. The actual percentage of inferential questions in each published form of Levels 3 through 10/12 ranges between 48% and 52%. (In each form at each level, either 23, 24, or 25 of the 48 questions are inferential.) The percentage of inferential questions in both forms of Level AR is 58%.

Classifying questions as literal or inferential in Levels 1 and 2 is more problematic, since there is no written question to compare with the text. Questions were classed as inferential if an inference seemed necessary to choose between the correct answer picture and either of the other two. The aim was that about 25% of the questions in Level 1 and about 35% of the questions in Level 2 should be inferential. The actual number and percentage of inferential questions in the published forms are:

Level 1, Form S, 10 questions (26%);
Level 2, Form S, 15 questions (38%);
Level 2, Form T, 13 questions (33%).

The literal/inferential classification of each Comprehension question in the published forms of Levels 1 through 10/12 and AR is given in Appendix A.

Types of Comprehension Answer Choices

In constructing comprehension tests, as in constructing vocabulary tests, it is important to consider the range of strategies that students may use when they do not fully understand the passage and do not know the answer to a question. (See the section "Types of Vocabulary Answer Choices," beginning on page 13, for a discussion of this issue.) For this reason, the *GMRT* Comprehension tests include many questions in which a wrong answer is the longest answer or the shortest answer; and the correct answers are essentially randomly distributed in the various answer positions, so that no one position contains a preponderance of correct answers.

Levels 3 through 10/12 and AR also include two types of wrong answers that are related to more complex strategies that students may apply when they do not fully understand the passage and do not know the answer to a question. These two types of wrong answers are *prior-knowledge* wrong answers and *text-phrase* wrong answers. Each of these types of wrong answers was included only when the desired question could readily accommodate that answer type without distorting the intent of the question or lowering its quality.

- **Prior-Knowledge Wrong Answers.** Prior-knowledge wrong answers are those wrong answers that may be chosen on the basis of what the reader already knows or believes rather than on the basis of information that can be obtained from the passage. Prior-knowledge wrong answers are clearly *wrong* answers; they may be counter to information in the passage, or they may be irrelevant to the passage. They may relate to the content of the passage, or they may simply seem plausible in relation to the question.

However, widely-held beliefs that were not directly contradicted by the passage were not used as wrong answers. A student who frequently chooses prior-knowledge wrong answers may not be using what he or she reads as a basis for revising preconceptions about what the text will say.[43]

- **Text-Phrase Wrong Answers.** Text-phrase wrong answers are those that include a phrase or salient content word that appears in the passage. Most text-phrase wrong answers make sense in relation to the question stem; some, however, simply repeat a particularly salient word or phrase from the passage. Text-phrase wrong answers were included to reveal reliance on a strategy of looking for answers that include words or phrases from the text rather than considering the question carefully and working out an answer by constructing an understanding of the text. Many correct answers also include words or phrases from the text (though many do not), so a student cannot get a good Comprehension score simply by avoiding answers that contain a word or phrase from the text.

Since prior-knowledge and text-phrase wrong answers were not systematically included in the Comprehension questions, a listing by question of these wrong answer types is not provided in this *Technical Report.* Also, as was the case with the Vocabulary wrong answer types, an extensive study of these answer types undertaken for the Third Edition showed that scores based on the types of wrong answers students chose were not very reliable. Many students, in other words, evidently do not choose consistently one or the other of these wrong answer types when they do not know the answer to a question.

Although tallying which of these types of wrong answers a student chooses is not justified as a way to understand a student's low Comprehension score, the authors' clinical experience indicates that these wrong answer types can be useful to keep in mind when working with a student individually and probing for the student's reasoning in choosing answers.[44] Interviews conducted by the authors with students at all the test levels showed that occasional students often do use these unproductive strategies.

Question Difficulty

Since the *GMRT* are designed for general use at specific grade levels, it is desirable for each test level to measure reading achievement accurately over much of the considerable range of achievement among students at the grade level(s) for which the test level is designed.

In addition, on a reading achievement test, it is important that nearly all students find some questions at the beginning of the test that seem to them to be within their capability. Therefore, several questions at the beginning of the test must be quite easy. To compensate for the easy questions at the beginning, and to measure well the achievement of good readers, a number of questions at the end of the test must be quite difficult. And to discourage guessing and keep students at all levels of achievement doing the best they can, the test passages and questions should generally progress gradually in difficulty, although occasional

relatively easy questions are useful for sustaining motivation. In addition to these considerations, the ideal distribution of question difficulties depends on other factors, including the intercorrelations of the questions, the number of answer choices, and the level(s) of ability at which the test should give the most accurate measurement.[45]

The blueprint for each test included a distribution of question difficulties designed to accommodate all these various considerations. Table 6 shows the target percentages for questions with four answer choices.[46] Those percentages were modified appropriately for tests in which the questions had three or five answer choices.

In developing questions for the field test, preliminary estimates of question difficulties were made by comparing new questions to questions of known difficulty from the Third Edition. After the field test, actual question difficulties replaced the estimates,[47] and blueprints of difficulty distributions similar to that in Table 6 were again used as guides in the development of the final test forms.[48] The difficulties (*p*-values) of all questions in each published level and form of the *GMRT* are given in Tables 30–40 in Appendix F.

Table 6. Desired Distribution of Question Difficulties

p-Value[a] Range	Percentage of Questions
96–100	0
91–95	4
86–90	4
81–85	4
76–80	6
71–75	8
66–70	14
61–65	14
56–60	14
51–55	12
46–50	8
41–45	4
36–40	4
31–35	4
Total	100
Average *p*-Value[49]	62

[a] *p*-Value: Percentage of students

Cultural Diversity

Bias Review

All test questions and all Comprehension passages were examined by African American, Asian, Hispanic, and Native American consultants for bias and possible offensiveness. Questions that a consultant thought might be biased or offensive were either eliminated or rewritten.[50] Passages that a consultant thought might be biased or offensive were eliminated. The consultants who examined the questions and passages were:

Laura Tapia Aitken, retired from William Patterson University, Wayne, NJ.
Angie Beauregard, Reading Specialist, Special Education/Bilingual Education, Mendenhall Elementary, Plano, TX.
Frank Ciriza, Evaluation Unit Manager, Assessment, Research and Reporting Team, San Diego City Schools, San Diego, CA.
Paul L. Dauphinais, School Psychologist, Special Education, Turtle Mountain Schools, Belcourt, ND.
Monte Dawson, Director, Monitoring and Evaluation Services, Alexandria City Public Schools, Alexandria, VA.
Roberta Dawson, Reading Specialist, Big Beaver Falls Middle School, Big Beaver Falls Area School District, Beaver Falls, PA.
Eddie Guitierrez, Teacher, Capitol High School, Santa Fe Public Schools, Santa Fe, NM.
Nelda Hobbs, Reading Specialist, Chicago Public Schools, and DePaul University, Chicago, IL.
LaUanah King-Cassell, Principal, St. James and St. John School, Baltimore, MD.
Patty Luke, Title 1 Coordinating Teacher, Seattle School District, Seattle, WA.
Julie Mitchell, Reading Specialist, Travis Middle School, Irving, TX.
Shelby Tallchief, Administrator, Indian Education; Indian Education Unit; Albuquerque Public Schools, Albuquerque, NM.
Fannie H. Tartt, Executive Director for Elementary Instruction, DeKalb County Schools, Decatur, GA.
Dianna J. Uchida, Science Teacher, Dunbar Vocational School, Chicago, IL.
Margaret Winstead, Title 1 Director, Moore Public Schools, Moore, OK.

In addition, prior to the field test, Jane W. Torrey, retired from Connecticut College, reviewed the Word Decoding questions in Levels 1 and 2 and the Letters and Letter-Sound Correspondences questions in Level BR for linguistic accuracy and for questions that might be confusing to speakers of an African American vernacular English. Several questions were revised on the basis of her analysis.

Statistical Bias Analysis (Differential Item Functioning)

Background

There have been many attempts, extending over more than 30 years, to develop statistical methods for detecting bias in tests. The concern for bias-free tests has led to a close examination of the question of what constitutes bias and to a general recognition that almost any test can be *used* in a biased way.[51]

The effort to develop statistical criteria has led to improved methods for detecting test questions that might be biased. These methods generally involve statistical techniques for comparing the success, on a particular test question, of individuals from two different groups. Such differences in success are referred to as *differential item functioning,* or *DIF.* The groups are usually a specified minority group (e.g., African Americans) and a non-minority group (or, often, all the test takers other than members of the minority group). The detection methods involve various ways of statistically comparing the performance on a particular test question of individuals who are otherwise equally able in relation to the ability or knowledge being tested.[52] Typically, the criterion of equal ability is a similar score on the total test.[53] For example, the performance on a particular test question of the individuals from a certain minority group might be compared with the performance on that question of other individuals who were not of that minority group, but who had the same total score on the test. Since individuals of equal ability in the two groups are compared, a difference in performance on the particular test question—the DIF—must be due to something other than level of ability. That something *might* be an unfairness in the test question.[54] Some of the statistical techniques for making such comparisons are more useful under some circumstances and other techniques are more useful under other circumstances.[55]

Demonstration of DIF, even when the comparison involves members of the minority group that are matched for ability with the other test takers, does not by itself demonstrate bias. A test question is biased when it can be further demonstrated that the difference in performance on the question can be attributed to some characteristic of the question that is not related to the domain of knowledge and skills that the test is intended to measure.[56]

A common example of actual bias is a word problem on a mathematics test when the test is given to students who do not read English well. If the test is intended to measure mathematical ability, the ability to read English is not part of what the test is designed to measure, and this question would be biased in this particular *use* of the test. This example also illustrates that bias is as much a matter of test use as of the nature of the questions. If the test were given to students who all read English well enough to understand the question, the question would not be biased in the same way.

Also, it is a characteristic of statistical analyses that they not only miss some cases, they also flag a certain percentage erroneously. On repeated administrations with other groups, some questions flagged for DIF on the first administration would not be flagged again, and some questions that were not flagged the first time would be flagged on a later administration.

Thus, DIF analysis is a useful procedure for calling attention to questions that may be biased. It can not, however, demonstrate that any particular question either is or is not biased, and, demonstration of DIF is only a first step in locating biased questions. The questions that display DIF are ordinarily then analyzed for their content to see if there is a substantive basis for the DIF. There is little evidence, however, that analysis of questions by content experts can produce valid judgments.[57] In the attempt to make sure that the *GMRT* is as free of bias as possible, questions were eliminated whenever there was strong statistical evidence of DIF, even when the authors and minority consultants could locate no basis for bias in the question.

DIF Procedure

As part of the field-test data analysis, the responses of African American students and of Hispanic students to each field-tested question were analyzed through comparison with the responses of a reference group consisting of all other students. The analysis used was the Mantel-Haenszel procedure for detecting differential item functioning.[58] The Mantel-Haenszel results were interpreted using a three-level classification based on chi-square tests of statistical significance and delta (a question difficulty metric)—the well-known system used by Educational Testing Service (ETS).[59] In this three-level system,

- Questions classified as A are considered to display little or no DIF and are considered appropriate for use in test construction.
- Questions classified as B are those that do not meet the criteria for either A or C and are used only if no A question is available to fill the content requirement of the test.
- Questions classified as C meet certain statistical criteria of effect size and statistical significance[60] and are to be used only if content experts consider them essential to meet the test specifications.

As described in the section "Field Testing," beginning on page 32, all *GMRT* field-test forms were administered at two grade levels, usually the grade level for which the test level was designed and the next higher grade. Questions that were ranked C at either grade level or that were ranked B at both grade levels were eliminated from consideration for use in the standardization.

The responses of male students and female students were also separately analyzed, using the same Mantel-Haenszel procedure. Questions with male/female DIF ratings of B or C were often retained, as long as counterbalancing questions were also retained, even though a basis in familiarity related to differences in the experiences of males and females could be proposed. For example, words and passages relating to science, business and industry, and sports were frequently somewhat easier for males than for females; words and passages relating to personal feelings, social relationships, the arts, and family life were frequently somewhat easier for females than for males. These topics were not avoided, but a balance of topics and of DIF ratings was considered in selecting questions and passages for the standardization test forms.

Diversity of Content

For both the field-test and the standardization editions, the passages written for Levels PR, 1, and 2 and the passages selected for the higher levels represented a range of cultural backgrounds. Passages were limited to those written in standard English or with only minor departures from standard English, but passages were chosen so that females and males of various ethnic groups would be represented in the test content—as characters in pictures and passages, and as authors of passages.

Answer Media

At Levels PR through 3, students in both the field testing and the standardization testing marked their answer choices directly in the test booklets. At Levels 4 through 10/12 and AR, students in the field testing and the standardization testing marked their answers on separate answer sheets. However, if a teacher or a school chooses, it is possible for students taking these upper levels of the Fourth Edition to mark their answers directly in the booklets, rather than on separate answer sheets. This option raises the question of the relative ease and accuracy of the two answer modes. In the standardization of the Second Edition of the *GMRT,* an extensive study of this question was undertaken in Grades 4–11. The results showed no consistent advantage of one answer mode over the other.[61]

Field Testing

More than 37,000 students participated in the field test for the Fourth Edition, which provided empirical data for evaluation of all test questions. As described below in the sections "Analysis of Field-Test Question Data," beginning on page 35, and "Question Selection," beginning on page 38, the results of this field test were used as one basis for the selection of questions for all Fourth Edition tests and subtests.

Field-Test Edition

The field-test edition included at least twice as many test forms for each test level as the published Fourth Edition. In addition to the *GMRT* materials being field tested, each field-test form also included an anchor test described below. For Level PR there were three field-test forms—two included all Level PR subtests, and one included only a Listening (Story) Comprehension subtest. This short third form was developed to be certain that there would be an ample pool of questions for this new subtest. For Levels BR and 1 there were two complete field-test forms and for Levels 2 through 10/12 and AR there were four complete forms.

As a basis for adjusting question difficulties for the possibly varying abilities of the groups taking different forms, all field-test forms included an anchor test. For students in Grades K through 7, this anchor test was an appropriate level of the Vocabulary test from Form K of the *Iowa Tests of Basic Skills*®.[62] For students in

Grade 9, the anchor test was the Form K Vocabulary test from Level 15 of the *Iowa Tests of Educational Development®*[63] *(ITED®),* and for students in Grades 10 and 12, the anchor test was the Form K Vocabulary test from Level 16 of the *ITED.*

With the exception of the short third form of Level PR, each field-test form included at least as many questions as the final forms. Both of the two complete field-test forms of Level PR were longer than the final Level PR form. This was also true of Level BR. In particular, extra stories were included for the Basic Story Words subtest for Level BR, since it is a new type of test for the Fourth Edition. In addition, the Comprehension tests at Levels 3 through 10/12 and AR were field tested with more questions than would appear in the final forms. Each field-test form at these levels contained 11 passages, as does each final form. However, the field-test forms contained between 56 and 60 questions, while each final form contains 48. The additional Comprehension questions were included so that individual questions could be eliminated from the set for a particular passage.

All comprehension passages and questions were new; none were reused from the Third Edition. A number of passages and questions that had been field tested for the Third Edition but not used in the published Third Edition were field tested again for the Fourth Edition. These passages and questions had worked well in the field test for the Third Edition; they had not been used in the published forms only because there was an excess of good questions and passages for some of the categories in the test blueprint.

Field-Test Administration

The main field testing was carried out in the fall of 1997. The tests were administered to regular classroom groups. For Levels 1 and higher, each tested classroom group received all the forms of one field-test level. In distributing the tests, the forms were "rotated," so that the form each student received was different from the form received by the preceding student. The testing was done without time limits, and teachers were requested to allow the students to keep working until "all but the slowest-working students" had finished all the questions. This was done so that the data analysis for questions at the end of each test would be based on essentially the same student sample as the data analysis for the earlier questions. The teachers' excellent cooperation in allowing time for all but the very slowest students to finish fulfilled this important condition for comparability of question data.

The field testing for Levels PR and BR was conducted somewhat differently. Because these two levels are orally administered, each form was administered to separate classroom groups that received only that one form. In addition, the teachers only noted the beginning and ending times for their administrations, rather than waiting until "all but the slowest-working students" had finished.

The teachers who administered the tests were encouraged to comment on the testing. These comments were carefully considered and were particularly helpful in revising the directions for preparing for testing and for administering the tests.

The schools selected for participation in the field test represented all regions of the country, large and small school districts, and public and non-public schools. To provide a sound basis for statistical DIF analysis, the sample was intentionally overweighted with schools with a high proportion of African American or Hispanic students. Students were asked to indicate, on a voluntary basis, their gender and ethnic/racial group. At the lower grades, the teacher was asked to provide that information on a voluntary basis. The number of students tested at each grade level in the fall of 1997 is shown in Table 7. Additional samples of 1,607 students in Kindergarten and 1,594 in Grade 1 were tested with Level PR and Level 1, respectively, in the spring of 1998.

Table 7. Fall 1997 Field-Test Sample

Grade	1	2	3	4	5	6	7	10	12	9[a]	Total
Number of Students	3,507	4,303	4,406	4,033	3,921	3,998	3,845	2,388	1,909	1,895	34,205

[a] These Grade 9 students took Level AR. See text.

The grade levels at which each test level was field tested are shown in Table 8.

Table 8. Grades at which Field Tests Were Administered

	Grades Tested	
Test Level	Fall	Spring
PR	1	K
BR	1, 2	
1	2	1
2	2, 3	
3	3, 4	
4	4, 5	
5	5, 6	
6	6, 7	
7/9	7, 10	
10/12	10, 12	
AR	9	

Most test levels were field tested at the grade level for which they were developed and also for the next higher grade. Level 7/9 was administered at Grades 7 and 10, Level 10/12 was administered at Grades 10 and 12, and Level AR was administered at Grade 9. Since Level AR was designed to be more difficult than Level 7/9 but less difficult than Level 10/12, Grade 9 represented an appropriate ability level for obtaining information about the relative difficulty and other statistical characteristics of the questions.

Levels PR and 1 followed a somewhat different testing plan. It would have been too frustrating to the students if Level PR had been administered to kindergarteners or Level 1 to first graders at the beginning of the school year. Therefore, Level PR was administered at Grade 1 in the fall and to Kindergarten students in the spring. Similarly, Level 1 was administered at Grade 2 in the fall and at Grade 1 in the spring.

Analysis of Field-Test Question Data

Field test data for each level were analyzed to provide estimates of question difficulties (*p*-values) and question-test correlations (biserial correlations). At Levels PR and BR, the criterion for the question-test correlations was the Total score. At Levels 1 and 2, the criterion was the Word Decoding score for Word Decoding questions, the Word Knowledge score for Word Knowledge questions, and the Comprehension score for Comprehension questions. At Levels 3 through 10/12 and AR, the criterion was the Vocabulary score for Vocabulary questions and the Comprehension score for Comprehension questions. Correlations among tests and subtests within forms were computed, and DIF analyses were performed, as described in the section "DIF Procedure" on page 31.

Data for the two different grade levels at which most test levels had been given were analyzed separately. However, for the purpose of selecting questions that would result in balanced forms for each test level, the two *p*-values for each question were averaged (via Difficulty Indices[64]), as were the two biserials (via Fisher's *z*). Since the different field-test forms had been rotated within each classroom, each field-test form had been taken by a different group of students at each grade level. In order to obtain difficulty indices that were comparable from form to form, anchor test scores of the groups taking the different forms at any one grade level were used to adjust the indices for differences in group ability.

Questions that had low biserials for the right answer or that had wrong answers with positive biserials were generally eliminated from consideration for the published test forms. Difficult questions that had one wrong answer with a positive biserial were considered, however, if the wrong answer biserial was very low and the right answer biserial was strongly positive.[65] Other things being equal, questions were favored for inclusion in the published forms when all the wrong answers were functional, i.e., were chosen by more than a very small percentage of students.

Statistical characteristics of the questions were by no means the only consideration in choosing questions to be included in the published forms. Other equally important considerations are described in the section "Question Selection" on page 38.

Analysis of Field-Test Administration Time

Teachers who administered the field-test forms were asked to provide information about the time it took students to complete each test or subtest. At Levels PR and BR, teachers were asked to record the times when each testing session started and stopped. At Levels 1 through 10/12 and AR, teachers were asked to record

- The number of students who were still working on the Word Decoding, the Word Knowledge, or the Vocabulary test after 20 minutes;
- The number of students who were still working on the Comprehension test after 30 minutes and again after 40 minutes (For Levels 1 and 2, there was only one such time check—at 35 minutes.);
- The time when each test started;
- The time when "all but the very slowest-working students have finished," at which time the teacher was to stop the test;
- The total number of students taking the test;
- The number of students who did not finish the test.

For Levels PR and BR, the times required for each testing session in the various classrooms were analyzed. The distribution of testing times for each testing session at each grade level and their mean and median were obtained. For Levels 1 through 10/12 and AR, the data were analyzed to determine, separately for each test (Word Decoding, Word Knowledge, Vocabulary, Comprehension) at each grade level, the distribution and the mean and median of

- The percentage of students in the various classroom groups who had finished the Word Decoding, Word Knowledge, or Vocabulary test within 20 minutes;
- The percentage who had finished the Level 1 or Level 2 Comprehension test within 35 minutes;
- The percentage who had finished the Comprehension test at Levels 3 and up within 30 minutes;[66]
- The time required for "all but the very slowest-working students" to finish.

An analysis of "questions not reached,"[67] as well as the teachers' reports of the number of students who did not finish the test, indicated that the teachers had generally been quite conscientious about allowing "all but the very slowest-working students" to finish.

These analyses were used to ensure that the number of questions included in each testing session in the published Fourth Edition of Levels PR and BR could be completed by most classes in about 20–25 minutes. At Levels 1 through 10/12 and AR, the analyses were used to ensure that most students would be able to complete all or nearly all of the questions within the allotted time limits. That the analyses were generally successful in these aims is shown in Table 26, "Form S Completion Rates," on page 65, which shows two different types of data indicating that most students had time to complete each of the Fourth Edition tests.

To meet the goals described in the preceding paragraph, the number of questions in the published forms was reduced,

- In the Levels 1 and 2 Word Decoding tests, to 43 questions from the 45 that had been in the Third Edition;
- In the Level 2 Word Knowledge test, to 43 questions from a planned 45. This was done so that the Word Knowledge and Word Decoding tests would have the same length;
- In the Levels 1 and 2 Comprehension tests, to 39 questions from the 48 that had been in the Third Edition. The new format required more time per question than the format in the Third Edition, but also gave excellent reliability, so, although the number of questions was reduced, high reliability was maintained, as shown in Tables 15 and 16 on pages 55 and 56.

At Levels 3 through 10/12 and AR, the numbers of Vocabulary (45) and Comprehension (48) questions that had been used in the Third Edition continued to be satisfactory from the standpoint of the amount of time required for most students to complete the tests.

Question Selection

Based on information from the field test, questions and passages were selected from those that had been field tested to provide the best possible array of content and difficulty for each test form. For those test levels with alternate forms, questions for the alternate forms of the tests were selected to make the alternate forms similar in many important respects. A balance of test content (such characteristics as test word difficulty, parts of speech, type of passage content, literal and inferential questions), of question difficulty, and of question-test correlation means that the two forms of any test should measure essentially the same thing. As described in the section "Cultural Diversity," questions or passages that consultants thought might be biased or offensive and questions with a strong indication of DIF were not used.

Levels PR and BR

The primary purpose of Level PR is to locate students whose background for reading is sufficiently limited that they may have difficulty in learning to read unless they receive a specially modified instructional program. For that reason, the selection of questions for Level PR was not aimed at making the subtests difficult enough to achieve the best possible discrimination among all the students in a typical group. Instead, the selection of a question depended on

- The authors' judgment as to the importance and teachability of the type of concept represented by the question;
- Sufficient difficulty to be useful in locating students with unusually weak backgrounds;
- A good correlation with the total test.

For the three letter-sound correspondences subtests of Level BR—Initial Consonants, Final Consonants, and Vowels—questions were chosen to provide a wide representation of those correspondences that are commonly taught in beginning reading instruction. Since Level BR may be given at both the beginning and end of Grade 1, questions were selected that test letter-sound correspondences that would be among the first taught as well as some that would be taught later. Similarly, for the Basic Story Words subtest, the stories that were selected provide a test of highly useful words that are typically learned early in reading instruction as well as highly useful words that are typically learned later in the year.

Levels 1 and 2 Word Decoding

For the Word Decoding tests of Levels 1 and 2, questions were chosen to provide a wide representation of those letter-sound correspondences that are commonly taught in the primary grades. Preference was given to questions in which the choice of a wrong answer clearly suggested failure to apply a specific decoding skill. While the letter-sound correspondences tested in Level 1 are generally easier than those tested in Level 2, some of the correspondences tested in Level 1 are more difficult, or are commonly taught later, than some of those in Level 2. That is because individual students learn these correspondences at widely differing rates.

The two forms of Level 2 were matched as closely as possible on the skills tested, given the available question pool, and were also closely matched on the following variables:

- Difficulty and discrimination indices;
- Harris-Jacobson grade levels of the test words;[68]
- Number of multisyllable test words;
- Numbers of nouns, verbs, and adjectives used as test words.

Level 2 Word Knowledge and Levels 3–10/12 and AR Vocabulary

For the Word Knowledge test of Level 2 and the Vocabulary tests of Levels 3 through 10/12 and AR, questions were chosen so that each test form would have an appropriate distribution of

- The grade level ratings given to test words by Harris and Jacobson[69] or, depending on test level, *The Living Word Vocabulary;*[70]
- Question difficulties;
- Parts of speech (See the section "Parts of Speech" on page 10).

The general usefulness of the tested word was also a major consideration.

Forms S and T of each test level were closely matched on the following variables:

- Difficulty and discrimination indices;
- Mean of the grade level ratings given to test words by Harris and Jacobson[71] or, depending on test level, *The Living Word Vocabulary;*[72]
- Ratio of multiword answer choices to single-word answer choices for both right and wrong answers.[73]

In addition, at Levels 3 through 10/12 and AR, the ratio of multiword correct answers to multiword wrong answers was kept roughly equal to $1/(k-1)$, where k is the number of answer choices. This was done so that the length of an answer choice could not be used as a clue to its correctness.

Levels 1–10/12 and AR Comprehension

For the Comprehension tests, passages were selected that would provide an appropriate range of readability and of the authors' ratings of reading difficulty extending above and below the grade level(s) for which the test level was designed. The selection of passages and questions was also guided by the test blueprints for

- Passage content classification;
- Narrative/exposition/setting passage classification;
- Literal/inferential question classification;
- Question difficulty.

Equally important considerations in selecting passages and questions were

- The quality of the passage in terms of the guidelines given in the section "Passage Characteristics in Levels 3 through 10/12 and AR" on page 17.
- The quality of the questions;
  - For Levels 3 through 10/12 and AR, the quality of the questions was judged by how well they followed the guidelines given in the section "Comprehension Questions at Levels 3 through 10/12 and AR" on page 19.
  - For Levels 1 and 2, the quality of the questions was judged by how well they followed the guidelines *relevant to picture choices* in the section "Comprehension Questions at Levels 3 through 10/12 and AR" on page 19 and also the guidelines given in the section "Picture Choices at Levels 1 and 2" on page 17.
- The authors' judgment as to the usefulness of the questions as an indication of a student's understanding of the passage.

The content and the narrative/exposition/setting classifications of passages and the literal/inferential classification of questions are shown in Appendix A.

Forms S and T of each test level were matched on

- Difficulty and discrimination indices;
- Average readability, supplemented by the authors' judgments of reading difficulty;
- Total length in words.

Also, to the extent possible given the available pool of passages, Forms S and T were balanced for numbers of

- Male and female characters and authors;
- Minority group characters and authors.

Question Sequence

At all test levels, Vocabulary and Comprehension questions were sequenced within forms very roughly according to question difficulty. However, a strict difficulty sequence was not followed for three reasons:

- In the field-test forms, restrictions were placed on the otherwise more-or-less random sequence of correct answer positions so as to avoid runs of correct answers in the same position and to distribute correct answer positions about equally within each sequential group of about 12 to 15 questions. In order to meet these same restrictions when assembling standardization forms, it is typically necessary either to move some correct answers from the positions they had occupied in the field test or to place questions in an order that is not strictly according to difficulty. Since the difficulty of a question depends partly on the order of the answer choices, preference was given to retaining the field-test position of the correct answers, although the positions of a few correct answers were changed.
- Comprehension questions about a given passage are seldom equal in difficulty. The most difficult question for a relatively easy passage is likely to be more difficult than an easy question for a more difficult passage. Thus, even if passages were sequenced according to their average difficulty, the questions would not be ordered according to difficulty.
- The most important reason for not ordering Comprehension passages strictly according to difficulty concerns student motivation. In order to arrange passages in a way that would maintain the students' interest and effort, the sequence of topics, types of content, and the subjective as well as objective difficulties of passages were considered.

STANDARDIZATION

A total of nearly 65,000 students from all parts of the country from both public and private schools were tested in the fall 1998 and spring 1999 standardizations of the Fourth Edition. The median fall testing date for students in grades Kindergarten through 12 was November 9. The median spring testing date was April 22. An additional sample of 3,059 students participated in a winter standardization of Level 1.[74] The median opening date for schools in the standardization of Levels PR through 10/12 was August 24.

Sample Selection

Levels PR through 10/12

The standardizations followed a stratified random sampling design. Three stratifying variables were used to classify public school districts across the nation:

- Geographic region (East, Midwest, South, and West);
- District enrollment;
- District socioeconomic status.

All information for these stratification variables was obtained from Quality Education Data, Inc. (QED).[75] School districts were assigned to one of four categories according to their district enrollment (smaller than 2,500; 2,500–9,999; 10,000–24,999; and larger than 24,999). Districts were also assigned to one of four socioeconomic (SES) categories on the basis of the percentage of students receiving a free or reduced price lunch (0%–5%, 6%–20%, 21%–40%, and more than 40%). These categories were used to build a 64-cell matrix of public school districts representing each combination of region, size, and SES. Private (including parochial) schools were grouped in an additional 65th cell.

Public school districts were randomly chosen from each of the 64 public school cells. These school districts were then contacted by representatives of the publisher and invited to participate. When a school district declined to participate, another school district was randomly chosen from that sampling cell and invited to participate. Schools within a district were typically selected by the district with the understanding that they should approximate as closely as possible the district's average reading achievement and racial/ethnic composition.

National demographics determined the proportion of students that should be included in each of the 64 cells. For larger-sized cells, no more than two classrooms, or about 50 students, in each of the grades Kindergarten through 12 were requested from each school district. For smaller-sized cells, the number of students requested was roughly proportional to the size of the cell.

Private and parochial schools for a 65th cell were randomly selected from two sources. The private schools were selected from the QED database, and the parochial schools were selected from Mahar.[76] The chosen schools were contacted by representatives of the publisher and invited to participate, until enough students were obtained at each grade to approximate the desired size of the cell—roughly 10% of the total sample. When a school declined to participate, another school was randomly chosen as a replacement.

Including the winter standardization of Level 1, a total of 65,059 students in 301 schools in 45 states took part in the standardization of Levels PR through 10/12. The school districts and schools that participated in the standardization are listed alphabetically by state in Appendix B. Table 9 shows, by grade and test level, the numbers of students in the fall, winter, and spring standardizations.

Table 9. Standardization Sample Sizes

Grade	Test Level	Fall *N*	Winter *N*	Spring *N*
K	PR			2,039
1	PR	1,884		
1	BR	2,645		2,507
1	1		3,059	2,203
2	2	3,680		3,970
3	3	3,584		3,419
4	4	3,126		3,294
5	5	2,770		3,126
6	6	2,538		2,440
7	7/9	2,445		1,873
8	7/9	2,228		1,964
9	7/9	1,450		2,193
10	10/12	1,327		1,634
11	10/12	986		872
12	10/12	862		941
Total		29,525	3,059	32,475
Grand Total				65,059

To obtain a nationally representative sample, the number of students included in each of the cells in the sampling matrix should be proportional to the national totals in these cells as aggregated from the QED database. Schools were randomly selected to fill the cells of the sampling matrix with numbers of students that would approximate the proportions in the cells determined by national demographics, but the final standardization sample did not completely correspond to those proportions. As a result, weighting of the numbers in the sample cells was required. The weightings of cases for the various cells were computed so as to make the total numbers of cases in the cells more nearly proportional to the numbers represented by national demographics.

Table 10 shows the percentage of students in each category of each stratification variable. The first column shows the percentages for the entire U.S. student population as determined from the QED database.[77] The other columns show the

percentages of students in the weighted sample and the actual percentages of students in the unweighted sample. Data on other characteristics of the standardization sample were obtained by a questionnaire given to each participating school. Aggregated data obtained from this questionnaire are given in Appendix E.

Table 10. Percentage of Students by Region, District Size, and SES Categories

	Population	Weighted Sample			Unweighted Sample		
Region		**Fall**	**Winter**	**Spring**	**Fall**	**Winter**	**Spring**
East	19.4	18.8	18.5	17.8	16.9	9.5	13.2
Midwest	25.2	22.4	25.6	24.9	20.8	21.9	22.6
South	32.4	34.9	31.9	33.2	38.2	34.4	41.4
West	23.0	23.9	24.0	24.1	24.1	34.2	22.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
District Size							
Small	20.4	22.3	21.7	22.1	38.6	21.2	26.2
Small–Medium	32.6	29.4	36.2	32.6	21.3	37.8	40.4
Medium–Large	19.3	19.4	16.8	19.6	23.8	13.9	14.4
Large	27.7	28.9	25.3	25.7	16.3	27.1	19.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SES							
Low	34.6	31.8	33.5	31.7	18.2	39.3	32.9
Low–Average	27.1	25.4	27.9	28.1	19.7	24.0	22.5
High–Average	25.6	27.5	25.8	24.4	48.2	28.0	29.8
High	12.7	15.3	12.8	15.8	13.9	8.7	14.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Public	90.0	89.9	90.0	90.0	96.1	74.2	89.4
Private/Parochial	10.0	10.1	10.0	10.0	3.9	25.8	10.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Level AR

The purpose of Level AR is to provide community colleges and training programs with a reading test suitable for screening entering students. In concert with other assessment, Level AR is intended to locate students who need to develop their reading skills before they begin regular classes. To be useful for this purpose, Level AR norms should reflect the range of reading skill typical of students entering the community college or training programs. For that reason, norms for Level AR were obtained only at one time of year—in the fall, when the majority of community college students first enter the college program. It was anticipated, however, that the most useful norms would usually be local norms that the particular community college or training program would establish. (A related procedure that correlates test scores with success in a particular program is described in the section "Which Individuals Meet the Requirements of Your Program?" in the *Manual for Scoring and Interpretation* for Level AR.)

Norms for Level AR were established by administering Level AR in the fall[78] to 2,824 students in 43 community colleges in 24 states. Participating community

colleges were randomly selected from the database in *Guidance Information System.*[79] The participating community colleges are listed in Appendix D. The community college students were enrolled in regular for-credit courses prerequisite to the freshman introductory English course. The community college standardization data were aggregated without weighting.

Equating Studies

In addition to the students who were tested in the standardizations, another 30,043 students participated in three equating studies. The equating samples came from a geographically dispersed, heterogeneous sample of schools across the United States. Some of the schools in the norming sample also participated in the equating studies. The school districts and schools that participated in the equating studies are listed alphabetically by state in Appendix C.

Each student in an equating study was tested twice within a three-week period.

- For the equating of adjacent test levels in the fall of 1998, each student was tested with two adjacent levels of Form S of the Fourth Edition—a level designed for the student's grade and the next lower level.

- For the equating of alternate forms in the spring of 1999, each student was tested with both Fourth Edition forms (Forms S and T) of the level designed for the student's grade.

- For the equating of the Third and Fourth Editions in the spring of 1999, each student was tested with comparable levels of Form K of the Third Edition and Form S of the Fourth Edition.

Each pair of tests was given in one order in some schools and in the reverse order in others; approximately half the students took one test first and half took the other test first. In schools that also participated in the standardization, the students took Form S of the level for their grade before taking the other level or form.

Equating of Levels

In the equating of adjacent test levels, students were tested twice with the Fourth Edition—once with the level of Form S designed for their grade and once with the next lower level of Form S. The out-of-level test used was the next lower level rather than the next higher level because the equating of levels was done in the fall (except for Level BR vs. Level 1). The next higher level might have been too difficult for students at that time—at the beginning of the school year. The adjacent levels study was designed to provide correlations between corresponding tests in the two levels and to provide data for the scaling of the Fourth Edition. The plan for this equating is shown in Table 11. Test levels in the same row in Table 11 were equated at the grade level shown for that row. Table 11 shows the sample sizes for this equating study, the maximum possible raw score on the test, the raw score means, standard deviations, and correlations between adjacent levels.

Table 11. Equating of Test Levels: Sample Sizes and Raw Score Means, Standard Deviations, and Correlations

Grade and Test	*N*	Max Score	Mean	*SD*	Max Score	Mean	*SD*	*r*
Grade 1			**Level BR**			**Level PR**		
Total	600	70	46.1	13.6	90	76.9	10.0	0.74
Grade 1[a]			**Level 1**			**Level BR**		
Total	1,940	82	50.5	17.9	70	56.8	11.4	0.83
Grade 2			**Level 2**			**Level 1**		
Word Decoding	364	43	31.4	10.0	43	35.8	7.4	0.88
Word Knowledge[b]		43	27.4	9.7	43	35.8	7.4	0.76
Comprehension		39	29.1	8.0	39	32.5	7.0	0.80
Total		125	87.9	25.7	82	68.3	13.6	0.88
Grade 3			**Level 3**			**Level 2**		
Vocabulary[c]	425	45	26.3	10.1	43	36.2	7.9	0.78
Vocabulary[d]		45	26.3	10.1	43	32.7	8.4	0.86
Comprehension		48	28.8	10.5	39	31.9	6.2	0.72
Total		93	55.1	19.8	125	100.8	20.6	0.88
Grade 4			**Level 4**			**Level 3**		
Vocabulary	1,083	45	25.8	9.5	45	32.4	9.1	0.87
Comprehension		48	27.5	10.3	48	32.9	9.8	0.82
Total		93	53.3	18.9	93	65.2	17.9	0.90
Grade 5			**Level 5**			**Level 4**		
Vocabulary	1,086	45	24.2	9.6	45	30.9	9.5	0.87
Comprehension		48	27.0	10.6	48	30.8	10.7	0.85
Total		93	51.1	19.1	93	61.8	19.2	0.91
Grade 6			**Level 6**			**Level 5**		
Vocabulary	713	45	23.4	9.3	45	27.6	9.2	0.86
Comprehension		48	29.0	9.7	48	29.2	10.3	0.85
Total		93	52.4	17.9	93	56.8	18.4	0.91
Grade 7			**Level 7/9**			**Level 6**		
Vocabulary	824	45	23.6	9.2	45	28.1	9.8	0.88
Comprehension		48	28.9	10.1	48	32.5	9.6	0.82
Total		93	52.5	18.2	93	60.7	18.4	0.90
Grade 8			**Level 7/9**			**Level 6**		
Vocabulary	1,030	45	29.0	9.4	45	32.9	9.3	0.89
Comprehension		48	32.8	10.1	48	35.7	9.3	0.80
Total		93	61.8	18.3	93	68.6	17.4	0.90
Grade 9			**Level 7/9**			**Level 6**		
Vocabulary	832	45	31.8	8.3	45	35.3	7.8	0.87
Comprehension		48	35.7	8.9	48	37.6	7.7	0.80
Total		93	67.4	16.0	93	72.9	14.4	0.90
Grade 10			**Level 10/12**			**Level 7/9**		
Vocabulary	539	45	25.8	8.4	45	33.2	7.8	0.86
Comprehension		48	28.1	9.7	48	36.7	8.6	0.75
Total		93	54.0	17.0	93	69.8	15.0	0.87
Grade 11			**Level 10/12**			**Level AR**		
Vocabulary	202	45	23.0	7.5	45	26.1	8.3	0.66
Comprehension		48	21.9	8.9	48	24.3	10.5	0.63
Total		93	44.9	14.4	93	50.4	16.2	0.69
Total *N*	9,638							

[a] Winter testing
[b] Level 1 test is Word Decoding
[c] Level 2 test is Word Decoding
[d] Level 2 test is Word Knowledge

An equipercentile procedure was used to define equivalent raw scores for corresponding tests and for Total on each pair of test levels. Smoothing of the equipercentile relationship was done using a fifth-degree polynomial, with final smoothing reviewed by graphing the scores.

The correlations between adjacent levels support the view that each level of the tests from Level 1 up through Level 10/12 measures essentially the same reading abilities as the adjacent level. At these levels, the correlations between test levels were all 0.87 or higher for Total. The correlation between Level PR and BR was much lower, as would be expected; these two tests were designed to serve two different purposes and to test different knowledge and skills. In the winter testing with Level 1 and Level BR, the correlation for Total was lower than that between Level 2 and Level 1. This also would be expected, since Level BR does not include a Comprehension test.

The correlation between the Level 2 Word Knowledge test and the Level 3 Vocabulary test was high (0.86). Evidently the different format of these two tests did not make much, if any, difference in the type of achievement measured by the two tests. The correlation between the Level 2 and Level 3 Comprehension tests was relatively low. The Level 2 Comprehension test is shorter (39 questions) and was relatively easy for the Grade 3 students, resulting in a somewhat restricted range of scores. In addition, as specified in the blueprint, the Level 2 Comprehension test included a smaller proportion of inferential questions. In this case, the different formats of the two tests may also have contributed to the lower correlation.

Equating of Third and Fourth Editions

The equating of the Third and Fourth Editions was designed to provide data to permit the conversion of scores on one edition to scores on the other, so schools that had been using the Third Edition could use the Fourth Edition without losing the continuity of year-to-year score comparisons. The plan for this equating is shown in Table 12. Test levels were equated at the grade level shown in the same row as the test level designations. Table 12 shows the sample sizes for this equating study, the maximum possible raw score on the test, the raw score means, standard deviations, and correlations between corresponding tests in the two editions. Level AR was not equated with a Third Edition test since there was no equivalent level in the Third Edition. The correlations between the two editions are high and very similar to the alternate-form reliabilities shown in Table 13 on page 50, except for Level 10/12, for which the sample sizes were too small and the scores too low for the correlations to be very meaningful. The high correlations indicate that the two editions measure essentially the same reading abilities.

An equipercentile procedure was used to define equivalent raw scores for each test and for Total on corresponding test levels of the two editions. Smoothing of the equipercentile relationship was done using a fifth-degree polynomial, with final smoothing reviewed by graphing the scores. Tables of corresponding PRs for converting scores on one edition to scores on the other are available through Riverside Customer Service in the publication *GMRT Third and Fourth Editions Score Comparisons.*

Table 12. Equating of Third and Fourth Editions: Sample Sizes and Raw Score Means, Standard Deviations, and Correlations

				Third Edition			Fourth Edition			
Level	Grade	Test	*N*	Max Score	Mean	*SD*	Max Score	Mean	*SD*	*r*
PRE vs PR	K	Total	1,032	98	80.4	14.1	90	66.2	14.8	0.89
R vs BR	1	Total	423	60	52.1	9.4	70	57.4	12.5	0.89
1 vs 1	1	Vocabulary[a]	1,091	45	32.9	9.5	43	30.2	9.4	0.90
		Comprehension		46	33.5	10.0	39	26.8	8.8	0.88
		Total		91	66.4	18.8	82	57.0	17.4	0.93
2 vs 2[b]	2	Vocabulary[a]	1,159	45	31.3	8.9	43	33.2	9.2	0.88
		Comprehension		46	36.1	8.9	39	30.3	7.3	0.85
		Total		91	67.3	17.0	125	92.6	23.9	0.92
3 vs 3	3	Vocabulary	1,133	45	30.2	9.5	45	30.2	9.3	0.89
		Comprehension		48	32.0	10.3	48	31.8	10.1	0.85
		Total		93	62.2	19.0	93	62.0	18.5	0.92
4 vs 4	4	Vocabulary	1,195	45	28.2	9.6	45	28.6	9.7	0.89
		Comprehension		48	29.6	10.2	48	28.9	10.8	0.86
		Total		93	57.8	18.9	93	57.5	19.5	0.92
5/6 vs 5	5	Vocabulary	1,273	45	25.6	9.6	45	25.7	9.4	0.88
		Comprehension		48	27.3	10.5	48	28.0	10.5	0.84
		Total		93	52.9	19.1	93	53.7	18.8	0.91
5/6 vs 6	6	Vocabulary	1,077	45	29.9	9.5	45	26.0	9.2	0.89
		Comprehension		48	30.1	10.2	48	31.8	9.0	0.82
		Total		93	60.0	18.6	93	57.8	17.1	0.91
7/9 vs 7/9	7	Vocabulary	497	45	24.7	8.8	45	25.4	9.3	0.89
		Comprehension		48	27.0	9.3	48	30.7	9.6	0.83
		Total		93	51.7	17.2	93	56.1	17.8	0.91
7/9 vs 7/9	8	Vocabulary	415	45	28.9	9.4	45	29.6	9.1	0.90
		Comprehension		48	30.2	9.5	48	34.0	8.8	0.84
		Total		93	59.2	17.9	93	63.5	16.9	0.92
7/9 vs 7/9	9	Vocabulary	169	45	29.2	9.5	45	30.4	9.1	0.89
		Comprehension		48	31.9	8.0	48	35.7	7.8	0.84
		Total		93	61.1	16.7	93	66.1	15.8	0.93
10/12	10	Vocabulary	56	45	17.1	7.6	45	18.6	8.4	0.77
		Comprehension		48	15.8	7.0	48	16.9	7.0	0.71
		Total		93	32.8	13.4	93	35.4	14.3	0.82
10/12	11	Vocabulary	43	45	19.0	10.1	45	18.3	9.9	0.90
		Comprehension		48	19.4	8.5	48	20.8	9.2	0.78
		Total		93	38.4	17.0	93	39.1	17.1	0.92
10/12	12	Vocabulary	53	45	19.4	9.1	45	18.6	9.1	0.81
		Comprehension		48	22.2	7.2	48	23.3	7.5	0.58
		Total		93	41.6	14.7	93	41.9	14.7	0.85
		Total *N*	9,616							

[a] Labeled Word Decoding in the Fourth Edition.
[b] Word Knowledge in the Fourth Edition had no equivalent in the Third Edition.

Equating of Forms

For Levels 2 through 10/12 and AR, there are two forms, S and T, of the Fourth Edition. The equating of these two forms at each level was designed to provide alternate-form reliability coefficients and equivalent derived scores for the two forms. The plan for this equating is shown in Table 13. Table 13 shows the sample sizes for this equating study, the maximum possible scores, the raw score means, standard deviations, and correlations between Forms S and T.

Table 13. Equating of Alternate Forms: Sample Sizes and Raw Score Means, Standard Deviations, and Correlations (Alternate-Form Reliabilities)

				Max	Form S		Form T		
Level	**Grade**	**Test**	***N***	**Score**	**Mean**	***SD***	**Mean**	***SD***	***r***
2	**2**	Word Decoding	1,163	43	33.5	9.1	33.9	9.0	0.92
		Word Knowledge		43	29.5	9.3	29.3	9.5	0.90
		Comprehension		39	30.3	7.3	30.1	7.4	0.86
		Total		125	93.4	24.2	93.2	24.2	0.95
3	**3**	Vocabulary	1,470	45	29.2	9.9	29.5	10.2	0.90
		Comprehension		48	30.8	10.5	30.4	10.9	0.87
		Total		93	60.0	19.3	59.8	20.2	0.93
4	**4**	Vocabulary	1,351	45	28.6	9.6	28.8	9.7	0.89
		Comprehension		48	29.1	10.6	30.5	10.6	0.86
		Total		93	57.7	19.1	59.2	19.4	0.92
5	**5**	Vocabulary	1,408	45	27.4	9.6	27.4	9.5	0.89
		Comprehension		48	29.9	10.2	30.0	10.5	0.86
		Total		93	57.3	18.9	57.4	19.1	0.93
6	**6**	Vocabulary	1,401	45	27.0	9.5	27.0	9.1	0.87
		Comprehension		48	31.8	9.2	31.0	9.7	0.82
		Total		93	58.8	17.4	58.0	17.8	0.91
7/9	**7**	Vocabulary	959	45	28.1	8.4	28.1	8.3	0.87
		Comprehension		48	32.5	9.1	31.8	10.0	0.83
		Total		93	60.7	16.1	59.9	17.2	0.90
7/9	**8**	Vocabulary	1,044	45	30.5	9.1	30.5	9.1	0.89
		Comprehension		48	34.1	9.2	34.4	9.7	0.83
		Total		93	64.6	17.3	64.9	17.8	0.90
7/9	**9**	Vocabulary	737	45	32.6	8.2	32.0	8.3	0.83
		Comprehension		48	35.6	9.5	35.2	9.7	0.80
		Total		93	68.1	16.3	67.3	16.4	0.88
10/12	**10**	Vocabulary	459	45	27.7	8.5	25.7	9.0	0.86
		Comprehension		48	28.4	10.6	27.3	10.8	0.83
		Total		93	56.0	17.9	53.0	18.5	0.90
10/12	**11**	Vocabulary	148	45	32.5	7.7	31.3	9.0	0.75
		Comprehension		48	33.6	10.1	32.4	10.2	0.74
		Total		93	66.1	16.2	63.7	18.1	0.81
10/12	**12**	Vocabulary	67	45	27.6	9.2	27.2	8.7	0.88
		Comprehension		48	24.8	11.0	25.2	11.1	0.89
		Total		93	52.4	18.6	52.4	18.2	0.93
AR	**CC**[a]	Vocabulary	734	45	30.6	9.3	31.1	8.9	0.89
		Comprehension		48	34.2	9.4	33.2	9.5	0.83
		Total		93	64.9	17.7	64.3	17.3	0.91
		Total *N*	10,941						

[a] Community college

An equipercentile procedure was used to relate the Word Decoding, Word Knowledge, Vocabulary, Comprehension, and Total raw scores on Form T to the corresponding raw scores on Form S of the same test level. Smoothing of the equipercentile relationship was done using a fifth-degree polynomial, with final smoothing reviewed by graphing the scores. Derived scores corresponding to raw scores on Form S were then assigned to equivalent raw scores on Form T.

The alternate-forms correlations (alternate-forms reliabilities) are all quite high. Except for Grades 9 and 11, the Total score reliabilities are all 0.90 or higher. The anomalous correlations at Grade 11 and the high correlations at Grade 12 are based on relatively small samples. The alternate-form reliabilities for the individual tests (e.g., Vocabulary, Comprehension) are also excellent. All, except those at Grade 11, are 0.80 or higher, with a median of 0.88.

Norms Development

Data from the equating of levels, described in the section "Equating of Levels" on page 46, provided the basis for developing a new scale for the extended scale scores (ESSs) of the Fourth Edition. The scale was developed according to item response theory (IRT), using the Rasch model. All analyses were performed using the WINSTEPS computer program.[80] This program provides parameter estimates and develops a raw score-to-ability relationship which is an essential component of the scaling process. The program can fix item (i.e., question)[81] parameters to preassigned values, ensuring that the final estimated parameters are on the required scale. To develop the scale,

1. A concurrent calibration that included all test items in Levels 3 through 10/12 and AR was run.
2. A separate calibration was run by extracting all the Level 3 item difficulties from the initial calibration and fixing them in a calibration involving Levels 1, 2, and 3.
3. All the Level 1 item difficulties were anchored in a joint calibration of Level BR and Level 1.
4. The Level BR item difficulties were anchored in a calibration of Levels PR and BR, thus putting all item difficulties on the same scale.

After the scale was developed, it was centered so that the median value for fall of Grade 5 would have an extended scale score (ESS) of 500. The scale was transformed to have a mean of 500 and a standard deviation of 30 by using the equation

$$ESS = 30(\theta) + 500$$

where θ is the ability estimate for each raw score.

The percentile rank (PR) associated with each ESS was calculated separately for Word Decoding, Word Knowledge/Vocabulary, Comprehension, and Total using cumulative weighted frequency distributions of the ESSs. This was done separately for the fall and spring standardizations. The PRs were based on the midpoints of the ESS frequencies. From these PRs the corresponding normal curve equivalents (NCEs) and stanines were assigned.

In addition to determining PR, NCE, and stanine norms for the fall and spring standardization dates (which were quartermonths 10 and 31), PR, NCE, and stanine norms were developed for all other quartermonths[82] by linear interpolation. Norms for quartermonths between quartermonth 31 of one grade and quartermonth 10 of the next higher grade were interpolated between those quartermonths.

The grade equivalent (GE) scale was set, separately for Word Decoding, Word Knowledge/Vocabulary, Comprehension, and Total, by first examining the within-grade weighted frequency distributions of the ESSs. The ESS associated with a PR of 50 in the fall was assigned a GE of G.2, where G is the grade level. The ESS associated with a PR of 50 in the spring was assigned a GE of G.7. From these relationships, a complete ESS to GE scale was created through linear interpolation. Once the GE scale was established, the GEs were applied, through their corresponding ESSs, to raw scores for other test levels.

The GE scale is based on the achievement of students as they progress through the grades of the public and related private and parochial schools. Although there are differences from district to district and from state to state, the core of the curriculum is generally fairly uniform. (It is generally possible, for example, for students to transfer from one district or state to the same grade in another.) And, although there is some school leaving, most students are in school from Grade 1 through Grade 12. Many high-school graduates, however, do not continue their schooling. Those who do, continue in a wide variety of studies and training programs. Thus, GEs beyond 12.9 would not be "grade equivalent" scores at all. They would not have the same meaning as GEs from 1.0 to 12.9 and so would undoubtedly be misleading. Therefore, all raw scores that would correspond to GEs higher than 12.9 are assigned a GE of PHS (post high school), meaning simply that they are higher than 12.9.

Since it would have been inappropriate to administer Level PR to students earlier than the spring of Kindergarten, there is only one data point (spring of Kindergarten) below Grade 1. As a result, there is no way of knowing how well students would have done earlier in Kindergarten and so no meaningful basis for extending the GE scale down through the Kindergarten year. All GEs lower than 1.0 are assigned a GE of K (Kindergarten), without a decimal for the month.

TEST CHARACTERISTICS

Item Difficulty

Data on item (i.e., question)[83] difficulties were obtained from analyses of the responses of the students in the standardization samples, weighted as described in the section "Sample Selection" on page 43. The *p*-values (proportion of correct responses) for Form S, which was the form used in the standardization, are shown in Tables 30–35 and 40 in Appendix F.

For Form T, *p*-values were estimated by using the data from the equating of alternate forms. An item analysis was performed on each of the Form T tests, and the resulting item difficulties were adjusted to reflect the national raw score mean for Form T. These estimated item difficulties are shown in Tables 36–40 in Appendix F.

Reliability

Reliability and error of measurement data for the tests of the Fourth Edition are presented in the subsection that follows. Various validity data are presented in the four subsequent subsections. Of course, most of the other information in this *Technical Report* is also materially related to validity. The guidelines and procedures for item and test development, the field testing and field-test data analysis, the data from the equating studies, and the data on reliability all have an important bearing on the validity of the tests.

Reliability Indices and Standard Errors of Measurement

From the item analyses described in the "Item Difficulty" section above, Kuder-Richardson Formula 20 (K-R 20) reliability coefficients were computed. These reliabilities are shown in Tables 14–16, along with maximum possible scores, average item difficulties (*p*-values), raw score means, standard deviations (*SD*s), and standard errors of measurement (*SEM*s). In addition to the K-R 20 reliability coefficients in Tables 14–16, alternate-forms reliability coefficients are given in Table 13.

Standard errors of measurement expressed in ESSs, along with K-R 20s and the means, standard deviations, and medians of the ESSs, are shown for Form S of Levels PR through 10/12 and AR in Table 17 and for Form T of Levels 2 through 10/12 and AR in Table 18. Standard errors of measurement in NCE units are given in Table 19.

Table 14. Raw Score Summary Statistics, Levels PR and BR

	Max Score	Average *p*-Value	Mean	*SD*	*SEM*	K-R 20
Level PR, Spring of Kindergarten						
Subtest 1, Literacy Concepts	20	0.82	16.44	3.11	1.4	0.80
Subtest 2, Oral Language/Phonological Awareness	20	0.67	13.32	4.36	1.8	0.83
Subtest 3, Letters and Letter-Sound Correspondences	30	0.82	24.55	4.85	1.6	0.89
Subtest 4, Listening (Story) Comprehension	20	0.66	13.23	4.32	1.8	0.82
Total	90	0.75	67.54	15.07	4.0	0.93
Level PR, Fall of Grade 1						
Subtest 1, Literacy Concepts	20	0.86	17.15	2.63	1.2	0.79
Subtest 2, Oral Language/Phonological Awareness	20	0.69	13.79	4.24	1.7	0.83
Subtest 3, Letters and Letter-Sound Correspondences	30	0.87	26.04	3.73	1.3	0.87
Subtest 4, Listening (Story) Comprehension	20	0.71	14.29	4.20	1.8	0.81
Total	90	0.79	71.27	13.39	3.5	0.93
Level BR, Fall of Grade 1						
Subtest 1, Letter-Sound Correspondences: Initial Consonants	15	0.64	9.56	3.64	1.5	0.84
Subtest 2, Letter-Sound Correspondences: Final Consonants	15	0.57	8.60	3.62	1.5	0.83
Subtest 3, Letter-Sound Correspondences: Vowels	15	0.46	6.85	3.65	1.5	0.82
Subtest 4, Basic Story Words	25	0.66	16.60	5.74	2.0	0.88
Total	70	0.59	41.61	15.09	3.4	0.95
Level BR, Spring of Grade 1						
Subtest 1, Letter-Sound Correspondences: Initial Consonants	15	0.84	12.53	2.34	0.9	0.84
Subtest 2, Letter-Sound Correspondences: Final Consonants	15	0.74	11.05	3.13	1.3	0.84
Subtest 3, Letter-Sound Correspondences: Vowels	15	0.71	10.60	3.30	1.4	0.83
Subtest 4, Basic Story Words	25	0.86	21.51	3.70	1.2	0.89
Total	70	0.80	55.69	11.84	2.6	0.95

Table 15. Form S Raw Score Summary Statistics, Levels 1–AR

		Fall					Spring				
	Max Score	Average *p*-Value	Mean	*SD*	*SEM*	K-R 20	Average *p*-Value	Mean	*SD*	*SEM*	K-R 20
Level 1, Grade 1											
Word Decoding	43						0.63	26.92	10.37	2.5	0.94
Comprehension	39						0.63	24.42	9.28	2.5	0.93
Total	82						0.63	51.34	19.75	3.9	0.96
Level 2, Grade 2											
Word Decoding	43	0.65	27.90	10.65	2.6	0.94	0.75	32.40	9.45	2.3	0.94
Word Knowledge	43	0.55	23.81	10.14	2.9	0.92	0.66	28.34	9.91	2.6	0.93
Comprehension	39	0.65	25.23	8.95	2.5	0.92	0.75	29.14	7.64	2.2	0.92
Total	125	0.62	76.94	28.25	4.9	0.97	0.72	89.88	24.83	4.3	0.97
Level 3, Grade 3											
Vocabulary	45	0.58	26.27	10.12	2.7	0.93	0.64	28.95	10.07	2.7	0.93
Comprehension	48	0.57	27.19	11.20	3.0	0.93	0.64	30.56	10.65	2.8	0.93
Total	93	0.57	53.46	20.37	4.1	0.96	0.64	59.51	19.88	4.0	0.96
Level 4, Grade 4											
Vocabulary	45	0.58	25.93	9.64	2.7	0.92	0.62	28.08	9.80	2.8	0.92
Comprehension	48	0.57	27.42	10.59	2.8	0.93	0.63	30.05	10.62	2.8	0.93
Total	93	0.57	53.35	19.13	3.8	0.96	0.63	58.13	19.49	3.9	0.96
Level 5, Grade 5											
Vocabulary	45	0.54	24.27	9.49	2.8	0.91	0.58	25.93	9.59	2.9	0.91
Comprehension	48	0.57	27.37	10.44	3.0	0.92	0.61	29.23	10.21	2.9	0.92
Total	93	0.56	51.64	18.64	4.2	0.95	0.59	55.16	18.68	4.2	0.95
Level 6, Grade 6											
Vocabulary	45	0.55	24.75	9.56	2.9	0.91	0.58	26.18	9.73	2.8	0.92
Comprehension	48	0.59	28.44	9.78	2.9	0.91	0.62	29.84	10.00	2.8	0.92
Total	93	0.57	53.19	17.93	4.0	0.95	0.60	56.02	18.42	4.1	0.95
Level 7/9, Grade 7											
Vocabulary	45	0.52	23.38	9.46	3.0	0.90	0.55	24.63	9.38	3.0	0.90
Comprehension	48	0.58	27.84	10.03	3.0	0.91	0.60	28.83	9.98	3.0	0.91
Total	93	0.55	51.22	18.11	4.4	0.94	0.57	53.46	18.04	4.4	0.94
Level 7/9, Grade 8											
Vocabulary	45	0.58	25.98	9.57	2.9	0.91	0.60	26.82	9.62	2.9	0.91
Comprehension	48	0.63	30.27	9.64	2.9	0.91	0.65	31.27	9.67	2.9	0.91
Total	93	0.60	56.25	17.90	4.0	0.95	0.62	58.09	18.14	4.1	0.95
Level 7/9, Grade 9											
Vocabulary	45	0.62	27.96	9.44	2.7	0.92	0.64	28.69	9.15	2.7	0.91
Comprehension	48	0.67	32.39	9.88	2.6	0.93	0.70	33.51	9.39	2.7	0.92
Total	93	0.65	60.35	17.66	3.9	0.95	0.67	62.20	17.50	3.9	0.95
Level 10/12, Grade 10											
Vocabulary	45	0.51	22.86	9.42	3.0	0.90	0.52	23.34	9.46	2.8	0.91
Comprehension	48	0.53	25.35	10.33	3.1	0.91	0.55	26.29	10.62	2.8	0.93
Total	93	0.52	48.21	19.00	4.2	0.95	0.53	49.63	19.12	4.3	0.95
Level 10/12, Grade 11											
Vocabulary	45	0.53	23.84	9.43	3.0	0.90	0.54	24.50	9.65	2.9	0.91
Comprehension	48	0.56	26.98	10.48	3.0	0.92	0.58	27.65	10.61	3.0	0.92
Total	93	0.55	50.82	19.00	4.2	0.95	0.56	52.15	18.96	4.2	0.95
Level 10/12, Grade 12											
Vocabulary	45	0.55	24.95	9.57	2.9	0.91	0.57	25.57	9.83	2.8	0.92
Comprehension	48	0.59	28.24	10.43	2.8	0.93	0.62	29.84	10.38	2.7	0.93
Total	93	0.57	53.19	18.28	4.1	0.95	0.60	55.41	18.91	4.2	0.95
Level AR, CC[a]											
Vocabulary	45	0.64	29.02	8.61	3.0	0.88					
Comprehension	48	0.71	33.88	8.49	2.8	0.89					
Total	93	0.68	62.90	15.84	4.2	0.93					

[a] Level AR is designed as a test for students entering community college or a training program at a similar level. Norms for Level AR were therefore obtained for first year students at the beginning (in the fall) of a community college program. The norms should apply also to students entering community college at other times of the year.

Table 16. Form T Raw Score Summary Statistics, Levels 2–AR

	Max Score	Fall Average *p*-Value	Fall Mean	Fall *SD*	Fall *SEM*	Fall K-R 20	Spring Average *p*-Value	Spring Mean	Spring *SD*	Spring *SEM*	Spring K-R 20
Level 2, Grade 2											
Word Decoding	43	0.65	28.13	10.76	2.6	0.94	0.76	32.59	9.43	2.3	0.94
Word Knowledge	43	0.54	23.40	10.14	2.7	0.93	0.65	28.11	10.04	2.7	0.93
Comprehension	39	0.64	24.93	9.10	2.6	0.92	0.74	28.89	7.75	2.3	0.91
Total	125	0.61	76.46	28.51	4.9	0.97	0.72	89.59	24.83	4.3	0.97
Level 3, Grade 3											
Vocabulary	45	0.59	26.33	10.43	2.8	0.93	0.65	29.10	10.36	2.5	0.94
Comprehension	48	0.55	26.57	11.65	2.9	0.94	0.63	30.11	11.12	2.9	0.93
Total	93	0.57	52.90	21.27	4.3	0.96	0.64	59.22	20.74	4.1	0.96
Level 4, Grade 4											
Vocabulary	45	0.58	26.03	9.76	2.8	0.92	0.63	28.19	9.88	2.8	0.92
Comprehension	48	0.60	28.85	10.70	2.8	0.93	0.65	31.37	10.57	2.8	0.93
Total	93	0.59	54.87	19.62	3.9	0.96	0.64	59.56	19.73	3.9	0.96
Level 5, Grade 5											
Vocabulary	45	0.54	24.28	9.38	2.8	0.91	0.58	26.00	9.42	2.8	0.91
Comprehension	48	0.57	27.34	10.69	2.8	0.93	0.61	29.34	10.48	2.8	0.93
Total	93	0.56	51.62	18.90	4.2	0.95	0.60	55.34	18.91	4.2	0.95
Level 6, Grade 6											
Vocabulary	45	0.55	24.88	9.20	2.9	0.90	0.58	26.19	9.33	2.8	0.91
Comprehension	48	0.57	27.49	10.15	3.0	0.91	0.60	28.98	10.42	2.9	0.92
Total	93	0.56	52.37	18.24	4.1	0.95	0.59	55.17	18.70	4.2	0.95
Level 7/9, Grade 7											
Vocabulary	45	0.52	23.30	9.29	2.8	0.91	0.54	24.47	9.25	2.8	0.91
Comprehension	48	0.57	27.19	10.37	2.9	0.92	0.59	28.25	10.36	2.9	0.92
Total	93	0.54	50.49	18.38	4.1	0.95	0.57	52.72	18.35	4.1	0.95
Level 7/9, Grade 8											
Vocabulary	45	0.57	25.81	9.45	2.8	0.91	0.59	26.61	9.51	2.9	0.91
Comprehension	48	0.62	29.77	10.11	2.9	0.92	0.64	30.85	10.17	2.9	0.92
Total	93	0.60	55.58	18.29	4.1	0.95	0.62	57.46	18.56	4.1	0.95
Level 7/9, Grade 9											
Vocabulary	45	0.62	27.82	9.35	2.8	0.91	0.63	28.48	9.08	2.7	0.91
Comprehension	48	0.67	31.96	10.41	2.8	0.93	0.69	33.07	9.95	2.8	0.92
Total	93	0.64	59.78	18.16	4.1	0.95	0.66	61.55	18.03	4.0	0.95
Level 10/12, Grade 10											
Vocabulary	45	0.46	20.92	9.74	2.9	0.91	0.47	21.26	9.81	2.8	0.92
Comprehension	48	0.50	24.15	10.48	3.0	0.92	0.52	25.04	10.74	3.0	0.92
Total	93	0.48	45.07	19.57	4.4	0.95	0.50	46.30	19.72	3.9	0.96
Level 10/12, Grade 11											
Vocabulary	45	0.49	21.98	9.78	2.9	0.91	0.50	22.45	10.02	2.8	0.92
Comprehension	48	0.55	26.27	10.60	3.0	0.92	0.56	26.77	10.69	3.0	0.92
Total	93	0.52	48.25	19.62	4.4	0.95	0.53	49.22	19.58	3.9	0.96
Level 10/12, Grade 12											
Vocabulary	45	0.52	23.47	9.94	2.8	0.92	0.54	24.14	10.26	2.7	0.93
Comprehension	48	0.58	28.06	10.49	3.0	0.92	0.61	29.18	10.34	2.9	0.92
Total	93	0.55	51.53	18.92	4.2	0.95	0.57	53.32	19.49	3.9	0.96
Level AR, CC[a]											
Vocabulary	45	0.66	29.63	8.21	2.7	0.89					
Comprehension	48	0.69	33.07	8.52	2.8	0.89					
Total	93	0.67	62.70	15.45	4.1	0.93					

[a] Level AR is designed as a test for students entering community college or a training program at a similar level. Norms for Level AR were therefore obtained for first year students at the beginning (in the fall) of a community college program. The norms should apply also to students entering community college at other times of the year.

Table 17. Form S ESS Summary Statistics and Standard Errors of Measurement

	Fall					Spring				
	Mean	*SD*	Median	*SEM*	K-R 20	Mean	*SD*	Median	*SEM*	K-R 20
Level PR, Grade K										
Total						326	38.2	324	10.1	0.93
Level PR, Grade 1										
Total	345	42.3	342	11.2	0.93					
Level BR, Grade 1										
Total	345	42.3	342	9.5	0.95	394	47.4	391	10.6	0.95
Level 1, Grade 1										
Word Decoding						396	51.9	392	12.7	0.94
Comprehension						394	46.0	391	12.2	0.93
Total						394	47.4	391	9.5	0.96
Level 2, Grade 2										
Word Decoding	427	50.0	423	12.2	0.94	452	52.1	451	12.8	0.94
Word Knowledge	424	43.7	421	12.4	0.92	446	47.5	445	12.6	0.93
Comprehension	423	45.1	423	12.8	0.92	445	44.5	445	12.6	0.92
Total	424	42.1	422	7.3	0.97	447	43.2	446	7.5	0.97
Level 3, Grade 3										
Vocabulary	461	42.4	459	11.2	0.93	472	43.5	471	11.5	0.93
Comprehension	461	40.5	460	10.7	0.93	475	40.5	474	10.7	0.93
Total	460	38.5	459	7.7	0.96	472	39.3	472	7.9	0.96
Level 4, Grade 4										
Vocabulary	482	39.0	481	11.0	0.92	491	41.1	491	11.6	0.92
Comprehension	486	38.7	484	10.2	0.93	496	40.8	494	10.8	0.93
Total	483	35.7	482	7.1	0.96	493	38.3	492	7.7	0.96
Level 5, Grade 5										
Vocabulary	501	36.9	500	11.1	0.91	508	38.3	506	11.5	0.91
Comprehension	502	38.2	500	10.8	0.92	509	37.7	508	10.6	0.92
Total	501	34.2	499	7.6	0.95	508	34.8	507	7.8	0.95
Level 6, Grade 6										
Vocabulary	515	37.4	513	11.2	0.91	520	38.7	519	10.9	0.92
Comprehension	516	37.1	515	11.1	0.91	522	38.9	521	11.0	0.92
Total	515	33.7	514	7.5	0.95	520	35.4	520	7.9	0.95
Level 7/9, Grade 7										
Vocabulary	526	36.2	525	11.4	0.90	531	36.4	530	11.5	0.90
Comprehension	528	35.6	526	10.7	0.91	531	35.7	531	10.7	0.91
Total	528	32.5	526	8.0	0.94	531	32.6	531	8.0	0.94
Level 7/9, Grade 8										
Vocabulary	536	38.1	535	11.4	0.91	540	38.7	539	11.6	0.91
Comprehension	536	35.1	536	10.5	0.91	540	35.8	540	10.8	0.91
Total	536	33.0	535	7.4	0.95	540	34.3	540	7.7	0.95
Level 7/9, Grade 9										
Vocabulary	544	38.5	543	10.9	0.92	547	37.6	546	11.3	0.91
Comprehension	546	38.9	544	10.3	0.93	550	38.0	548	10.7	0.92
Total	546	35.1	544	7.8	0.95	549	35.8	547	8.0	0.95
Level 10/12, Grade 10										
Vocabulary	550	36.2	549	11.4	0.90	552	36.4	551	10.9	0.91
Comprehension	555	36.4	552	10.9	0.91	559	38.2	555	10.1	0.93
Total	553	34.1	550	7.6	0.95	556	34.3	554	7.7	0.95
Level 10/12, Grade 11										
Vocabulary	554	36.5	553	11.5	0.90	556	37.4	555	11.2	0.91
Comprehension	561	37.9	558	10.7	0.92	563	38.6	561	10.9	0.92
Total	557	34.2	556	7.6	0.95	559	34.3	558	7.7	0.95
Level 10/12, Grade 12										
Vocabulary	558	37.3	557	11.2	0.91	561	39.4	558	11.1	0.92
Comprehension	566	38.6	564	10.2	0.93	572	40.4	567	10.7	0.93
Total	563	33.7	560	7.5	0.95	566	36.2	562	8.1	0.95
Level AR, CC[a]										
Vocabulary	560	36.5	557	12.6	0.88					
Comprehension	580	32.5	579	10.8	0.89					
Total	569	30.9	567	8.2	0.93					

[a] Level AR is designed as a test for students entering community college or a training program at a similar level. Norms for Level AR were therefore obtained for first year students at the beginning (in the fall) of a community college program. The norms should apply also to students entering community college at other times of the year.

Table 18. Form T ESS Summary Statistics and Standard Errors of Measurement

	Fall					Spring				
	Mean	*SD*	Median	*SEM*	K-R 20	Mean	*SD*	Median	*SEM*	K-R 20
Level 2, Grade 2										
Word Decoding	427	50.0	423	12.2	0.94	452	52.1	451	12.8	0.94
Word Knowledge	424	43.7	421	11.6	0.93	446	47.5	445	12.6	0.93
Comprehension	423	45.1	423	12.8	0.92	445	44.5	445	13.3	0.91
Total	424	42.1	422	7.3	0.97	447	43.2	446	7.5	0.97
Level 3, Grade 3										
Vocabulary	461	42.4	459	11.2	0.93	472	43.5	471	10.7	0.94
Comprehension	461	40.5	460	9.9	0.94	475	40.5	474	10.7	0.93
Total	460	38.5	459	7.7	0.96	472	39.3	472	7.9	0.96
Level 4, Grade 4										
Vocabulary	482	39.0	481	11.0	0.92	491	41.1	491	11.6	0.92
Comprehension	486	38.7	484	10.2	0.93	496	40.8	494	10.8	0.93
Total	483	35.7	482	7.1	0.96	493	38.3	492	7.7	0.96
Level 5, Grade 5										
Vocabulary	501	36.9	500	11.1	0.91	508	38.3	506	11.5	0.91
Comprehension	502	38.2	500	10.1	0.93	509	37.7	508	10.0	0.93
Total	501	34.2	499	7.6	0.95	508	34.8	507	7.8	0.95
Level 6, Grade 6										
Vocabulary	515	37.4	513	11.8	0.90	520	38.7	519	11.6	0.91
Comprehension	516	37.1	515	11.1	0.91	522	38.9	521	11.0	0.92
Total	515	33.7	514	7.5	0.95	520	35.4	520	7.9	0.95
Level 7/9, Grade 7										
Vocabulary	526	36.2	525	10.9	0.91	531	36.4	530	10.9	0.91
Comprehension	528	35.6	526	10.1	0.92	531	35.7	531	10.1	0.92
Total	528	32.5	526	7.3	0.95	531	32.6	531	7.3	0.95
Level 7/9, Grade 8										
Vocabulary	536	38.1	535	11.4	0.91	540	38.7	539	11.6	0.91
Comprehension	536	35.1	536	9.9	0.92	540	35.8	540	10.1	0.92
Total	536	33.0	535	7.4	0.95	540	34.3	540	7.7	0.95
Level 7/9, Grade 9										
Vocabulary	544	38.5	543	11.6	0.91	547	37.6	546	11.3	0.91
Comprehension	546	38.9	544	10.3	0.93	550	38.0	548	10.7	0.92
Total	546	35.1	544	7.8	0.95	549	35.8	547	8.0	0.95
Level 10/12, Grade 10										
Vocabulary	550	36.2	549	10.9	0.91	552	36.4	551	10.3	0.92
Comprehension	555	36.4	552	10.3	0.92	559	38.2	555	10.8	0.92
Total	553	34.1	550	7.6	0.95	556	34.3	554	6.9	0.96
Level 10/12, Grade 11										
Vocabulary	554	36.5	553	11.0	0.91	556	37.4	555	10.6	0.92
Comprehension	561	37.9	558	10.7	0.92	563	38.6	561	10.9	0.92
Total	557	34.2	556	7.6	0.95	559	34.3	558	6.9	0.96
Level 10/12, Grade 12										
Vocabulary	558	37.3	557	10.6	0.92	561	39.4	558	10.4	0.93
Comprehension	566	38.6	564	10.9	0.92	572	40.4	567	11.4	0.92
Total	563	33.7	560	7.5	0.95	566	36.2	562	7.2	0.96
Level AR, CC[a]										
Vocabulary	560	36.5	557	12.1	0.89					
Comprehension	580	32.5	579	10.8	0.89					
Total	569	30.9	567	8.2	0.93					

[a] Level AR is designed as a test for students entering community college or a training program at a similar level. Norms for Level AR were therefore obtained for first year students at the beginning (in the fall) of a community college program. The norms should apply also to students entering community college at other times of the year.

Table 19. Standard Errors of Measurement in NCEs

Test Level, Grade	Fall Form S	Fall Form T	Spring Form S	Spring Form T		Fall Form S	Fall Form T	Spring Form S	Spring Form T
Level PR, Grade K					**Level 7/9, Grade 7**				
Total			5.6		Vocabulary	6.7	6.3	6.7	6.3
					Comprehension	6.3	6.0	6.3	6.0
Level PR, Grade 1					Total	5.2	4.7	5.2	4.7
Total	5.6								
					Level 7/9, Grade 8				
Level BR, Grade 1					Vocabulary	6.3	6.3	6.3	6.3
Total	4.7		4.7		Comprehension	6.3	6.0	6.3	6.0
					Total	4.7	4.7	4.7	4.7
Level 1, Grade 1									
Word Decoding			5.2		**Level 7/9, Grade 9**				
Comprehension			5.6		Vocabulary	6.0	6.3	6.3	6.3
Total			4.2		Comprehension	5.6	5.6	6.0	6.0
					Total	4.7	4.7	4.7	4.7
Level 2, Grade 2									
Word Decoding	5.2	5.2	5.2	5.2	**Level 10/12, Grade 10**				
Word Knowledge	6.0	5.6	5.6	5.6	Vocabulary	6.7	6.3	6.3	6.0
Comprehension	6.0	6.0	6.0	6.3	Comprehension	6.3	6.0	5.6	6.0
Total	3.6	3.6	3.6	3.6	Total	4.7	4.7	4.7	4.2
Level 3, Grade 3					**Level 10/12, Grade 11**				
Vocabulary	5.6	5.6	5.6	5.2	Vocabulary	6.7	6.3	6.3	6.0
Comprehension	5.6	5.2	5.6	5.6	Comprehension	6.0	6.0	6.0	6.0
Total	4.2	4.2	4.2	4.2	Total	4.7	4.7	4.7	4.2
Level 4, Grade 4					**Level 10/12, Grade 12**				
Vocabulary	6.0	6.0	6.0	6.0	Vocabulary	6.3	6.0	6.0	5.6
Comprehension	5.6	5.6	5.6	5.6	Comprehension	5.6	6.0	5.6	6.0
Total	4.2	4.2	4.2	4.2	Total	4.7	4.7	4.7	4.2
Level 5, Grade 5					**Level AR, CC**[a]				
Vocabulary	6.3	6.3	6.3	6.3	Vocabulary	7.3	7.0		
Comprehension	6.0	5.6	6.0	5.6	Comprehension	7.0	7.0		
Total	4.7	4.7	4.7	4.7	Total	5.6	5.6		
Level 6, Grade 6									
Vocabulary	6.3	6.7	6.0	6.3					
Comprehension	6.3	6.3	6.0	6.0					
Total	4.7	4.7	4.7	4.7					

[a] Community college

Correlations among Tests

Correlations among the component tests or subtests and Total were computed within level for Form S from the raw scores of the students in the standardization sample who took all the tests or subtests in the form. These correlations are shown in Tables 20–22. These tables also show the reliability of the difference scores (differences between pairs of scores for the various tests or subtests). The correlations in Tables 20–22 and the K-R 20 reliabilities in Tables 14 and 15 were used to compute the reliabilities of differences.

Table 20. Raw Score Subtest Intercorrelations (above Diagonal) and Reliabilities of Differences (below Diagonal) Levels PR and BR

	Level PR, Spring of Kindergarten *N* = 2039				
Subtest	**LC**	**PA**	**LS**	**SC**	**Tot**
Subtest 1, Literacy Concepts (**LC**)		0.59	0.58	0.53	0.80
Subtest 2, Oral Language/Phonological Awareness (**PA**)	0.55		0.62	0.50	0.84
Subtest 3, Letters and Letter-Sound Correspondences (**LS**)	0.63	0.63		0.45	0.84
Subtest 4, Listening (Story) Comprehension (**SC**)	0.60	0.65	0.73		0.76
	Level PR, Fall of Grade 1 *N* = 1884				
	LC	**PA**	**LS**	**SC**	**Tot**
Subtest 1, Literacy Concepts (**LC**)		0.57	0.56	0.57	0.79
Subtest 2, Oral Language/Phonological Awareness (**PA**)	0.56		0.60	0.57	0.85
Subtest 3, Letters and Letter-Sound Correspondences (**LS**)	0.61	0.63		0.50	0.82
Subtest 4, Listening (Story) Comprehension (**SC**)	0.55	0.59	0.69		0.81
	Level BR, Fall of Grade 1 *N* = 2645				
	IC	**FC**	**V**	**SW**	**Tot**
Subtest 1, Letter-Sound Corresp: Initial Consonants (**IC**)		0.76	0.68	0.76	0.89
Subtest 2, Letter-Sound Corresp: Final Consonants (**FC**)	0.33		0.70	0.76	0.89
Subtest 3, Letter-Sound Corresp: Vowels (**V**)	0.47	0.41		0.71	0.86
Subtest 4, Basic Story Words (**SW**)	0.43	0.42	0.48		0.93
	Level BR, Spring of Grade 1 *N* = 2507				
	IC	**FC**	**V**	**SW**	**Tot**
Subtest 1, Letter-Sound Corresp: Initial Consonants (**IC**)		0.74	0.71	0.76	0.88
Subtest 2, Letter-Sound Corresp: Final Consonants (**FC**)	0.39		0.72	0.70	0.88
Subtest 3, Letter-Sound Corresp: Vowels (**V**)	0.44	0.41		0.73	0.89
Subtest 4, Basic Story Words (**SW**)	0.45	0.54	0.48		0.91

Table 21. Raw Score Test Intercorrelations (above Diagonal) and Reliabilities of Differences (below Diagonal) Levels 1 and 2, Form S

					Spring (*N* = 2203)			
Level 1, Grade 1					**WD**	**C**	**Total**	
Word Decoding (**WD**)						0.83	0.96	
Comprehension (**C**)					0.59		0.95	
	Fall (*N* = 3680)				Spring (*N* = 3970)			
Level 2, Grade 2	**WD**	**WK**	**C**	**Total**	**WD**	**WK**	**C**	**Total**
Word Decoding (**WD**)		0.86	0.82	0.95		0.86	0.80	0.95
Word Knowledge (**WK**)	0.50		0.81	0.95	0.57		0.80	0.95
Comprehension (**C**)	0.62	0.60		0.92	0.66	0.63		0.91

Table 22. Raw Score Test Intercorrelations (above Diagonal) and Reliabilities of Differences (below Diagonal) Levels 3–AR, Form S

	Fall			Spring		
	V	**C**	**Total**	**V**	**C**	**Total**
Level 3, Grade 3		(*N* = 3584)			(*N* = 3419)	
Vocabulary (**V**)		0.83	0.95		0.81	0.95
Comprehension (**C**)	0.58		0.96	0.62		0.96
Level 4, Grade 4		(*N* = 3126)			(*N* = 3294)	
Vocabulary (**V**)		0.80	0.95		0.81	0.95
Comprehension (**C**)	0.60		0.96	0.61		0.96
Level 5, Grade 5		(*N* = 2770)			(*N* = 3126)	
Vocabulary (**V**)		0.80	0.94		0.78	0.94
Comprehension (**C**)	0.58		0.95	0.62		0.95
Level 6, Grade 6		(*N* = 2538)			(*N* = 2440)	
Vocabulary (**V**)		0.76	0.94		0.77	0.94
Comprehension (**C**)	0.63		0.94	0.65		0.94
Level 7/9, Grade 7		(*N* = 2445)			(*N* = 1873)	
Vocabulary (**V**)		0.76	0.93		0.75	0.93
Comprehension (**C**)	0.60		0.94	0.62		0.94
Level 7/9, Grade 8		(*N* = 2228)			(*N* = 1964)	
Vocabulary (**V**)		0.77	0.94		0.76	0.94
Comprehension (**C**)	0.61		0.94	0.62		0.94
Level 7/9, Grade 9		(*N* = 1450)			(*N* = 2193)	
Vocabulary (**V**)		0.77	0.94		0.74	0.93
Comprehension (**C**)	0.65		0.95	0.68		0.94
Level 10/12, Grade 10		(*N* = 1327)			(*N* = 1634)	
Vocabulary (**V**)		0.79	0.94		0.79	0.94
Comprehension (**C**)	0.55		0.95	0.60		0.96
Level 10/12, Grade 11		(*N* = 986)			(*N* = 872)	
Vocabulary (**V**)		0.79	0.94		0.72	0.92
Comprehension (**C**)	0.58		0.95	0.70		0.94
Level 10/12, Grade 12		(*N* = 862)			(*N* = 941)	
Vocabulary (**V**)		0.77	0.93		0.72	0.92
Comprehension (**C**)	0.64		0.95	0.72		0.94
Level AR, CC[a]		(*N* = 2824)				
Vocabulary (**V**)		0.68	0.91			
Comprehension (**C**)	0.65		0.92			

[a] Community college

As reliabilities of differences in test scores go, those shown in Tables 20–22 are relatively high. They justify using score differences, if large, to locate students who may need special help. For example, the *GMRT* publications *Linking Testing to Teaching* provide guidance for using differences between Vocabulary and Comprehension scores to locate students who might profit from special help with vocabulary building or with comprehension. Of particular interest are the good reliabilities of differences between the Level 2 Word Decoding and Word Knowledge scores as shown in Table 21. Although the format of these two tests is the same, and students must identify words in both tests, there are important differences between the tests in the nature of the test words and of the wrong answers. The good reliabilities of the difference scores provide a basis for identifying imbalances between students' ability to use decoding skills and their knowledge of word meanings.

Stability of Scores: Fall-Spring Correlations

Students in several of the schools in the standardization sample took Form S of the level for their grade in both fall and spring. All schools that participated in the fall standardization sample were invited to participate again in the spring. Although many schools declined, a large fraction of the students who participated in the fall standardization did participate again in the spring. The spring testing of these students permitted the computation of fall-spring correlations.

To locate the same students' scores in the two standardizations, a scoring program matched the student information on the scoring documents for the two testings. Matching was based on student name, date of birth, gender, and grade. Additional matching was performed by hand. The number of students at each grade whose scoring documents could be matched and the correlations between their fall and spring scores are shown in Tables 23–25. The correlations between the fall and spring scores indicate the extent of stability or change in the students' relative standing.

Table 23. Correlations between Fall and Spring Total Test Raw Scores at Grade 1

Test Level	*N*	Fall or Spring	*r*	Mean	*SD*
PR	280	Fall	0.74	69.46	12.19
BR		Spring		54.47	12.05
PR	160	Fall	0.66	73.70	10.79
1		Spring		50.63	17.59
BR	610	Fall	0.77	38.90	14.38
BR		Spring		57.07	12.71
BR	78	Fall	0.90	37.27	15.03
1		Spring		49.12	18.17

Table 24. Correlations between Fall and Spring Raw Scores at Grade 2 for Level 2, Form S (*N* = 906)

	Word Decoding			Word Knowledge			Comprehension			Total		
	r	Mean	*SD*	*r*	Mean	*SD*	*r*	Mean	*SD*	*r*	Mean	*SD*
Fall	0.86	25.44	10.81	0.86	22.48	10.25	0.82	24.86	9.23	0.90	72.77	28.65
Spring		31.39	10.13		27.80	10.42		29.64	8.27		88.83	27.32

Table 25. Correlations between Fall and Spring Raw Scores at Grades 3–12, Form S

Test Level	Grade	*N*	Fall or Spring	Vocabulary			Comprehension			Total		
				r	Mean	*SD*	*r*	Mean	*SD*	*r*	Mean	*SD*
3	3	601	Fall	0.88	22.49	10.43	0.84	25.92	11.08	0.90	48.41	20.60
			Spring		27.68	10.66		30.24	10.93		58.00	20.52
4	4	476	Fall	0.90	25.26	9.50	0.85	27.42	10.61	0.92	52.68	19.07
			Spring		29.64	9.28		31.68	10.66		61.32	19.11
5	5	463	Fall	0.85	23.21	9.28	0.83	26.87	10.24	0.89	50.09	18.54
			Spring		27.03	9.68		30.51	10.44		57.54	18.96
6	6	464	Fall	0.91	22.06	9.13	0.83	29.30	9.66	0.91	51.37	17.48
			Spring		25.04	9.55		32.58	9.20		57.62	17.45
7/9	7	285	Fall	0.88	23.59	9.06	0.83	30.72	9.19	0.91	54.31	17.23
			Spring		26.39	9.38		32.62	9.41		59.01	17.77
7/9	8	257	Fall	0.90	26.28	9.21	0.86	31.39	10.02	0.93	57.66	18.07
			Spring		29.13	9.56		33.57	9.80		62.70	18.29
7/9	9	237	Fall	0.90	24.62	9.99	0.74	28.36	10.43	0.88	52.98	19.03
			Spring		27.41	8.91		31.21	10.03		58.62	17.78
10/12	10	112	Fall	0.89	23.24	7.95	0.77	26.64	9.12	0.88	49.88	15.73
			Spring		25.55	8.19		28.75	9.28		54.30	16.62
10/12	11	87	Fall	0.89	25.75	8.47	0.78	29.54	8.86	0.90	55.29	16.26
			Spring		27.51	8.27		30.09	10.70		57.60	17.96
10/12	12	100	Fall	0.75	24.25	9.82	0.58	26.37	11.23	0.71	50.62	19.63
			Spring		26.03	10.33		27.13	12.11		53.16	21.18

The correlations between fall and spring scores for Grade 1 are generally smaller than at the other grades. This difference is to be expected, since many of the students had received little or no reading instruction when first tested at the beginning of Grade 1, while other students had received considerable reading instruction. Thus, the students' fall scores were partly a reflection of the amount of reading instruction they had received prior to fall of first grade. By spring, nearly all students had received considerable reading instruction, and the amount of instruction they had received prior to fall of Grade 1 was a much less important influence on their scores.

The correlations between fall and spring scores can be compared with the alternate-form reliability coefficients shown in Table 13. In most cases, the alternate-form reliabilities, based on tests taken within three weeks of each other, are somewhat higher than the fall-spring correlations.

Validity

Completion Rates

All the *GMRT* tests are designed as power tests;[84] they measure the students' knowledge of concepts related to reading, their knowledge of decoding skills and word meanings, and their understanding of what they read. The tests are not intended to measure how quickly the students can respond to the test questions or how rapidly they can skim through printed material. A test of reading speed can sometimes provide very useful information, but a *comprehension* test that is too speeded gives ambiguous scores. For example, a low score on a speeded test may be the result of a student's gaining

- Only a superficial understanding of many passages and answering many questions with poor accuracy, or
- A thorough understanding of fewer passages and answering relatively few questions with good accuracy.

The *Gates-MacGinitie Reading Tests* give the great majority of students time to apply fully their powers of word decoding, word knowledge, and reading comprehension. Table 26 shows this characteristic of the tests.[85] The completion rates shown in Table 26 were computed from the numbers of questions answered by students in the standardization sample and are therefore shown only for Form S, which was the standardization form. Given that the tests in Form T are the same length as those in Form S, follow the same content blueprint, have identical formats, and are closely matched in difficulty, completion rates for Form T should be very similar to those for Form S.

In the fall, more than 80% of the students at all grade levels completed each of the tests they took, except for the Comprehension test at Grade 4 and Grade 5, which 79% of the students completed. In the spring, more than 85% of the students at all grade levels completed each of the tests they took. Three-quarters of the questions on each test were answered by at least 90% of the students in the fall and by at least 95% of the students in the spring. Data on completion rates are not given for Level PR and Level BR because all the students taking those levels should be able to try all the questions. For those levels, the teacher is directed to "give [the students] enough time to do the best they can with each question, without pausing so long that they become inattentive."

Table 26. Form S Completion Rates

	Grade												
	1	2	3	4	5	6	7	8	9	10	11	12	CC
	Test Level												
Test	1	2	3	4	5	6	7/9	7/9	7/9	10/12	10/12	10/12	AR
Fall													
Percentage of Students Completing the Test													
Word Decoding		91											
Word Knowledge/Vocabulary		93	91	85	87	90	92	96	96	97	96	95	97
Comprehension		93	81	79	79	84	85	89	90	91	90	91	86
Percentage of Students Completing 75% of the Questions													
Word Decoding		97											
Word Knowledge/Vocabulary		97	96	94	95	96	97	98	98	98	97	97	98
Comprehension		97	92	92	91	95	95	96	94	95	94	95	94
Spring													
Percentage of Students Completing the Test													
Word Decoding	93	97											
Word Knowledge/Vocabulary		98	95	92	94	97	97	99	99	99	99	99	
Comprehension	90	97	88	88	87	94	94	96	93	91	93	93	
Percentage of Students Completing 75% of the Questions													
Word Decoding	97	99											
Word Knowledge/Vocabulary		99	98	97	98	99	99	100	99	99	100	100	
Comprehension	96	99	95	95	95	98	99	99	98	95	97	97	

Ceiling and Floor Data

In most schools, the range of achievement at each grade level is very great. A test that shows the level that the very best student can reach will usually be so difficult that the poorest students can only guess at many of the answers; a test that provides meaningful scores for the lowest achieving students will usually give many of the better students perfect scores, so the full range of their achievement is not measured. A general achievement test should provide meaningful scores for the great majority of students, so that only the students with extremely high or extremely low achievement cannot obtain scores fully indicative of their achievement level.

As described in the section "Question Difficulty," the difficulties of the questions in the *GMRT* were carefully estimated prior to the field test so that the distributions of question difficulties obtained in the field test would enable the authors to construct final test forms with the desired difficulty characteristics. As a result, all *GMRT* test forms are appropriate for the range of reading achievement found in most classes across the country. In those cases in which a test of different difficulty is needed, the readily-available out-of-level norms encourage the use of an easier or more difficult test level.

The difficulty distributions of the Fourth Edition are generally excellent. This characteristic is shown in Tables 27–29. The ceiling and floor data for Levels 1 through 10/12 and AR are presented in Tables 27 and 28 and for Levels PR and BR in Table 29.

Levels 1 through 10/12 and AR

For Levels 1 through 10/12 and AR, the PR of a perfect score is 99 for all tests in all levels at the grade levels for which the tests are designed, except for the Level 2 Word Decoding test, for which this PR was 98 (Form S) and 97 (Form T) in the fall and 94 (both forms) in the spring. By the spring of second grade most students have a very good command of basic decoding skills, and the Level 2 Word Decoding test is not very difficult. The Level 2 Word Decoding test could have been made more difficult by having it test students' knowledge of relatively esoteric decoding rules. Since the test was designed to locate those students who need continued help with the most productive decoding rules, however, only decoding rules that were deemed to be generally useful were included.

At the other end of the distributions, the PR of a chance level score in the spring is 10 or lower for nearly all tests at the grades for which the test was designed. In the fall, the Comprehension test of both forms of Levels 2 and 3 was more difficult, with PRs of a chance level score being somewhat higher than 10. By spring, however, the PR of a chance level score for both levels had dropped to 7 or less. The Comprehension test in Level 1 and both tests in Level 10/12, particularly Comprehension, were relatively difficult—PRs of a chance level score are relatively high—though not excessively so.

Table 27. Form S Ceiling and Floor Data

	Grade												
	1	2	3	4	5	6	7	8	9	10	11	12	CC
	Test Level												
Test	1	2	3	4	5	6	7/9	7/9	7/9	10/12	10/12	10/12	AR
Fall													
Ceiling—PR of a Perfect Score													
Word Decoding		98											
Word Knowledge/Vocabulary		99	99	99	99	99	99	99	99	99	99	99	99
Comprehension		99	99	99	99	99	99	99	99	99	99	99	99
Ceiling—PR of One Wrong													
Word Decoding		94											
Word Knowledge/Vocabulary		99	99	99	99	99	99	99	99	99	99	99	99
Comprehension		98	99	99	99	99	99	99	99	99	99	99	99
Floor—PR of a Chance Score													
Word Decoding		6											
Word Knowledge/Vocabulary		10	7	5	6	5	6	5	3	7	6	5	2
Comprehension		12	13	8	8	6	7	4	3	10	8	7	1
Spring													
Ceiling—PR of a Perfect Score													
Word Decoding	99	94											
Word Knowledge/Vocabulary		99	99	99	99	99	99	99	99	99	99	99	
Comprehension	99	98	99	99	99	99	99	99	99	99	99	99	
Ceiling—PR of One Wrong													
Word Decoding	96	86											
Word Knowledge/Vocabulary		96	99	99	99	99	99	99	99	99	99	99	
Comprehension	97	95	99	99	99	99	99	99	99	99	99	99	
Floor—PR of a Chance Score													
Word Decoding	6	2											
Word Knowledge/Vocabulary		5	5	3	4	5	5	4	2	6	5	4	
Comprehension	15	4	5	5	6	6	7	4	2	9	8	6	

Table 28. Form T Ceiling and Floor Data

	Grade											
	2	3	4	5	6	7	8	9	10	11	12	CC
	Test Level											
Test	2	3	4	5	6	7/9	7/9	7/9	10/12	10/12	10/12	AR
Fall												
Ceiling—PR of a Perfect Score												
Word Decoding	97											
Word Knowledge/Vocabulary	99	99	99	99	99	99	99	99	99	99	99	99
Comprehension	99	99	99	99	99	99	99	99	99	99	99	99
Ceiling—PR of One Wrong												
Word Decoding	94											
Word Knowledge/Vocabulary	99	99	99	99	99	99	99	99	99	99	99	99
Comprehension	98	99	99	99	99	99	99	99	99	99	99	99
Floor—PR of a Chance Score												
Word Decoding	7											
Word Knowledge/Vocabulary	10	9	5	6	4	5	4	3	12	10	9	1
Comprehension	14	15	8	9	7	7	4	3	14	11	10	1
Spring												
Ceiling—PR of a Perfect Score												
Word Decoding	94											
Word Knowledge/Vocabulary	99	99	99	99	99	99	99	99	99	99	99	
Comprehension	98	99	99	99	99	99	99	99	99	99	99	
Ceiling—PR of One Wrong												
Word Decoding	85											
Word Knowledge/Vocabulary	96	99	99	99	99	99	99	99	99	99	99	
Comprehension	96	99	99	99	99	99	99	99	99	99	99	
Floor—PR of a Chance Score												
Word Decoding	3											
Word Knowledge/Vocabulary	5	6	4	4	4	5	3	2	11	10	8	
Comprehension	5	7	5	6	7	7	4	2	13	11	7	

Levels PR and BR

The ceiling and floor data for Levels PR and BR are presented and discussed separately because, for these two test levels, subtest scores are tabled as stanines rather than as PRs.[86] Also, Levels PR and BR are functionally somewhat different from the higher levels. The subtests of Level PR are designed to be relatively easy, so they can most effectively help locate students whose background for learning to read is weak—students who will be likely to need special attention. Level BR is designed so that it may be given at both the beginning and end of Grade 1. It is given at the end of Grade 1 primarily to classes that have made less than average progress, and therefore it must be relatively easy at the end of Grade 1. Level BR is designed to test basic elements of reading—letter-sound correspondences and the reading in context of very common words that may not follow basic decoding rules. That these intentions were realized is evident from Table 29.

Level PR

On two of the subtests of Level PR, the stanines of a perfect score do indeed reach 9 at both the spring of Kindergarten and the fall of Grade 1. Those stanines drop to 8 when only a single question is missed, however, suggesting that the scores fall rapidly, as indeed they do, when only a few questions are missed. Since the stanine of a chance score is either 1 or 2 at both testing times, most raw scores are assigned to students with average stanines (5) or less. The Oral Language Concepts and Listening (Story) Comprehension subtests were the most difficult. Even those subtests, however, assign 15 of the 20 raw score points to stanines of 5 or less at both the spring of Kindergarten and the fall of Grade 1.

The PRs for Total on Level PR confirm this analysis of subtest scores. At the spring of Kindergarten, 78 of the 90 Total raw score points correspond to PRs below the upper quartile. At the fall of Grade 1, 83 of the 90 Total raw score points correspond to PRs below the upper quartile.

Level BR

In the fall of Grade 1, the stanines of a perfect score on three of the Level BR subtests reach 9 and one reaches 8. Although these stanines drop one point when only a single question is missed, the floors of the subtests—the stanines of the chance scores—are relatively high (2, 3, 4, and 2). Thus, although some students in the fall of Grade 1 get all or nearly all of the questions correct on each of the subtests, many students find Level BR rather difficult at the beginning of the year. That is to be expected, since, in many schools, a concerted effort to teach students to read does not begin before the fall of Grade 1.

In the spring of Grade 1, the stanines of a perfect score on the four Level BR subtests are only 7, 7, 8, and 7. Most of the score range above a chance score gives stanines between 1 and 5. Level BR discriminates well, therefore, among students in the lower half of the achievement distribution, fulfilling its purpose of providing a good measure of achievement of basic reading skills for those classes that have made less than average progress. Thus, the difficulty of Level BR is close to ideal. It could not have been made easier in the fall of Grade 1 without making it so easy in the spring that it would not provide sound assessment in below-average classes.

The PRs for Total on Level BR confirm this analysis of subtest scores. In the fall of Grade 1, the two highest possible scores (a perfect score and only one wrong) correspond to a PR of 99. A student who misses 16 of the 70 questions is still at the upper quartile (PR of 75). There is room to show a lot of growth in basic reading skills. In the spring of Grade 1, even a perfect score corresponds to a PR of only 98, and a student who misses only four questions is at the upper quartile. The preponderance of the Total score range is available for placing the achievement of students below the upper quartile.

Table 29. Ceiling and Floor Data for Levels PR and BR

Level PR			
Spring of Kindergarten		**Fall of Grade 1**	
Ceiling—Stanine of a Perfect Score			
Literacy Concepts	8	Literacy Concepts	7
Oral Language Concepts (Phonological Awareness)	9	Oral Language Concepts (Phonological Awareness)	9
Letters & Letter-Sound Correspondences	8	Letters & Letter-Sound Correspondences	8
Listening (Story) Comprehension	9	Listening (Story) Comprehension	9
Ceiling—Stanine of One Wrong			
Literacy Concepts	6	Literacy Concepts	6
Oral Language Concepts (Phonological Awareness)	8	Oral Language Concepts (Phonological Awareness)	8
Letters & Letter-Sound Correspondences	7	Letters & Letter-Sound Correspondences	6
Listening (Story) Comprehension	8	Listening (Story) Comprehension	8
Floor—Stanine of a Chance Score			
Literacy Concepts	1	Literacy Concepts	1
Oral Language Concepts (Phonological Awareness)	2	Oral Language Concepts (Phonological Awareness)	1
Letters & Letter-Sound Correspondences	1	Letters & Letter-Sound Correspondences	1
Listening (Story) Comprehension	2	Listening (Story) Comprehension	2
Level BR			
Fall of Grade 1		**Spring of Grade 1**	
Ceiling—Stanine of a Perfect Score			
Initial Consonants	8	Initial Consonants	7
Final Consonants	9	Final Consonants	7
Vowels	9	Vowels	8
Basic Story Words	9	Basic Story Words	7
Ceiling—Stanine of One Wrong			
Initial Consonants	7	Initial Consonants	5
Final Consonants	8	Final Consonants	6
Vowels	8	Vowels	7
Basic Story Words	8	Basic Story Words	6
Floor—Stanine of a Chance Score			
Initial Consonants	2	Initial Consonants	1
Final Consonants	3	Final Consonants	1
Vowels	4	Vowels	2
Basic Story Words	2	Basic Story Words	1

Contributions of Test Design and Development

Nearly every feature of the design of the *GMRT* and nearly every procedure followed in their development is an important contributor to the validity of the *GMRT* as survey measures of reading achievement. Many of these features of design and development were described in the preceding sections. Some of them are listed again here with specific reference to their contributions to validity.

Test Design

- The validity of the *GMRT* is rooted in the overall design of the series, which measures the progression of students' understandings and skills in reading from Kindergarten through high school.
  - Reading development is assessed as it progresses from background knowledge that is important for learning to read, to understanding appropriately sophisticated expository and narrative prose. Each level tests what is central to reading development at the grade(s) for which the level is intended. Thus, achievement is measured through the use of tasks and materials congruous with the students' stage of reading growth.
  - Comprehension is assessed as it progresses from understanding stories read aloud, to reading simple stories and expository text, to reading increasingly mature, age-appropriate text. The Comprehension test at each level tests understanding of the range of materials students in the associated grades are learning to read. The concepts and inferences tested and the difficulty and tone of the writing are age appropriate.
  - Reading vocabulary is assessed as it progresses from use of important letter-sound correspondences and reading frequently-used words ("sight words") in context, to knowledge of common word meanings, to progressively more developed vocabulary. Thus, vocabulary development, also, is tracked by tests that assess the kinds of learnings that are basic to the stage of the students' reading development.
- New tests and testing formats are based on research findings and on the authors' assessment of their practical usefulness. Pilot studies of these new tests and formats were conducted to ensure that the tasks and the directions for administering them would be clear and that the tasks would be appropriate in difficulty.
- Exceptional care in test construction make the measurements relatively free of unintended influences. For example, in the Basic Story Words subtest of Level BR, the words in the context sentences for these stories were selected from the same list of common words as the test words, so the context sentences would be clear in meaning. But none of the test words were used in the contexts, thus assuring that students could not learn to recognize a right answer by having it read to them in the context for one of the other questions.
- Careful analysis of the time required for answering the questions means that nearly all students have ample opportunity to show how well they can read.

Level PR and Level BR

Only questions that the field test demonstrated were effective and easily understood by the students were included in Levels PR and BR. Question types that did not meet these standards were eliminated. All questions in the first three subtests of both levels follow a simple, easy-to-understand format, and distinctive pictures and silhouettes guide the students so that they will keep working at the proper place. Thus, the students are able to do their best on each question without the distraction of any difficulty in keeping the place or of understanding a new or puzzling format.

There is one subtest in each of these levels that is new for the Fourth Edition—Listening (Story) Comprehension in Level PR and Basic Story Words in Level BR. Both of these new subtests were fully pilot tested before the field test, and the field test was used to select the stories that proved most useful in assessing the students' comprehension of stories and ability to identify frequently-used words correctly.

Word Decoding, Word Knowledge, and Vocabulary Tests

- Relatively familiar words were used as the test words for the Word Decoding tests of Levels 1 and 2. The selection was guided by various word frequency counts and by the authors' considerable experience, so that the tests are as independent of word knowledge as possible and thus more valid as measures of decoding.

- The Word Decoding tests provide some of the same kinds of information as an analysis that involves recording a student's oral reading errors and noting patterns of errors[87]—a type of analysis that is widely used by clinicians to help students who are having difficulty with decoding skills. While a Word Decoding test is not a substitute for recording oral reading errors, it has some important advantages. Although it does not involve reading words in context, it provides a more direct assessment of the student's use of specific skills uninfluenced by the cues and misleads that verbal context can inject. Also, information on the skills that may be weak is gathered for the entire class at one time, rather than for individual students, one at a time.

- Written guidelines for picture specifications and repeated editing of the pictures that depict correct answer words in the Word Decoding and Word Knowledge tests avoid ambiguity in the relation of the picture to the answer choices.

- Comprehensive rules for the use and reuse of words in the Word Knowledge and Vocabulary tests make the questions independent of each other and limit irrelevant influences on the students' responses.

- Analyses of the relative frequency of English parts of speech guided the selection of test words in the Word Knowledge and Vocabulary tests, so that the tested words are appropriately representative of grammatical uses of words.

- Answer choices for the Word Knowledge and Vocabulary tests take account of various strategies that can interfere with the accurate measurement of the students' knowledge of word meanings.
  - Length or position of answer choices is not a useful cue to right answers.
  - Inclusion of three types of wrong answers—visual similarity, miscue, and association—limits the usefulness of students' irrelevant answer strategies.
  - Wrong answer choices, other than visual similarity wrong answers, are at least as familiar as the correct answer, adding to their effectiveness as wrong answers.

Comprehension Tests at Levels 1 and 2

- The format for the Comprehension tests at Levels 1 and 2 allows the students to focus on reading the passages without the additional reading load of
  - Written questions and answer choices and
  - The complexities of trying to relate written questions to the passage and to the alternative answer choices.
- The careful preparation of the stories and informational texts by the test authors and by five children's authors provides test passages that maintain interest and that are characteristic of texts that students in Grades 1 and 2 are learning to read.
- The 10 passages in each test form permit the inclusion of a wide range of topics that will not favor students with specific backgrounds or interests. The 10 passages can also present a good balance of fiction, natural science, and social science and of narrative and expository prose.
- Comprehensive specifications and thorough editing of the three answer-choice pictures for each text segment resulted in pictures that are understandable and unambiguous. In addition, each wrong answer picture was checked to be sure that it could not reasonably be taken as an illustration of the passage, and the sequence of answer panels was examined to be sure that students could not figure out the story from the pictures alone.

Comprehension Tests at Levels 3 through 10/12 and AR

- The participation of several teachers and former teachers in the selection of test passages resulted in a selection of test passages that are varied in content and style and representative of reading that students do.
- Explicit written guidelines for the selection of passages helped ensure that the passages selected were appropriate in content and presentation.

- To provide a representative selection of reading tasks, each test includes 11 passages, so that
  - The likelihood that a student will score well or poorly because of extensive knowledge, or lack of knowledge, about a particular topic, plot, or theme is small. The score is more representative of the student's general ability to construct meaning than if the score depended on just a few passages.
  - Fiction, natural science, social science, and the humanities are all represented in the passage content in a balanced and developmental way.
  - The reading tasks include an appropriate balance of narrative, expository, and setting passages.
- A balance of literal and inferential questions is maintained at all levels in order to assess the students' abilities both to construct literal meaning and make appropriate inferences.
- Guidelines for construction of questions about the passages helped to ensure that the questions are significant and indicative of understanding.
  - Correct answers were designed not to be evident from prior knowledge, from the answer to another question, from the repetition of a word or phrase from the passage, or from a plot or concept that can be deduced from the set of questions.
  - The questions are clear and relatively simple; they assess understanding of the passages, not the ability to read the questions.
- Wrong answer choices for the Comprehension tests at Levels 3 through 10/12 and AR are designed to prevent students' getting good scores by following strategies that are irrelevant to the construction of meaning.
  - Length or position of answer choices is not a useful cue to right answers.
  - Inclusion of prior-knowledge and text-phrase wrong answers limits the usefulness of students' irrelevant answer strategies. Students who rely on prior knowledge rather than constructing meaning from the passage and students who answer questions simply by matching a phrase in an answer choice with a phrase in the passage are likely to choose wrong answers rather than correct ones.
  - All answer choices fit the questions grammatically, so wrong answers cannot be eliminated on the basis of their lack of grammatical fit.
- Qualitative analysis of age appropriateness and reading difficulty, plus the quantitative analysis of readability with three readability formulas, ensures that the passages included in the various test levels are appropriate in content, tone, and reading difficulty for the range of interests and reading ability at the grade level(s) for which each test is designed.

Question Difficulty

At each stage of test development, test forms were constructed that conformed closely to a blueprint of desired difficulties. As a result, all *GMRT* test forms are appropriate for the range of reading achievement found in most classes across the country. They are thus free of ceiling and floor limitations on their validity for most classes in most schools. Out-of-level norms encourage use of out-of-level tests when a test of different difficulty is needed.

Cultural Diversity

- Examination of the field-test forms by 15 reviewers representing various ethnic groups from various parts of the country was part of a strong effort to make sure that the test scores are not influenced by irrelevant factors, such as biased questions or content that might distract students from doing their best.
- Review by a sociolinguist of the Word Decoding questions in Levels 1 and 2 and the Letters and Letter-Sound Correspondences questions in Level BR for linguistic accuracy and for questions that might be confusing to speakers of an African American vernacular English contributes to the validity of those tests as measures of decoding skills.
- Statistical bias analysis (analysis of Differential Item Functioning) was used to check for questions that might be unfair to African American or Hispanic students. Questions with any strong suggestion of DIF were not used.
- Comprehension test passages at all levels were chosen so that females and males of various ethnic groups would be represented in test content—as characters in pictures and passages and as authors of passages. Test materials that students see as meaningful and relevant are likely to provide more valid scores.

Field Testing

Extensive field-test data collection and data analysis contributed to the validity of the tests through providing rigorous empirical data that guided the selection of questions and passages for the published tests. The teachers' cooperation in allowing time for all but the very slowest students to finish made the data for questions at the end of each test comparable to the data for the earlier questions.

- Field-test data analysis indicated that the high reliability of the tests meant that the tests could be short enough for students' motivation to be maintained throughout the testing. Motivation to do well is important to the validity of a test.

- Data on reliability permitted adjustments in the length of some tests to ensure that almost all students are able to answer all or nearly all of the questions within the time allowed. As a result, even students who work relatively slowly are able to show how well they can understand what they read.
- Comments from teachers on procedures for administering the tests helped in revising the directions for administration. Test administration that goes smoothly, that gives students a clear understanding of the task without an overload of information, and that does not sap motivation by being confusing or protracted enhances the validity of any test.

Question Selection

- Questions that had a low correlation between the right answer and the total score, or a positive correlation between a wrong answer and the total score, or that were extremely easy or extremely difficult were generally eliminated from consideration in selecting questions. Eliminating such questions improves validity as well as reliability. There is usually something unclear or otherwise faulty about a question that the best readers miss. And questions that no one misses or that no one gets right add no information to a test score.
- Questions for alternate forms of the tests were selected to make the alternate forms similar in many important respects. Equivalent test content (such characteristics as test word difficulty, parts of speech, type of passage content, literal and inferential questions), equivalent question difficulty, and equivalent question-test correlation mean that the two forms of any test measure essentially the same thing—an essential requirement for the validity of the tests.
- For Level PR, questions were selected that measure concepts that the authors considered teachable. Since the aim of Level PR is to locate students with background weaknesses that can be strengthened to improve success in initial reading, concepts that a student can learn are central to the validity of the test.
- For Level BR, questions for the three letter-sound correspondences subtests were chosen to represent those correspondences that are commonly taught in beginning reading instruction. Test words for the Basic Story Words subtest are unusually useful words that are typically learned in Grade 1.
- For the Levels 1 and 2 Word Decoding tests, the questions are based on decoding skills commonly taught in the primary grades and on test words that are generally known by the students in speech, so that missed questions are likely to represent decoding skills that the students do not yet know or do not use.

- For the Level 2 Word Knowledge and Levels 3 through 10/12 and AR Vocabulary tests, the words tested were chosen to be words that are useful to students with the typical range of reading ability at the grade level(s) for which the test level is designed. Unusual or specialized words that might function in a test of vocabulary knowledge but that would not be very useful in actual reading were not included.
- For the Levels 1 through 10/12 and AR Comprehension tests, quantitative or classifiable characteristics, such as passage and question difficulty and passage and question content, were important in building the test forms. In addition, however, the authors' judgments of the quality of the passage (see the section "Passage Characteristics in Levels 3 through 10/12 and AR" on page 17) and of the usefulness of the questions as indicators of the students' understanding of the passage were factors in the selection of passages and questions. Thus, test validity is based both on technical characteristics and on the authors' experience of what is useful in teaching and testing comprehension.

Other Evidence of Validity

Several studies of the relationships of Total scores on the Third Edition of the *Gates-MacGinitie Reading Tests* with other tests, grade point averages, and students' letter grades in reading were conducted. These studies are reported in the *Technical Report* for the Third Edition.[88] The results of these studies of the Third Edition are relevant to the validity of the Fourth Edition because

- The Total score correlations between the Third Edition and the Fourth Edition were very high (Table 12 on page 49);
- The design of the two editions was very similar;
- The procedures for developing the Fourth Edition tests were essentially the same as those for developing the Third Edition.

For these reasons, these studies of the Third Edition are summarized in this section.

Correlations with Other Reading Tests

In general, the correlations between the Third Edition and other reading tests were high, but in most cases not as high as the alternate-form reliabilities for the Third Edition. These results were to be expected. Although all the reading tests with which the Third Edition was compared had good reliability and all of them assessed reading achievement in a meaningful and objective way, the other comprehension tests were not based on the same blueprint or on selections from actual published sources. Also, the other vocabulary tests did not use the same criteria for selecting vocabulary test words.

Correlations between the Third Edition and the PSAT Verbal section, SAT Verbal section, and ACT English test were fairly high, but not as high as the correlations with the other reading tests. The PSAT and SAT Verbal sections and the ACT English test include tasks that are not strictly reading. Also, the groups that take them are relatively select, so the correlations were attenuated by a more restricted range. The correlations with the mathematics tests were considerably lower and typical of correlations between reading achievement and mathematics achievement.

Correlations with Course Grades

Studies of the relationships between scores on the Third Edition of the *Gates-MacGinitie Reading Tests* and teacher-assigned course grades generally showed that the correlations up through Grade 6 were quite high. At Grades 3–6, the correlations between reading course grades and Comprehension were somewhat higher, on the average, than those between reading course grades and Vocabulary. At Grades 1 and 2, however, the correlations between Comprehension and reading course grades were essentially the same, on the average, as those between Word Decoding (called "Vocabulary" in the Third Edition) and reading course grades.

In Grades 7 and 8, the correlations between the Third Edition and reading course grades were much lower, on the average, than in Grade 6 and below. In many schools, it is mainly students whose reading achievement is low who attend classes specifically labeled "Reading." Thus, in Grades 7 and 8, the range of reading achievement in many reading courses is quite limited, and low correlations are the usual consequence of a restricted range.

Correlations with "reading" course grades were generally higher than those with either "language" or "English" course grades. Typically, language and English courses require a great deal of reading, but the grades in such courses are also based on other aspects of achievement, such as writing, spelling, English usage, and oral language.

Correlations with grade point averages (GPAs) were generally as high as those with language or English, even though the GPA typically includes subject areas, such as art and mathematics, that do not depend as heavily on reading ability. At Grade 5 and above, where GPAs frequently pool the judgments of more than one teacher, the median correlations of Third Edition Total score with GPA were consistently higher than those with either language or English. Those relatively high correlations were probably partly due to the greater reliability of GPAs, which frequently pool the judgments of more than one teacher.

References and Notes

1. Bruner, J. (2000). Reading for possible worlds. *National Reading Conference Yearbook, 49,* 31–40.

2. Moats, L. C. (2000). *Speech to print.* Baltimore, MD: Brookes.

   Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

3. MacGinitie, W. H., & MacGinitie, R. (1989). *Technical Report for the Third Edition of the Gates-MacGinitie Reading Tests.* Rolling Meadows, IL: Riverside Publishing.

4. Cunningham, P. M., Moore, S. A., Cunningham, J. W., & Moore, D. W. (2000). *Reading and writing in elementary classrooms: Strategies and observations* (4th ed.). New York: Addison Wesley Longman.

5. Dolch, E. W. (1936). A basic sight vocabulary. *Elementary School Journal, 36,* 456–460.

6. Johns, J. L. (1976). Updating the Dolch basic sight vocabulary. *Reading Horizons, 16,* 104–111.

7. MacGinitie, W. H., & MacGinitie, R. (1989). *Technical Report for the Third Edition of the Gates-MacGinitie Reading Tests.* Rolling Meadows, IL: Riverside Publishing.

8. Instructional Programs and Services Division. (1983). *Basal reading series cross-reference guide.* Olympia, WA: Superintendent of Public Instruction.

   Although this guide was published in 1983, it proved particularly useful for development of the letter-sound correspondences subtests of Level BR and the Word Decoding tests of Levels 1 and 2 because it provides, by grade level of introduction, an extensive list of specific letter-sound correspondences. That the information in this publication is still up-to-date is indicated by how fully it parallels the sequence of skills instruction described in more recent publications, such as

   Chall, J. S., & Popp, H. M. (1996). *Teaching and assessing phonics: Why, what, when, how.* Cambridge, MA: Educators Publishing Service.

   Fry, E. B., Kress, J. E., & Fountoukidis, D. L. (2000). *The reading teacher's book of lists* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

   Chall and Popp describe a teaching sequence in general terms and describe, again in general terms, the phonics instruction appropriate for kindergarten, and for first grade, second grade, and third grade. They note that "most published phonics programs . . . use a sequence similar to the one we suggest" (p. 17). Fry, et. al. include a "Suggested Phonics Teaching Order" that lists specific decoding skills and a recommended sequence for teaching them, but that does not assign the skills to grade levels. These publications, and most others, parallel, as far as they go, the more specific sequence and

grade level designations in the publication from the Instructional Programs and Services Division.

9. Harris, A. J., & Jacobson, M. D. (1982). *Basic reading vocabularies.* New York: Macmillan.

   This reference is still exceptionally useful. Lists of written words that are the most frequent do not change much over time. While technical and academic terms and words used in marketing may change, simple, common words are relatively stable in the written vocabulary, since they are needed for expressing common relationships and describing everyday objects and events. Also, many of the commonest words are integral to English grammar. Certain more recent word counts, such as *The educator's word frequency guide,* are less useful for some purposes because they list inflected forms as separate words. (Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science Associates.)

10. Instructional Programs and Services Division. (1983). *Basal reading series cross-reference guide.* Olympia, WA: Superintendent of Public Instruction. See Note 8.

11. Such a sequence is given by Chall, J. S., & Popp, H. M. (1996). *Teaching and assessing phonics: Why, what, when, how.* Cambridge, MA: Educators Publishing Service. Many other authors give similar lists that are much alike in general outline.

12. Harris, A. J., & Jacobson, M. D. (1982). *Basic reading vocabularies.* New York: Macmillan.

13. Dale, E., & O'Rourke, J. (1976). *The living word vocabulary.* Chicago: Field Enterprises Educational Corporation.

    Dale, E., & O'Rourke, J. (1979). *Supplement to existing 1976 edition of the Living Word Vocabulary.* Chicago: Field Enterprises Educational Corporation.

14. Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book.* Boston: Houghton Mifflin.

15. Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science Associates.

16. "A List of 3000 Words Known by Students in Grade 4" compiled by Edgar Dale (Revised 1983). In Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall Readability Formula.* Cambridge, MA: Brookline Books.

17. Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

18. Francis, W. N. (1958). *The structure of American English.* New York: Ronald Press.

19. These percentages are given in Table 5.5 on page 547 of Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

20. Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 681–744). New York: Longman.

21. Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

22. Aborn, M., Rubenstein, H., & Sterling, T. D. (1959). Sources of contextual constraint upon words in sentences. *Journal of Experimental Psychology, 57,* 171–180.

23. Fantauzzo, P. D. (1996). Using standardized tests, observations, and nontraditional assessment techniques to identify specific factors in reading. In L. R. Putnam (Ed.), *How to become a better reading teacher* (pp. 101–112). Englewood Cliffs, NJ: Merrill.

24. MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000–2002). *Linking testing to teaching: A classroom resource for reading assessment and instruction.* Rolling Meadows, IL: Riverside Publishing. There are separate publications for various test levels. The publications for Level 3, Levels 4–6, and Levels 7/9–10/12 include a brief description of how teachers can interview students about their strategies for answering the Vocabulary questions.

25. Harris, A. J., & Jacobson, M. D. (1982). *Basic reading vocabularies.* New York: Macmillan.

26. Dale, E., & O'Rourke, J. (1976). *The living word vocabulary.* Chicago: Field Enterprises Educational Corporation; and Dale, E., & O'Rourke, J. (1979). *Supplement to existing 1976 edition of the Living Word Vocabulary.* Chicago: Field Enterprises Educational Corporation.

27. The professional writers of passages that were used in the published forms were Kaye C. Benson, Jan Gleiter, Guadelupe V. Lopez, Laureen Mar, and Kathleen Thompson.

28. Maria, K. (1990). Reading comprehension instruction: Issues and strategies. Parkton, MD: York Press.

29. Johnston, P. H. (1984). Assessment in reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 147–182). New York: Longman.

30. See, for example, Dymock, S. J. (1999). Learning about text structure. In G. B. Thompson & B. Nicholson (Eds.), *Learning to read: Beyond phonics and whole language* (pp. 174–192). New York: Teachers College Press.

    Britton, B. K. & Black, J. B. (1985). *Understanding expository text.* Hillsdale, NJ: Erlbaum, describes attempts at analyzing the structure of expository texts, the cognitive demands of processing them, and the role of world knowledge. The chapter by Black, "An Exposition on Understanding Expository Text" (pp. 249–267) provides a good background to this work.

31. Venezky, R. L. (2000). The origins of the present-day chasm between adult literacy needs and school literacy instruction. *Scientific Studies of Reading, 4,* 19–39. (Reprinted from *Visible Language, 16,* 113–136, 1982.)

32. Faigley, L., & Meyer, P. (1983). Rhetorical theory and readers' classification of text types. *Text, 3,* 305–325.

33. Faigley & Meyer, op. cit.

34. Dry, H. (1981). Sentence aspect and the movement of narrative time. *Text, 1,* 233–240.

35. Chall, J. S., & Dale, E. (1995). *Manual for the use of the new Dale-Chall Readability Formula.* Cambridge, MA: Brookline Books.

36. Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11,* 513–516.

    Fry, E. B. (1977). Fry's readability graph: Clarifications, validity, and extension to Level 17. *Journal of Reading, 21,* 242–252.

37. Spache, G. D. (1978). *Good reading for poor readers* (Revised 10th ed.). Champaign, IL: Garrard.

38. Dale-Chall grade level estimates of 5–6, 7–8, 9–10, 11–12, and 13–15 were treated as 5.5, 7.5, 9.5, 11.5, and 14 in computing the averages.

39. Spache, G. D. (1978). *Good reading for poor readers* (Revised 10th ed.). Champaign, IL: Garrard.

40. Baylor, B. (1978). *The other way to listen.* New York: Charles Scribner's Sons.

41. Rohmer, H. (1982). *The legend of Food Mountain.* San Francisco: Children's Book Press.

42. In computing these correlations, the one Fry and four Dale-Chall estimates above grade 13 were changed to grade 13, since the authors had assigned a reading difficulty of grade 13 to all passages they regarded as having a reading difficulty above grade 12.

43. Maria, K., & MacGinitie, W. H. (1982). Reading comprehension disabilities: Knowledge structures and non-accommodating text processing strategies. *Annals of Dyslexia, 32,* 33–59.

44. MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000–2002). *Linking testing to teaching: A classroom resource for reading assessment and instruction.* Rolling Meadows, IL: Riverside Publishing. There are separate publications for various test levels. The publications for Level 3, Levels 4–6, and Levels 7/9–10/12 include a brief description of how teachers can interview students about their strategies for answering the Comprehension questions.

45. Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 130–159). Washington, DC: American Council on Education.

46. In developing Level PR, less specific difficulty guidelines were used. The aim was to produce a relatively easy test that would be most useful for locating students who might have difficulty learning to read if weaknesses in the tested areas of background for reading were not taken into account.

47. Since Levels 2 through 10/12 are designed to be used in the fall or spring or at any time throughout the school year, difficulty estimates for midyear

were used in selecting questions for the final test forms. Such estimates should be the best basis for establishing difficulty distributions that would be suitable throughout the year. The midyear difficulty estimates were obtained by linear interpolation from the difficulty indices obtained in the fall field test at a given grade and the fall field test at the next higher grade. (See the section "Field-Test Administration" beginning on page 33 for a description of the field-test design.)

48. The mean *p*-value resulting from the distribution shown in Table 6 would be about optimal for producing a four-choice test of high reliability when scores are corrected for guessing. When scores are not corrected for guessing, as is the case with the *GMRT,* a somewhat higher mean *p*-value would theoretically be optimal [Feldt, L. S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education, 6,* 37–48; Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrica, 17,* 181–193.]. Such a higher *p*-value is not appropriate for the *GMRT,* however, since the correction for guessing is generally unsatisfactory, and, when distractors are carefully devised, as in the *GMRT,* to appeal to students' misunderstandings, the correction overcorrects considerably [Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 130–159). Washington, DC: American Council on Education.].

    A second reason that a difficulty blueprint like that in Table 6 is appropriate for the *GMRT* is that the average reading achievement of groups of students covers a very wide range. A difficulty distribution like that in Table 6 gives useful derived scores for students with low, average, or very good reading achievement for their grade level (see the section "Ceiling and Floor Data," on page 65) but may give less useful derived scores for students with *very* low reading achievement. Since it is now common practice for schools with reading levels below average to administer an easier, out-of-level test, and since out-of-level norms are provided for the *GMRT,* on-level tests that are appropriate in difficulty for schools in which the reading level is about average, and that include many students with very good reading achievement, should be the most useful.

49. Throughout the test development process, whenever *p*-values were averaged, the averaging was done through transforming the *p*-values to difficulty indices on an equal-interval scale. The transformation was the same as that used in converting PRs to NCEs.

50. Exceptions were a few questions in Levels PR, BR, 1, and 2 in which various common animals were pictured. One Native American consultant noted that many of these pictured animals were of special significance to one or more of the many groups of Native Americans. The concerns of this consultant were accommodated to the extent possible. However, the names of these animals (e.g. bear, snake) are especially useful in many questions as names likely to be known by nearly all students, so, when no suitable substitute could be found to fulfill the intent of the question, some of these pictures representing animal names were retained.

51. Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201–219). Macmillan: New York.

52. Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17*(1), 31–44.

53. It is generally recommended that the question being analyzed *not* be excluded from the total test score when conducting a DIF analysis, e.g., Lewis, C. (1993). A note on the value of including the studied item in the test score when analyzing test items for DIF. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 317–320). Hillsdale, NJ: Erlbaum.

54. Penfield, R. D., & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice, 19*(3) 5–15.

55. Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Newbury Park, CA: Sage.

56. Penfield & Lam, op. cit.

57. Clauser & Mazor, op. cit.

58. Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Erlbaum.

59. Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–348). Hillsdale, NJ: Erlbaum.

60. Zieky, op. cit.

61. MacGinitie, W. H. (1981). *Technical Summary for the Second Edition of the Gates-MacGinitie Reading Tests.* New York: Teachers College Press.

62. Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Iowa tests of basic skills,* Form K. Rolling Meadows, IL: Riverside Publishing.

63. Feldt, L. S., Forsyth, R. A., Ansley, T. N., & Alnot, S. D. (1994). *Iowa tests of educational development,* Form K. Rolling Meadows, IL: Riverside Publishing.

64. Davis, F. B. (1949). *Item-analysis data: Their computation, interpretation, and use in test construction.* Cambridge, MA: Graduate School of Education, Harvard University.

65. Attali, Y., & Fraenkel, T. (2000). The point-biserial as a discrimination index for distractors in multiple-choice items: Deficiencies in usage and an alternative. *Journal of Educational Measurement, 37,* 77–86. Although this article deals with the point-biserial, the general argument applies also to the biserial.

66. Some teachers expressed confusion as to whether students who had completed the Comprehension test by the first (30-minute) time check should be included among the students who had completed the test by the second (40-minute) time check. While, in many cases, it was clear how the teacher had interpreted the instructions, in other cases it was not.

Therefore, the data for the first time check (plus the time for "all but the very slowest-working students" to finish) were the main data considered in establishing the number of questions for the Comprehension tests.

67. The second blank question after the last attempted question was considered to be "not reached." All questions beyond this question were also "not reached." This is a commonly used criterion, since it is impossible to tell from an answer document whether the first question beyond the last answered question was not reached or simply omitted.

68. Harris, A. J., & Jacobson, M. D. (1982). *Basic reading vocabularies.* New York: Macmillan.

69. Harris & Jacobson, op. cit.

70. Dale, E., & O'Rourke, J. (1976). *The living word vocabulary.* Chicago: Field Enterprises Educational Corporation; and Dale, E., & O'Rourke, J. (1979). *Supplement to existing 1976 edition of the Living Word Vocabulary.* Chicago: Field Enterprises Educational Corporation.

71. Harris & Jacobson, op. cit.

72. Dale, & O'Rourke, op. cit.

73. Does not apply to Level 2 Word Knowledge, since all answer choices at that level were single words.

74. The median testing date for this winter standardization was February 24. The norms were adjusted by linear interpolation to represent the period January 24–31. This was done so the winter norms would apply to the same period as the winter norms for Level BR and all higher test levels. See the section "Appropriate Norms for Your Testing" in the *Manual for Scoring and Interpretation* for Level 1.

75. Quality Education Data, Inc. (1998). Denver, CO: Author.

76. Mahar, M. (Ed.). (1994). *NCEA/Ganley's Catholic Schools in America.* Montrose, CO: Fisher Publishing.

77. Quality Education Data, Inc. (1998). Denver, CO: Author.

78. The median standardization testing date for Level AR was October 26.

79. *Guidance information system* (Version 3.0) [Computer software]. (1998). Rolling Meadows, IL: Riverside Publishing.

80. Linacre, J. M., & Wright, B. D. (1998). *WINSTEPS* [Computer software]. Chicago: MESA Press.

81. For familiarity and ease of reading, the word *question* has been used in earlier sections of this report. In the sections "Norms Development" and "Test Characteristics", however, the equivalent term *item* will be used. That is because *item* is standard usage in many phrases describing norms development and the statistical characteristics of tests.

82. See the section "Adjusted Norms" of the *Manual for Scoring and Interpretation* for any level of the *GMRT,* Fourth Edition, for a listing of the quartermonth dates.

83. See Note 82.

84. *Power test* is used here in the sense of the classical distinction between power tests and speed tests.

85. In determining the percentage of students who had completed the test when the testing was stopped, a student was counted as having completed the test if the last question attempted by the student was the last or next-to-last question in the test. This is a commonly-used criterion, since it is impossible to tell whether the last question, if not attempted, was not reached or simply omitted.

86. The individual tests in Levels PR and BR are referred to as "subtests" because they are short and therefore less reliable than the longer individual tests in Levels 1 through 10/12 and AR. If the subtest scores for Levels PR and BR were given as PRs, with 99 score points, the PRs might imply greater precision than would be warranted.

87. See, for example, Chapter 5, "Recording Oral Reading," in Roller, C. M. (1998). *So . . . what's a tutor to do?* Newark, DE: International Reading Association.

88. MacGinitie, W. H., & MacGinitie, R. (1989). *Technical Report for the Third Edition of the Gates-MacGinitie Reading Tests.* Rolling Meadows, IL: Riverside Publishing.

APPENDIX A
Characteristics of Comprehension Passages

Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]	Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]
	Level 1, Form S						
1–4	FI	N	L,L,L,I				
5–8	FI	S	L,L,L,L				
9–12	SS	N	L,L,L,I				
13–16	FI	N	L,I,L,I				
17–20	NS	N	L,L,L,I				
21–24	NS	E	L,L,L,I				
25–28	FI	N	L,L,I,I				
29–32	NS	E	L,L,L,L				
33–36	FI	N	L,L,L,I				
37–39	SS	E	L,L,I				
	Level 2, Form S				**Level 2, Form T**		
1–4	FI	N	L,L,I,I	1–4	FI	N	I,L,L,L
5–8	SS	E	I,L,L,L	5–8	NS	E	L,L,L,L
9–12	SS	N	L,L,I,L	9–12	FI	N	L,I,L,I
13–16	NS	N	I,L,L,I	13–16	SS	E	L,L,I,I
17–20	FI	N	I,L,L,I	17–20	SS	N	I,I,I,I
21–24	FI	S	L,I,I,I	21–24	NS	N	L,L,L,L
25–28	NS	E	L,I,I,L	25–28	FI	S	L,L,L,L
29–32	FI	N	I,L,L,I	29–32	SS	E	L,I,L,L
33–36	SS	E	L,L,L,L	33–36	FI	N	L,I,L,I
37–39	NS	E	L,L,L	37–39	NS	E	L,I,L
	Level 3, Form S				**Level 3, Form T**		
1–4	FI	N	L,I,I,I	1–5	FI	N	I,L,I,L,I
5–8	SS	E	L,L,L,L	6–8	FI	N	L,I,I
9–13	SS	E	I,L,I,I,I	9–13	NS	E	I,I,I,I,L
14–16	NS	E	I,L,L	14–16	SS	E	I,L,I
17–19	FI	N	I,I,I	17–21	FI	N	I,I,L,L,L
20–25	FI	N	L,I,I,L,L,I	22–27	SS	N	I,L,L,I,L,I
26–29	NS	E	L,L,L,L	28–30	NS	E	L,L,I
30–34	FI	N	L,L,L,I,I	31–35	SS	E	L,L,L,I,L
35–39	FI	N	I,I,L,I,I	36–40	FI	N	I,L,I,I,I
40–44	SS	N	I,I,L,L,I	41–43	NS	E	L,L,L
45–48	NS	E	L,I,I,L	44–48	FI	N	L,I,L,I,I

[a] FI = fiction; SS = social science; NS = natural science; HU = humanities.

[b] The order in which the questions appear following the passage. For example, L,L,I,L means that the first two questions are literal questions, the third question is an inferential question, and the fourth (last) question is a literal question.

Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]	Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]
Level 4, Form S				**Level 4, Form T**			
1–3	FI	N	L,I,I	1–4	FI	N	I,L,I,I
4–6	SS	E	L,L,L	5–7	NS	E	L,I,L
7–12	FI	N	I,L,L,L,L,L	8–13	SS	N	I,L,L,L,L,I
13–17	SS	N	I,L,L,I,I	14–17	NS	E	L,L,L,L
18–22	FI	N	L,I,L,L,I	18–21	FI	S	L,I,L,I
23–27	NS	E	I,L,L,L,I	22–27	FI	N	I,I,L,I,L,I
28–30	FI	S	L,I,L	28–31	SS	E	I,I,L,I
31–34	NS	E	L,I,I,L	32–35	FI	N	L,I,I,I
35–38	FI	N	I,I,I,I	36–39	NS	E	L,L,L,L
39–42	SS	E	L,L,L,L	40–42	SS	N	L,I,I
43–48	NS	N	I,I,I,I,I,I	43–48	FI	N	I,I,I,L,I,I
Level 5, Form S				**Level 5, Form T**			
1–3	FI	N	L,I,L	1–5	FI	S	L,L,L,I,I
4–6	FI	N	L,L,L	6–10	SS	N	I,L,I,L,L
7–11	FI	S	L,L,L,L,L	11–13	HU	E	I,I,I
12–14	SS	E	L,I,L	14–18	NS	E	I,I,L,L,I
15–19	NS	E	I,I,I,L,I	19–21	SS	N	L,L,I
20–25	HU	N	L,L,L,I,I,L	22–27	FI	N	I,I,L,I,L,I
26–30	NS	E	I,I,L,L,I	28–31	NS	E	I,L,I,L
31–34	SS	E	I,I,I,I	32–36	FI	N	I,L,L,I,L
35–38	SS	N	L,L,I,L	37–40	FI	N	I,L,I,I
39–43	FI	N	I,L,I,I,I	41–43	SS	E	I,L,I
44–48	NS	E	I,I,I,I,L	44–48	NS	E	L,L,I,L,L
Level 6, Form S				**Level 6, Form T**			
1–3	SS	E	I,I,L	1–3	SS	N	L,L,I
4–7	SS	N	L,L,I,I	4–6	FI	S	L,L,L
8–11	NS	E	L,L,L,L	7–10	FI	N	L,L,L,I
12–14	FI	S	L,I,L	11–15	HU	E	L,L,L,I,L
15–19	FI	N	L,L,L,I,I	16–21	FI	N	L,I,I,I,I,I
20–23	SS	E	I,L,I,I	22–26	SS	E	L,I,I,I,I
24–27	NS	E	I,L,L,I	27–29	NS	E	I,L,I
28–33	FI	N	I,I,L,I,I,I	30–34	NS	E	I,I,L,I,L
34–38	HU	N	I,I,I,L,I	35–38	NS	E	I,L,I,I
39–42	FI	N	L,L,L,L	39–43	FI	N	L,L,I,I,I
43–48	NS	E	L,I,L,I,L,I	44–48	SS	N	L,L,L,L,L
Level 7/9, Form S				**Level 7/9, Form T**			
1–5	FI	N	I,L,I,I,I	1–3	FI	N	I,I,I
6–9	FI	N	I,L,I,I	4–6	SS	E	L,L,I
10–15	SS	E	L,L,L,I,L,I	7–12	FI	N	L,L,I,L,L,I
16–19	NS	N	I,I,I,L	13–17	SS	N	I,I,I,I,I
20–23	FI	N	I,L,L,I	18–21	HU	E	L,L,L,I
24–28	NS	N	L,L,L,L,I	22–27	NS	E	I,L,L,L,L,L
29–31	FI	S	I,L,I	28–30	FI	S	L,I,I
32–35	NS	E	I,L,L,L	31–35	NS	E	L,I,L,L,L
36–38	SS	E	I,I,L	36–39	SS	N	I,I,I,I
39–43	SS	E	L,L,I,L,L	40–42	NS	E	I,L,L
44–48	HU	E	L,I,I,L,I	43–48	FI	N	L,L,L,I,I,I

[a] FI = fiction; SS = social science; NS = natural science; HU = humanities.

[b] The order in which the questions appear following the passage. For example, L,L,I,L means that the first two questions are literal questions, the third question is an inferential question, and the fourth (last) question is a literal question.

Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]	Passage (Question Numbers)	Passage Type[a]	Narrative, Exposition, or Setting	Literal or Inferential in Question Order[b]
Level 10/12, Form S				**Level 10/12, Form T**			
1–4	FI	N	I,L,L,I	1–3	FI	N	L,I,I
5–8	SS	N	L,I,I,L	4–6	NS	E	I,I,I
9–14	HU	E	L,L,I,L,I,I	7–12	FI	N	I,L,I,I,I,I
15–19	FI	S	I,I,I,I,I	13–15	FI	S	I,I,I
20–23	NS	E	I,I,I,L	16–18	HU	E	I,L,I
24–28	FI	N	L,I,L,I,I	19–24	NS	E	L,L,L,L,L,L
29–33	NS	E	L,L,L,L,I	25–28	SS	N	I,L,I,I
34–37	SS	E	L,I,I,L	29–33	FI	N	L,L,L,L,L
38–42	FI	N	I,L,I,L,L	34–38	SS	E	L,L,L,L,I
43–45	SS	E	L,L,I	39–43	NS	E	L,L,L,I,L
46–48	NS	E	L,I,L	44–48	SS	E	L,L,I,I,I
Level AR, Form S				**Level AR, Form T**			
1–4	SS	E	L,L,L,I	1–6	FI	N	I,I,L,L,I,I
5–9	FI	N	I,L,I,I,I	7–9	SS	E	L,I,I
10–14	NS	E	L,I,I,L,I	10–14	FI	S	I,I,I,I,L
15–19	FI	S	I,I,I,I,I	15–18	NS	N	I,I,L,L
20–23	NS	E	I,L,I,I	19–21	HU	E	L,I,I
24–29	FI	N	I,I,L,L,L,I	22–26	SS	E	L,L,I,I,I
30–32	SS	E	L,L,I	27–30	SS	E	L,I,L,I
33–36	FI	N	I,I,I,L	31–35	FI	N	L,I,I,I,I
37–40	HU	E	L,L,I,I	36–38	NS	E	L,I,L
41–43	SS	N	L,L,I	39–43	FI	N	L,L,I,I,I
44–48	NS	E	I,L,I,L,L	44–48	NS	E	L,L,L,I,L

[a] FI = fiction; SS = social science; NS = natural science; HU = humanities.

[b] The order in which the questions appear following the passage. For example, L,L,I,L means that the first two questions are literal questions, the third question is an inferential question, and the fourth (last) question is a literal question.

APPENDIX B
Schools Participating in the Standardization of Levels PR through 10/12

School District	School	City
Alabama		
Vestavia Hills City School District	Vestavia Hills Elementary School-Central	Birmingham
Vestavia Hills City School District	Vestavia Hills Elementary School-East	Birmingham
Alaska		
Archdiocese of Anchorage Education Office	Saint Mary's School	Kodiak
Saint Mary's City School District	Saint Mary's School District	Saint Mary's
Arizona		
Aguila School District 3	Aguila Elementary School	Aguila
Cartwright School District	Cartwright Elementary School	Phoenix
Casa Grande Elementary District 4	Casa Grande Elementary District 4	Casa Grande
Deer Valley Unified School District #97	Deer Valley Unified School District #97	Phoenix
Flowing Wells Unified School District 8	Robert Richardson Elementary	Tucson
Greenlee County Office of Education	Greenlee County Alternative School	Clifton
Mesa Unified School District 3	Franklin East Elementary School	Mesa
Mesa Unified School District 4	Red Mountain Ranch Elementary School	Mesa
Paradise Valley Unified School District 68	Boulder Creek Elementary School	Phoenix
Paradise Valley Unified School District 69	Larkspur Elementary School	Phoenix
Sunnyside Unified School District #12	Sunnyside High School	Tucson
Arkansas		
Atkins School District	Atkins School	Atkins
Fountain Hill School District	Fountain Hill School	Fountain Hill
Hazen School District	Hazen Elementary School	Hazen
Ouachita School District	Ouachita Elementary School	Donaldson
Stone County School District 1	Timbo School	Timbo
Union School District 2	Union School	El Dorado
California		
Archdiocese of Los Angeles Education Office	Verbum Dei High School	Los Angeles
Beverly Hills Unified School District	El Rodeo Elementary School	Beverly Hills
Charter Oak Unified School District	Charter Oak Unified School District	Covina
Diocese of Monterey Education Office	Holy Cross School Elementary and Junior High School	Santa Cruz
Diocese of Monterey Education Office	Old Mission School	San Luis Obispo
Downey Unified School District	Edward C. Lewis Elementary	Downey
El Monte City School District	Durfee Elementary School	El Monte
El Monte City School District	Wilkerson Elementary School	El Monte
Etiwanda School District	Etiwanda School District	Etiwanda
Hermosa Beach City School District	Hermosa Valley School	Hermosa Beach
Hermosa Beach City School District	Hermosa View School	Hermosa Beach
Lompoc Unified School District	Miguelito Elementary School	Lompoc
Los Angeles Unified School District Crenshaw-Dorsey	Watts Learning Center	Los Angeles
Los Angeles Unified School District Eagle Rock-Frank-Marshall	John Marshall High School	Los Angeles
Los Angeles Unified School District Grant-Van Nuys	Sylvan Park Elementary School	Van Nuys

School District	School	City
Los Angeles Unified School District Hamilton-Palisades-University	Palisades Elementary Charter School	Pacific Palisades
Los Angeles Unified School District Hamilton-Palisades-University	Palms Elementary School	Los Angeles
Los Angeles Unified School District Jordan-Locke	Alain LeRoy Locke Senior High School	Los Angeles
Los Angeles Unified School District Jordan-Locke	Gompers Middle School	Los Angeles
Los Angeles Unified School District Jordan-Locke	Ritter Elementary School	Los Angeles
McFarland Unified School District	Browning Road Elementary School	McFarland
Montebello Unified School District	La Merced Intermediate School	Montebello
Palos Verdes Peninsula Unified School District	Cornerstone at Pedregal School	Palos Verdes
Palos Verdes Peninsula Unified School District	Miraleste Intermediate School	Palos Verdes Estates
Placer Hills Union School District	Weimar Hills Middle School	Weimar
Redondo Beach Unified School District	Redondo Beach Unified School District	Redondo Beach
Riverside Garden School	Riverside Garden School	Riverside
Riviera Hall Lutheran School	Riviera Hall Lutheran School	Redondo Beach
San Antonio SDA Junior Academy	San Antonio SDA Junior Academy	Ontario
San Gabriel Unified School District	Coolidge Elementary School	San Gabriel
San Gabriel Unified School District	Gabrielino High School	San Gabriel
San Gabriel Unified School District	Jefferson Middle School	San Gabriel
San Gabriel Unified School District	Roosevelt Elementary School	San Gabriel
San Gabriel Unified School District	Washington Elementary School	San Gabriel
San Gabriel Unified School District	Wilson Elementary School	San Gabriel
Saugus Union Elementary School District	Emblem Elementary School	Saugus
West Covina Unified School District	Wescove Elementary School	West Covina
Colorado		
Academy School District 20	Academy Edison Elem. School	Colorado Springs
Alamosa School District R-11-J	Polston Primary School	Alamosa
Cherry Creek School District 5	Village East Community Elementary School	Aurora
Cotopaxi School District R-3	Cotopaxi Consolidated School	Cotopaxi
Diocese of Pueblo Education Office	Saint Columba School	Durango
Jefferson County School District R-1	Belmar Elementary School	Lakewood
Jefferson County School District R-1	Columbine Senior High School	Littleton
Jefferson County School District R-1	Mandalay Middle School	Westminister
Jefferson County School District R-1	Moore Middle School	Arvada
Jefferson County School District R-1	Wilmore-Davis Elementary School	Wheat Ridge
Van Dellen Christian School	Van Dellen Christian School	Denver
Connecticut		
New Haven City School District	Bishop Woods Elementary School	New Haven
New Haven City School District	Worthington Hooker Elementary School	New Haven
Simsbury School District	Tootin' Hills Elementary School	West Simsbury
Delaware		
Appoquinimink School District	Townsend Elementary School	Townsend
District of Columbia		
District of Columbia Public Schools	Ellington School of the Arts	Washington
District of Columbia Public Schools	Hine Junior High School	Washington
Nannie H. Burroughs School	Nannie H. Burroughs School	Washington

School District	School	City
Florida		
Archdiocese of Miami Education Office	Saint Mary Cathedral School	Miami
Baker County School District	MacClenny Elementary School	MacClenny
Diocese of Orlando Education Office	Our Lady of Lourdes School	Daytona Beach
Diocese of Orlando Education Office	Resurrection Catholic School	Lakeland
Diocese of Orlando Education Office	Saint Thomas Aquinas School	Saint Cloud
Gadsden County School District	Carter-Parramore Middle School	Quincy
Gadsden County School District	Havana Middle School	Havana
Hernando County School District	J. D. Floyd Elementary School	Spring Hill
Jefferson County School District	Howard Middle School	Monticello
Jefferson County School District	Jefferson County High School	Monticello
Lee County School District	J. Colin English Elementary School	North Ft. Myers
Leon County Schools	Canopy Oaks Elementary School	Tallahassee
Seminole County School District	English Estates Elementary School	Fern Park
Seminole County School District	Longwood Elementary School	Longwood
Seminole County School District	Pine Crest Elementary School	Sanford
State of Florida	A. D. Henderson University School	Boca Raton
Washington County School District	Vernon Middle School	Vernon
Georgia		
Griffin-Spalding County School District	Griffin-Spalding County School District	Griffin
Irwin County School District	Irwin County Elementary School	Ocilla
Irwin County School District	Irwin County High School	Ocilla
Hawaii		
Kauai School District	Waimea Canyon School	Waimea
Idaho		
Mullan School District 392	Mullan Elementary School	Mullan
Mullan School District 392	Mullan Junior Senior High School	Mullan
Nampa School District 131	Skyview High School	Nampa
Sugar-Salem Joint District 322	Thomas D. Kershaw Intermediate School	Sugar City
Illinois		
Belleville Public School District 118	Belleville PS 118	Belleville
Berwyn North School District 98	Jefferson Elementary School	Berwyn
Carmi-White County School District 5	Jefferson Elementary School	Carmi
Carmi-White County School District 5	Lincoln Elementary School	Carmi
Central A & M Community School District 21	Central A & M High School	Moweaqua
Chicago Heights Elementary School District 170	Greenbriar Elementary School	Chicago Heights
Chicago Public School District-Region 6	Amelia Earhart Elementary School	Chicago
Community Unit School District 300	Dundee Highlands Elementary School	Dundee
Diocese of Joliet Education Office	Our Lady of Peace School	Darien
Diocese of Joliet Education Office	Saint Joseph School	Addison
Dolton-Riverdale Area School District 148	Dolton-Riverdale Area School District 148	Riverdale
Hampton School District #29	Hampton Elementary School	Hampton
Harlem Unit School District 122	Marquette ElementarySchool	Machesney
Main Township High School District 207	Maine West High School	Des Plaines
Schaumburg Community Consolidated School District 54	Hanover Highlands Elementary School	Hanover Park
Indiana		
Gary Community School District	Thomas A. Edison Middle School	Gary
Gary Community School District	West Side High School	Gary
Saint Peter Lutheran School	Saint Peter Lutheran School	North Judson

School District	School	City
Kansas		
Diocese of Wichita Education Office	Saint James Catholic School	Augusta
Diocese of Wichita Education Office	Saint Joseph Catholic School	Wichita
Diocese of Wichita Education Office	Saint Margaret Mary Catholic School	Wichita
Diocese of Wichita Education Office	Saint Patrick Catholic School	Wichita
Diocese of Wichita Education Office	Saint Peter Catholic School	Schulte
Madison-Virgil Unified School District 386	Madison Elementary School	Madison
Madison-Virgil Unified School District 386	Madison Junior Senior High School	Madison
Northern Valley Unified School District 212	Almena Elementary School	Almena
Northern Valley Unified School District 212	Northern Valley High School	Almena
Pratt Unified School District 382	Pratt Unified School District 382	Pratt
Wichita Public Schools	Chester Lewis Open Magnet Elementary School	Wichita
Wichita Unified School District 259	Benton Elementary School	Wichita
Kentucky		
Hardin County School District	Rineyville Elementary School	Rineyville
Kenton County School District	Piner Elementary School	Morning View
Kenton County School District	RC Hinsdale Elementary School	Edgewood
Kenton County School District	Summit View Middle School	Independence
Pike County School District	Johns Creek Elementary School	Pikeville
Pike County School District	Majestic Knox Creek Elementary School	Majestic
Pike County School District	Pike County Central High School	Pikeville
Raceland-Worthington Independent School District	Campbell Elementary School	Raceland
Raceland-Worthington Independent School District	Worthington Elementary School	Worthington
Louisiana		
Concordia Parish School District	Vidalia Lower Elementary School	Vidalia
Diocese of Lafayette Education Office	Saint Michael Elementary School	Crowley
Monroe City School District	Monroe City School District	Monroe
Maine		
Baileyville Union 107	Woodland Elementary School	Baileyville
Maryland		
Anne Arundel County Public Schools	Anne Arundel County Public Schools	Ft. Meade
Massachusetts		
Blue Hills Region Voc. School District	Blue Hills Region Tech High School	Canton
Danvers School District	Danvers School District	Danvers
Greenfield Public School District	North Parish Elementary School	Greenfield
Haverhill Public Schools	Tilton Elementary School	Haverhill
Lawrence School District	Haverhill Street School	Lawrence
Marshfield School District	Eames Way Elementary School	Marshfield
Nashoba Regional School District	Florence Sawyer School	Bolton
North Brookfield School District	North Brookfield Elementary School	North Brookfield
Revere Public Schools	James Garfield Community Magnet School	Revere
Revere Public Schools	Paul Revere Elementary School	Revere
Michigan		
Diocese of Kalamazoo Education Office	Lake Michigan Catholic School	Saint Joseph
Marshall Public School District	Marshall Middle School	Marshall
Plymouth-Canton Community Schools	Bird Elementary School	Plymouth
Traverse City Area Public Schools	Traverse City East Junior High School	Traverse City
Warren Consolidated School District	Warren Consolidated School District	Warren

School District	School	City
Minnesota		
Blooming Prairie Independent School District 756	Blooming Prairie Elementary School	Blooming Prairie
Round Lake School District 516	Round Lake School	Round Lake
Saint John's Lutheran School	Saint John's Lutheran School	Maple Grove
Mississippi		
Aberdeen School District	Aberdeen Elementary School	Aberdeen
Diocese of Jackson Education Office	Sacred Heart School	Southaven
Diocese of Jackson Education Office	Saint Elizabeth School	Clarksdale
Diocese of Jackson Education Office	Saint Mary School	Jackson
Diocese of Jackson Education Office	Saint Patrick School	Meridian
Grenada School District	Grenada School District	Grenada
Indianola School District	Indianola School District	Indianola
Madison County School District	Madison County School District	Flora
New Albany Public School District	Ford Elementary School	New Albany
New Albany Public School District	Mattie Thompson Elementary School	New Albany
New Albany Public School District	New Albany Middle School	New Albany
Quitman County School District	Quitman County Elementary School	Lambert
Shaw School District	McEvans Elementary School	Shaw
West Tallahatchie School District	R. H. Bearden Elementary School	Sumner
Wilkinson County School District	Wilkinson County School District	Woodville
Yazoo City Municipal School District	Yazoo City Municipal School District	Yazoo City
Missouri		
Francis Howell School District	Francis Howell High School	Saint Charles
Rockwood School District R6	Chesterfield Elementary School	Chesterfield
Rockwood School District R6	Ridge Meadows Elementary School	Balwin
Waynesville School District R6	Waynesville Middle School	Waynesville
Montana		
Bozeman School District 7	Bozeman High School	Bozeman
Ophir School District 72	Ophir Elementary School	Gallatin Gateway
South Stacey School District 90	South Stacey Elementary School	Volborg
Springhill School District 20	Springhill Elementary School	Belgrade
Sunburst School District 2	Sunburst Elementary School	Sunburst
Nebraska		
Archdiocese of Omaha Education Office	All Saints Catholic School	Omaha
Beemer School District 55	Bemmer School	Beemer
Custer County Schools	Custer County District 17 School	Weissert
Custer County Schools	Custer County District 63 School	Miller
Custer County Schools	Custer County District 66 School	Broken Bow
Custer County Schools	Custer County District 75 School	Broken Bow
Custer County Schools	Custer County District 153 School	Gothenburg
Custer County Schools	Custer County District 164 School	Broken Bow
Custer County Schools	Custer County District 169 School	Mason City
Custer County Schools	Custer County District 234 School	Berwyn
Custer County Schools	Custer County District 256 School	Oconto
Custer County Schools	Custer County District 284 School	Gothenburg
Custer County Schools	Custer County District 523 School	Broken Bow
Lancaster County Schools	Lancaster County District 13 School	Lincoln
Lincoln Public School District 1	Lincoln Public School District 1	Lincoln
Louisville School District 32	Louisville School	Louisville
Winnebago Public School District 17	Winnebago School	Winnebago
Nevada		
Las Vegas Junior Academy	Las Vegas Junior Academy	Las Vegas

School District	School	City
New Hampshire		
District 49 Governor Wentworth Regional School District	Carpenter Elementary School	Wolfeboro
District 49 Governor Wentworth Regional School District	Tuftonboro Central School	Tuftonboro
Salem Christian School	Salem Christian School	Salem
New Jersey		
Camden Diocese Education Office	Saint Peter Celestine School	Cherry Hill
Egg Harbor City School District	Charles Spragg Elementary School	Egg Harbor
Elizabeth School District	Christopher Columbus Elementary School 15	Elizabeth
Elizabeth School District	Marquis De Lafayette Elementary School 6	Elizabeth
Hillside School District	Hillside High School	Hillside
Saddle River Day School	Saddle River Day School	Saddle River
New Mexico		
Albuquerque School District-Northeast Region	Bel-Air Elementary School	Albuquerque
Albuquerque School District-Northeast Region	Dennis Chavez Elementary School	Albuquerque
Albuquerque School District-Northwest Region	Cochiti Elementary School	Albuquerque
Farmington Municipal School District 5	Piedra Vista High School	Farmington
Las Vegas City School District	Las Vegas City School District	Las Vegas
Penasco Independent School District	Penasco Elementary School	Penasco
Penasco Independent School District	Penasco Junior Senior High School	Penasco
New York		
Buffalo City Schools	PS 31 Early Childhood Center	Buffalo
Central Islip Union School District	Francis J. O'Neill Elementary School	Central Islip
Cheektowaga-Sloan Union Free School District	John F. Kennedy Middle School	Cheektowaga
Cheektowaga-Sloan Union Free School District	Theodore Roosevelt Primary School	Cheektowaga
Cheektowaga-Sloan Union Free School District	Woodrow Wilson Elementary School	Sloan
Depew Union Free School District	Cayuga Heights Elementary School	Depew
Depew Union Free School District	Depew High School	Depew
East Ramapo Central School District	M.L. Colton Intermediate School	Spring Valley
East Ramapo Central School District	Summit Park Primary School	New City
Jamestown City School District	Jamestown City School District	Jamestown
Madrid-Waddington Central School District	Madrid-Waddington Central School	Madrid
Onteora Central School District	Phoenicia Elementary School	Phoenicia
Trinity Lutheran School	Trinity Lutheran School	Hicksville
Williamsville Central School District	Maple East Elementary School	Williamsville
North Dakota		
Salund School District 10	Salund Elementary School	McLeod
Ohio		
Akron Public Schools	Crouse Elementary School	Akron
Akron Public Schools	Schumacher Academy	Akron
Barnesville Exempted Village School District	Barnesville Elementary School	Barnesville
Barnesville Exempted Village School District	Barnesville Middle School	Barnesville
Canton City School District	Claredon Elementary School	Canton
Dawson-Bryant Local School District	Dawson-Bryant Elementary School	Ironton
Dawson-Bryant Local School District	Dawson-Bryant Middle School	Coal Grove
Fairport Harbor Exempt Village School District	Fairport Harbor Junior Senior High School	Fairport Harbor
Fairport Harbor Exempt Village School District	McKinley Elementary School	Fairport Harbor
Huron City School District	Woodlands Elementary School	Huron

School District	School	City
Oklahoma		
Ada School District 19	Hayes Elementary School	Ada
Ada School District 19	Washington Elementary School	Ada
Archdiocese of Oklahoma City Education Office	All Saints Catholic School	Norman
Archdiocese of Oklahoma City Education Office	Saint John Nepomuk School	Yukon
Archdiocese of Oklahoma City Education Office	Saint Philip Neri School	Midwest City
Bray-Doyle School District 42	Bray-Doyle Elementary School	Bray
Bray-Doyle School District 42	Bray-Doyle High School	Bray
Clinton School District 99	Clinton High School	Clinton
Clinton School District 99	Clinton Middle School	Clinton
Durant School District I-72	Durant School District I-72	Durant
Laverne School District	Laverne Elementary School	Laverne
Laverne School District	Laverne Junior Senior School	Laverne
Marietta School District 16	Marietta High School	Marietta
Moore School District I-2	Moore School District I-2	Moore
Mustang Public School District I-69	Mustang Trails Elementary School	Mustang
Mustang Public School District I-69	Mustang Valley Elementary School	Oklahoma City
Oakdale Elementary School District I-69	Oakdale Elementary School	Edmond
Oklahoma Christian School	Oklahoma Christian School	Edmond
Oklahoma City School District I-89	Oklahoma City School District I-89	Oklahoma City
Oologah-Talala School District 4	Oologah-Talala School District 4	Oologah
Ponca City Public School District	Ponca City Public School District	Ponca City
Rocky Mountain School District 24	Rocky Mountain Elementary School	Stillwell
Tulsa Independent School District 1	Cherokee Elementary School	Tulsa
Wynona School District 30	Wynona School	Wynona
Yukon School District I-27	Central Elementary School	Yukon
Oregon		
Sherwood School District 88J	Sherwood School District 88J	Sherwood
Pennsylvania		
Archdiocese of Philadelphia Education Office	Mercy Vocational High School	Philadelphia
Delaware County Intermediate Unit 25	Delaware County Intermediate Unit 25	Media
Harmony Area School District	Harmony Area School	Westover
Lebanon School District	Lebanon School District	Lebanon
Philadelphia School District	Philadelphia School District	Philadelphia
Scranton City School District	Scranton City School District	Scranton
Selinsgrove Area School District	Selinsgrove Elementary School	Newport
Wyoming Valley West School District	State Street Elementary Center	Larksville
Rhode Island		
Glocester School District	West Glocester Elementary School	Chepachet
South Carolina		
Charleston County School District 10-Saint Andrews	Springfield Elementary School	Charleston
Charleston County School District 20-Charleston	Rivers Middle School	Charleston
Diocese of Charleston Education Office	Saint Anne School	Sumter
Diocese of Charleston Education Office	Summerville Catholic School	Summerville
South Dakota		
Rutland School District 39-4	Rutland School	Rutland
Stickney 1-2	Stickney School	Stickney
Tennessee		
Oak Ridge School District	Linden Elementary School	Oak Ridge
Union Academy	Union Academy	Laconia
Wilson County School District	Wilson County School District	Lebanon

School District	School	City
Texas		
Azle Independent School District	Azle Independent School District	Azle
Blue Ridge Independent School District	Blue Ridge High School	Blue Ridge
Blue Ridge Independent School District	Blue Ridge Middle School	Blue Ridge
Bosqueville Independent School District	Bosqueville School	Waco
Goose Creek Consolidated Independent School District	David Crockett Elementary School	Baytown
Grand Prairie Independent School District	Kennedy Middle School	Grand Prairie
Grand Prairie Independent School District	Lee Middle School	Grand Prairie
Houston Independent School District-South Area	Grissom Elementary School	Houston
San Antonio Independent School District	San Antonio Independent School District	San Antonio
Silsbee Independent School District	Silsbee High School	Silsbee
Spring Branch Independent School District	Cornerstone Academy	Houston
Spring Branch Independent School District	Northbrook Middle School	Houston
Spring Branch Independent School District	Spring Branch Education Center	Houston
Sunnyvale Independent School District	Sunnyvale School	Sunnyvale
Vysehrad Independent School District	Vysehrad Elementary School	Hallettsville
Utah		
Cache County School District	Willow Valley Middle School	Hyrum
Davis County School District	Windridge Elementary School	Kaysville
Morgan County School District	Morgan Middle School	Morgan
San Juan School District	Monument Valley High School	Monument Valley
Vermont		
Moretown School District	Moretown Elementary School	Moretown
South Burlington School District 16	Chamberlin Elementary School	South Burlington
South Burlington School District 16	South Burlington School District 16	South Burlington
Washington		
Archdiocese of Seattle Education Office	Saint Edward's School	Seattle
Central Kitsap School District 401	Silverdale Elementary School	Silverdale
Clover Park School District 400	Hillside Elementary School	Fort Lewis
Diocese of Spokane Education Office	All Saints School	Spokane
Diocese of Spokane Education Office	Saint Patrick School	Pasco
North Beach School District 64	North Beach High School	Ocean Shores
North Beach School District 64	North Beach Middle School	Ocean Shores
Yelm Community School District 2	McKenna Elementary School	McKenna
Yelm Community School District 2	Mill Pond Intermediate School	Yelm
Yelm Community School District 2	Yelm High School	Yelm
Yelm Community School District 2	Yelm Middle School	Yelm
Yelm Community School District 2	Yelm Prairie Elementary School	Yelm
Wisconsin		
Christ the Lord Lutheran School	Christ the Lord Lutheran School	Brookfield
Good Shepherd Lutheran School	Good Shepherd Lutheran School	West Bend
Somerset School District	Somerset Elementary School	Somerset
Somerset School District	Somerset Middle School	Somerset
Wyoming		
Park County School District 16	Meeteetse School	Meeteetse
Sweetwater County School District 2	Wilson Elementary School	Green River

APPENDIX C
Schools Participating in the Equating Studies

School District	School	City
Alabama		
Morgan County Schools	Falkville Elementary	Falkville
Arizona		
Cedar Unified School District #25	Jeddito School	Keams Canyon
Flowing Wells Schools	Homer Davis Elementary	Tucson
Gilbert Public School	Islands Elementary	Gilbert
Gilbert Unified School District	Finley Farms Elementary	Gilbert
Arkansas		
Bismarck	Bismarck School	Bismarck
Dardanelle School District	Dardanelle School District	Dardanelle
Greenland School District	Greenland Public School	Greenland
Lakeside	Lakeside Primary	Hot Springs
California		
Archdiocese of Los Angeles	Our Lady of Fatima School	Artesia
Bakersfield Adventist Academy	Bakersfield Adventist Academy School	Bakersfield
Cabrillo College	Cabrillo College	Aptos
Conejo Valley Unified School District	Acacia	Thousand Oaks
Crafton Hills College	Crafton Hills College	Yucaipa
Diocese of San Jose	Holy Family Educational Center	San Jose
Grossmont-Cuyamaca Community College District	Grossmont-Cuyamaca Community College District	El Cajon
Monterey Diocese	Junipero Serra School	Carmel
Morgan Hill Unified School District	Morgan Hill Unified School District	Morgan Hill
Mountain Elementary School District	Mountain Elementary School	Soquel
Mountain View School District	Bubb School	Mountain View
Orange County Christian School	Orange County Christian School	Anaheim
Palo Verde College	Palo Verde College	Blythe
Southern California Conference	Southern California Conference School	Glendale
San Ramon Unified School District	Pine Valley Middle School	Danville
San Ramon Valley Christian Academy	San Ramon Valley Christian Academy	Danville
Southeastern California Conference Seventh-day Adventist	Victor Valley Seventh-day Adventist School	Victorville
Southern California Conference of Seventh-day Adventist	South Bay Junior Academy	Glendale
Temple City Unified	Cloverly Temple City	Temple City
West Covina Hills Adventist School	West Covina Hills Adventist School	West Covina
Colorado		
Agate School District	Agate School District	Agate
Jefferson County Schools	Arvada Middle	Golden
Connecticut		
Andover	Andover	Andover
Granby Public Schools	Granby Public Schools	Granby
New Haven	West Hills Middle School	New Haven
Trumbull	Daniels Farm School	Trumbull
Trumbull	Trumbull	Trumbull
District of Columbia		
District of Columbia Public Schools	District of Columbia Public Schools	Washington

School District	School	City
Florida		
Archdiocese of Miami	Corpus Christi Catholic School	Miami
Archdiocese of Miami	Saint Francis Xavier School	Miami
Diocese of Orlando	All Souls Catholic School	Orlando
Diocese of Orlando	Holy Redeemer Catholic School	Kissimmee
Diocese of Orlando	Saint Peter School	Deland
Florida National College	Florida National College	Miami
Florida National College	Florida National College-Hialeah	Hialeah
Hillsborough County Schools	Walden Lake Elementary	Plant City
Leon County School District	Belle Vue Middle School	Tallahassee
Leon County Schools	Buck Lake Elementary School	Tallahassee
Saint James Cathedral School	Saint James Cathedral School	Orlando
Georgia		
Berrien County	Berrien High School	Nashville
Coweta County School System	Coweta County School System	Newnan
Richmond County Board of Ed	Richmond County Board of Ed	Augusta
Idaho		
Pocatello District #25	Lewis and Clark Elementary	Pocatello
Illinois		
Annawan Unit 226	Annawan Grade School	Annawan
Chicago Public Schools	Clemente Community Academy HS	Chicago
Germantown School District #60	Germantown School District #60	Germantown
Illinois Eastern Community Colleges-Olney	Illinois Eastern Community Colleges-Olney	Olney
LaGrange District 105	Seventh Avenue School	LaGrange
Lebanon Community Unit School District #9	Summerfield Grade School	Summerfield
Lisle Community Unit School District #202	Tate Woods School	Lisle
Prairie State College	Prairie State College	Chicago Heights
Prairieview District 192	Prairieview District 192	Royal
Schaumburg School District	Dirksen School	Schaumburg
Schaumburg School District	Dr. Thomas Dooley School	Schaumburg
Schaumburg School District	Frost Junior High School	Schaumburg
Schaumburg School District	Keller Junior High School	Schaumburg
Schaumburg School District	MacArthur Elementary School	Hoffman Estates
Schaumburg School District	Mead Junior High School	Elk Grove
Schaumburg School District	Robert Frost Junior High	Schaumburg
Schaumburg School District	Winston Churchill Elementary School	Schaumburg
Saint Paul's Lutheran School	Saint Paul's Lutheran School	Brookfield
Indiana		
Anderson Community Schools	Westvale Elementary School	Anderson
Bethel Christian School	Bethel Christian School	Princeton
Columbus Seventh-day Adventist School	Columbus Seventh-day Adventist School	Columbus
Diocese of Evansville	Saint John School	Newburgh
Metropolitan School District of Wayne Township	Garden City Elementary School	Indianapolis
Metropolitan School District of Wayne Township	Maplewood Elementary School	Indianapolis
Metropolitan School District of Wayne Township	McClelland Elementary School	Indianapolis
Metropolitan School District of Wayne Township	Rhoades Elementary School	Indianapolis
Metropolitan School District of Wayne Township	Westlake Elementary School	Indianapolis
Orleans Community Schools	Orleans Elementary School	Orleans
School City of East Chicago	School City of East Chicago	East Chicago
Trinity Lutheran School	Trinity Lutheran School	Fort Wayne

School District	School	City
Kansas		
Blue Valley School District	Blue Valley	Overland Park
Cowley College	Cowley College	Arkansas City
Kentucky		
Hazard Community College	Hazard Community College	Hazard
Louisiana		
Delgado Community College	Delgado Community College	New Orleans
Diocese of Lafayette	Saint Ignatius Elementary	Grand Coteau
Morehouse Parish Schools	Morehouse Parish Schools	Bastrop
Winn Parish	Winnfield Intermediate School	Winnfield
Maryland		
Anne Arundel County Public Schools	Anne Arundel County Public Schools-2	Fort Meade
Anne Arundel County Public Schools	Anne Arundel County Public Schools-3	Fort Meade
Anne Arundel County Public Schools	Chesapeake Bay Middle School	Pasadena
Lindale/Brooklyn Park Middle School	Lindale/Brooklyn Park Middle School	Linthicum
Massachusetts		
Bristol Community College	Bristol Community College	Fall River
Michigan		
Great Lakes College	Great Lakes College	Bad Axe
Memorial Lutheran Ingham	Memorial Lutheran School	Williamston
Plymouth-Canton Community Schools	Gallimore Elementary School	Canton
Plymouth-Canton Community Schools	Field Elementary	Canton
Minnesota		
LeCenter Public Schools	LeCenter Public Schools	LeCenter
Mississippi		
Greenville Public Schools	Greenville Public Schools	Greenville
Grenada School District	Grenada School District	Grenada
Holmes Community College	Holmes Community College	Ridgeland
Itawamba Community College	Itawamba Community College	Fulton
Jackson County Schools	East Central Upper Elementary School	Pascagoula
Leake County Schools	South Leake School	Walnut Grove
Lee County	Guntown Middle School	Guntown
Lee County	Verona School	Verona
Meridian Public Schools	Carver Middle School	Meridian
Pascagoula Municipal School District	College Park Elementary School	Gautier
Pascagoula Municipal Separate Schools	Beach Elementary School	Pascagoula
Pascagoula Municipal Separate Schools	Central Elementary School	Pascagoula
Presbyterian Day School	Presbyterian Day School	Cleveland
Shannon Elementary School	Shannon Elementary School	Shannon
South Panola	Batesville Intermediate School	Batesville
Missouri		
Diocese of Jefferson City	Saint Martin School	Jefferson City
Francis Howell School District	Saeger Accelerated Middle School	Saint Charles
Marshall Public Schools	Marshall Public Schools	Marshall
Trinity Lutheran	Trinity Lutheran-Cape District	Cape Girardeau
Waynesville R-VI	East Elementary	Waynesville
Waynesville R-VI	Thayer Elementary School	Fort Leonard
Waynesville R-VI	Williams Elementary School	Waynesville
Waynesville R-VI	Wood Middle School	Waynesville
Montana		
Little Big Horn College	Little Big Horn College	Crow Agency

School District	School	City
Nebraska		
Hampton Lutheran School	Hampton Lutheran School	Hampton
Immanuel Lutheran School	Immanuel Lutheran School	Columbus
New Hampshire		
EG Sherburne School	EG Sherburne School	Pelham
Gov. Wentworth Regional School District	Kingswood Reg. Middle School	Wolfeboro Falls
New Jersey		
Irvington	Grove Saint School	Irvington
Logan Township	Logan Township	Swedesboro
Saddle Brook	Long Memorial School	Saddle Brook
New Mexico		
Albuquerque Public Schools	Adobe Acres Elementary School	Albuquerque
Albuquerque Public Schools	Jackson Middle School	Albuquerque
Albuquerque Public Schools	S.Y. Jackson Elementary School	Albuquerque
Jal Public Schools	Jal Elementary School	Jal
New York		
Chittenango Central Schools	Chittenango Central Schools	Chittenango
East Ramapo Central School District	Colton Elementary School	Spring Valley
East Ramapo Central School District	Eldorado Elementary School	Spring Valley
East Ramapo Central School District	Elmwood Elementary School	Spring Valley
East Ramapo Central School District	Lime Kiln Elementary School	Spring Valley
North Carolina		
Catawba Valley Community College	Catawba Valley Community College	Hickory
Oklahoma		
Cashion Public School	Cashion Public School	Cashion
Marlow Public Schools	Marlow Public Schools	Marlow
Yukon Public Schools	Lakeview Middle School	Yukon
Yukon Public Schools	Yukon Mid-High School	Yukon
Oregon		
Central Oregon Community College	Central Oregon Community College	Bend
Pennsylvania		
Beaver Area School District	Beaver Area School District	Beaver
Carmichaels Area	Carmichaels Area	Carmichaels
Centennial School District	Leary Elementary School	Warminster
Diocese of Harrisburg	Holy Family School	Harrisburg
Ebenezer Faith Christian School	Ebenezer Faith Christian School	Plymouth
Harrisburg Diocese	Saint Joseph School	Mechanicsburg
Kennett Consolidated School District	Kennett High School	Kennett Square
Philadelphia	Andrew Hamilton School	Philadelphia
Philadelphia School District	George Washington HS	Philadelphia
Portage Area School District	Portage Area School District	Portage
West Chester Area School District	Fugett Middle School	West Chester
South Carolina		
Diocese of Charleston	Blessed Sacrament School	Charleston
Diocese of Charleston	Saint Anthony of Padua Catholic School	Greenville
Tennessee		
Knoxville Diocese	Saint John Neumann School	Knoxville
West Carroll Special School District	West Carroll Special School District	Trezevant
Texas		
Harlandale Independent School District	Harlandale High School	San Antonio
North East Independent School District	Harmony Hills Elementary	San Antonio
Olfen Independent School District	Olfen Independent School District	Rowena
Silsbee Independent School District	Silsbee Middle School	Silsbee

School District	School	City
Utah		
Davis County School District	Doxey Elementary School	Sunset
Davis County School District	Layton Elementary School	Layton
Davis County School District	Knowlton Elementary School	Farmington
Davis County School District	Monte Vista Elementary School	Farmington
Davis County School District	South Clearfield Elementary	Clearfield
Park City School District	Treasure Mountain Middle School	Park City
Washington		
Issaquah School District	Issaquah School District	Issaquah
LaCenter School District 101	LaCenter School	LaCenter
Skykomish	Skykomish School	Skykomish
Wisconsin		
Bryant Stratton College-Milwaukee Campus	Bryant Stratton College-Milwaukee Campus	Milwaukee
Germantown School District	MacArthur School	Germantown
Milwaukee Public Schools	Rufus King High School	Milwaukee
Sheboygan County Christian High	Sheboygan County Christian High	Sheboygan
Saint Paul's Ev. Lutheran School	Saint Paul's Ev. Lutheran School	Wonewoc
Wyoming		
Fremont County School District #24	Shoshoni Elementary	Shoshoni

APPENDIX D
Community Colleges Participating in the Standardization of Level AR

District or Authority	College	City
Arkansas		
Rich Mountain Community College	Rich Mountain Community College	Mena
California		
Cabrillo College	Cabrillo College	Aptos
Fremont-Newark Community College District	Ohlone Community College	Fremont
Grossmont-Cuyamaca Community College District	Grossmont College	El Cajon
Hartnell College	Hartnell College	Salinas
Palo Verde Community College District	Palo Verde College	Blythe
San Bernardino Community College District	Crafton Hills College	Yucaipa
San Jose Community College District	Evergreen Valley College	San Jose
Delaware		
Delaware Tech & Community College	Delaware Tech & Community College-Stanton Campus	Newark
Florida		
Florida National College	Florida National College	Hialeah
Georgia		
University System of Georgia	Gainesville College	Gainesville
Hawaii		
University of Hawaii	Kapi'olani Community College	Honolulu
Illinois		
Illinois Eastern Community Colleges	Illinois Eastern Community Colleges-Olney Central College	Olney
Prairie State College	Prairie State College	Chicago Heights
Kansas		
Cowley County Community College	Cowley County Community College	Arkansas City
Kentucky		
Hazard Community College	Hazard Community College	Hazard
Louisiana		
Delgado Community College	Delgado Community College	New Orleans
Massachusetts		
Bristol Community College	Bristol Community College	Fall River
Michigan		
Great Lakes College	Great Lakes College	Bad Axe
Minnesota		
Central Lakes College	Central Lakes College	Brainerd
Mississippi		
Holmes Community College	Holmes Community College-Ridgeland Campus	Ridgeland
Itawamba Community College	Itawamba Community College-Fulton Campus	Fulton
Montana		
Little Big Horn College	Little Big Horn College	Crow Agency

District or Authority	College	City
New Jersey		
Camden County College	Camden County College	Blackwood
North Carolina		
Anson Community College	Anson Community College	Polkton
Catawba Valley Community College	Catawba Valley Community College	Hickory
Durham Technical Community College	Durham Technical Community College	Durham
Southeastern Community College	Southeastern Community College	Whiteville
Ohio		
Bowling Green State University	Bowling Green State University-Firelands College	Huron
Hocking College	Hocking College	Nelsonville
Northwestern College	Northwestern College	Lima
Oklahoma		
Oklahoma City Community College	Oklahoma City Community College	Oklahoma City
Platt College	Platt College	Tulsa
Oregon		
Central Oregon Community College	Central Oregon Community College	Bend
Southwestern Oregon Community College	Southwestern Oregon Community College	Coos Bay
Pennsylvania		
Butler County Community College	Butler County Community College	Butler
Texas		
Blinn College	Blinn College	Brenham
Central Texas College	Central Texas College	Killeen
Grayson County College	Grayson County College	Denison
Texas State Technical College	Texas State Technical College	Harlingen
Virginia		
Johnson and Wales University	Johnson and Wales University	Norfolk
Wytheville Community College	Wytheville Community College	Wytheville
Wisconsin		
Bryant Stratton College	Bryant Stratton College-Milwaukee Campus	Milwaukee

APPENDIX E
Standardization Questionnaire

A questionnaire was sent to each participating school at the time the school agreed to participate. The person coordinating the testing at the school was asked to fill out the questionnaire. The questionnaire stated that the questions were being asked in order to obtain information about the student sample participating in the *Gates-MacGinitie Reading Tests* National Standardization Study. The schools were assured that the information would be aggregated and used only as descriptive data for the entire sample—that individual students, classes, schools, and/or districts would not be identified.

The questions on the questionnaire and the aggregated responses of the participating schools are shown below. The percentages in question 6 were computed from response averages, without weighting for number of students tested.

1. Do you currently use Riverside products? [*N* = 301]

60.5% Yes **39.5%** No

2. Do you currently use the *Gates-MacGinitie Reading Tests* ? [*N* = 301]

24.6% Yes **75.4%** No

3. How many students are enrolled in your school/district? [*N* = 265]

1,174 (average of schools/districts)

4. Which of the following categories best describes the community that your school/district serves? (Check one) [*N* = 301]

20.6% Urban **39.2%** Suburban **40.2%** Rural

5. What is the size of your district? (Check one) [*N* = 294]

35.7% Less than 1,000 students
10.5% 1,000 to 2,499 students
13.6% 2,500 to 4,999 students
12.9% 5,000 to 9,999 students
15.0% 10,000 to 24,999 students
12.2% 25,000 or more students

6. Percents of groups composing school enrollment: [*N* = 269]

16.4% African-American/Black, not of Hispanic origin
4.0% Asian/Pacific Islander
14.2% Hispanic
2.8% Native American
61.7% White, not of Hispanic origin
0.9% Other

APPENDIX F
Item Difficulties

List of Tables

Table 30. Item Difficulties (*p*-values)
Level PR, Form S

Level PR															
Kindergarten, Spring								Grade 1, Fall							
Lit[a]		Oral[b]		LSC[c]		LC[d]		Lit[a]		Oral[b]		LSC[c]		LC[d]	
Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*
1	0.98	21	0.82	41	0.98	71	0.86	1	0.95	21	0.85	41	0.98	71	0.88
2	0.80	22	0.84	42	0.89	72	0.66	2	0.80	22	0.88	42	0.93	72	0.72
3	0.90	23	0.69	43	0.91	73	0.54	3	0.92	23	0.72	43	0.93	73	0.61
4	0.95	24	0.72	44	0.92	74	0.65	4	0.92	24	0.78	44	0.94	74	0.73
5	0.95	25	0.71	45	0.97	75	0.80	5	0.94	25	0.76	45	0.97	75	0.83
6	0.79	26	0.64	46	0.95	76	0.65	6	0.84	26	0.63	46	0.96	76	0.70
7	0.54	27	0.52	47	0.93	77	0.71	7	0.65	27	0.52	47	0.96	77	0.80
8	0.80	28	0.54	48	0.96	78	0.66	8	0.85	28	0.67	48	0.96	78	0.69
9	0.69	29	0.53	49	0.88	79	0.55	9	0.78	29	0.60	49	0.92	79	0.62
10	0.90	30	0.46	50	0.94	80	0.80	10	0.92	30	0.54	50	0.97	80	0.84
11	0.85	31	0.90	51	0.86	81	0.76	11	0.87	31	0.90	51	0.94	81	0.75
12	0.88	32	0.73	52	0.82	82	0.66	12	0.91	32	0.69	52	0.89	82	0.70
13	0.91	33	0.73	53	0.86	83	0.56	13	0.93	33	0.75	53	0.91	83	0.63
14	0.86	34	0.64	54	0.87	84	0.73	14	0.89	34	0.65	54	0.91	84	0.79
15	0.73	35	0.66	55	0.82	85	0.80	15	0.83	35	0.68	55	0.87	85	0.82
16	0.80	36	0.68	56	0.72	86	0.67	16	0.83	36	0.69	56	0.82	86	0.71
17	0.79	37	0.70	57	0.69	87	0.77	17	0.84	37	0.69	57	0.75	87	0.79
18	0.72	38	0.69	58	0.77	88	0.49	18	0.80	38	0.70	58	0.85	88	0.62
19	0.79	39	0.60	59	0.69	89	0.48	19	0.84	39	0.57	59	0.79	89	0.56
20	0.81	40	0.52	60	0.78	90	0.43	20	0.84	40	0.52	60	0.89	90	0.50
				61	0.90							61	0.95		
				62	0.81							62	0.91		
				63	0.92							63	0.94		
				64	0.85							64	0.94		
				65	0.66							65	0.68		
				66	0.83							66	0.90		
				67	0.63							67	0.73		
				68	0.52							68	0.53		
				69	0.69							69	0.81		
				70	0.53							70	0.51		

[a] Literacy Concepts
[b] Oral Language Concepts
[c] Letters and Letter-Sound Correspondences
[d] Listening (Story) Comprehension

Table 31. Item Difficulties (*p*-values)
Level BR, Form S

Level BR, Grade 1[a]																
Fall								Spring								
Initial[b]		Final[c]		Vowels		BSW[d]		Initial[b]		Final[c]		Vowels		BSW[d]	
Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*	Item No.	*p*
1	0.79	26	0.83	16	0.30	46	0.67	1	0.90	26	0.94	16	0.66	46	0.91
2	0.84	27	0.85	17	0.45	47	0.67	2	0.93	27	0.95	17	0.72	47	0.91
3	0.72	28	0.66	18	0.52	48	0.88	3	0.88	28	0.83	18	0.75	48	0.94
4	0.72	29	0.53	19	0.38	49	0.82	4	0.90	29	0.77	19	0.62	49	0.94
5	0.69	30	0.57	20	0.31	50	0.69	5	0.87	30	0.85	20	0.63	50	0.90
6	0.49	31	0.71	21	0.83	51	0.70	6	0.70	31	0.83	21	0.95	51	0.90
7	0.53	32	0.69	22	0.48	52	0.68	7	0.76	32	0.87	22	0.78	52	0.86
8	0.35	33	0.59	23	0.57	53	0.70	8	0.53	33	0.70	23	0.89	53	0.88
9	0.59	34	0.63	24	0.29	54	0.61	9	0.81	34	0.74	24	0.59	54	0.83
10	0.58	35	0.58	25	0.26	55	0.66	10	0.83	35	0.76	25	0.61	55	0.85
11	0.86	36	0.40			56	0.46	11	0.95	36	0.60			56	0.73
12	0.80	37	0.42	41	0.61	57	0.73	12	0.93	37	0.55	41	0.80	57	0.91
13	0.56	38	0.43	42	0.52	58	0.48	13	0.80	38	0.60	42	0.71	58	0.79
14	0.55	39	0.35	43	0.64	59	0.66	14	0.89	39	0.51	43	0.85	59	0.81
15	0.49	40	0.36	44	0.36	60	0.86	15	0.85	40	0.55	44	0.54	60	0.94
16	See Vowels[a]			45	0.33	61	0.35	16	See Vowels[a]			45	0.50	61	0.65
						62	0.73							62	0.91
						63	0.52							63	0.79
						64	0.77							64	0.89
						65	0.79							65	0.93
						66	0.56							66	0.86
						67	0.71							67	0.91
						68	0.75							68	0.91
						69	0.56							69	0.75
						70	0.59							70	0.81

[a] Some item numbers for Level BR are not listed consecutively in this table because the items of the Vowels subtest are all grouped in the "Vowels" column, even though five of the items are administered at the end of the first testing session, five at the beginning of the second testing session, and five at the end of the second testing session.
[b] Initial Consonants and Consonant Clusters
[c] Final Consonants and Consonant Clusters
[d] Basic Story Words

Table 32. Item Difficulties (*p*-values)
Levels 1, 2, and 3, Form S

Item No.	Level 1, Spring		Level 2, Grade 2						Level 3, Grade 3				Item No.
			Fall			Spring			Fall		Spring		
	WD[a]	Comp	WD[a]	WK[b]	Comp	WD[a]	WK[b]	Comp	Voc	Comp	Voc	Comp	
1	0.90	0.79	0.89	0.81	0.86	0.94	0.90	0.93	0.94	0.88	0.94	0.91	1
2	0.85	0.70	0.88	0.83	0.80	0.92	0.92	0.88	0.91	0.54	0.92	0.65	2
3	0.86	0.68	0.83	0.81	0.75	0.89	0.89	0.84	0.87	0.68	0.91	0.76	3
4	0.80	0.85	0.84	0.69	0.82	0.91	0.81	0.87	0.82	0.81	0.88	0.87	4
5	0.82	0.78	0.81	0.73	0.78	0.88	0.83	0.87	0.83	0.74	0.88	0.79	5
6	0.85	0.77	0.77	0.65	0.67	0.87	0.76	0.78	0.86	0.72	0.88	0.79	6
7	0.73	0.73	0.77	0.73	0.78	0.79	0.82	0.88	0.64	0.60	0.69	0.70	7
8	0.72	0.78	0.72	0.77	0.73	0.79	0.87	0.83	0.77	0.64	0.79	0.74	8
9	0.72	0.46	0.76	0.66	0.73	0.85	0.73	0.83	0.76	0.73	0.84	0.80	9
10	0.85	0.53	0.66	0.67	0.73	0.74	0.78	0.83	0.67	0.51	0.76	0.57	10
11	0.68	0.71	0.68	0.77	0.57	0.76	0.84	0.68	0.65	0.57	0.71	0.63	11
12	0.64	0.70	0.70	0.63	0.71	0.80	0.72	0.81	0.74	0.50	0.79	0.57	12
13	0.84	0.68	0.72	0.63	0.67	0.81	0.72	0.78	0.65	0.52	0.74	0.59	13
14	0.67	0.73	0.71	0.60	0.79	0.80	0.71	0.86	0.72	0.49	0.76	0.59	14
15	0.71	0.74	0.61	0.69	0.73	0.74	0.79	0.81	0.59	0.77	0.70	0.86	15
16	0.64	0.67	0.73	0.72	0.47	0.85	0.86	0.59	0.64	0.73	0.68	0.80	16
17	0.81	0.33	0.65	0.64	0.73	0.79	0.72	0.84	0.73	0.43	0.81	0.49	17
18	0.66	0.55	0.70	0.61	0.69	0.81	0.74	0.80	0.64	0.60	0.67	0.68	18
19	0.67	0.72	0.73	0.53	0.74	0.86	0.65	0.83	0.60	0.72	0.64	0.78	19
20	0.71	0.49	0.68	0.55	0.56	0.80	0.67	0.65	0.63	0.74	0.71	0.80	20
21	0.62	0.51	0.69	0.58	0.73	0.78	0.72	0.84	0.45	0.61	0.52	0.68	21
22	0.61	0.65	0.64	0.56	0.58	0.74	0.64	0.69	0.57	0.59	0.64	0.67	22
23	0.56	0.72	0.61	0.57	0.58	0.69	0.67	0.69	0.61	0.67	0.68	0.73	23
24	0.53	0.64	0.61	0.46	0.70	0.74	0.58	0.83	0.63	0.69	0.66	0.75	24
25	0.60	0.70	0.62	0.48	0.41	0.72	0.64	0.53	0.55	0.51	0.65	0.61	25
26	0.58	0.56	0.64	0.48	0.31	0.73	0.58	0.41	0.60	0.62	0.67	0.68	26
27	0.54	0.65	0.67	0.45	0.55	0.76	0.53	0.65	0.61	0.65	0.67	0.72	27
28	0.54	0.63	0.62	0.64	0.59	0.72	0.76	0.67	0.57	0.64	0.68	0.71	28
29	0.63	0.62	0.50	0.50	0.74	0.62	0.63	0.84	0.49	0.61	0.54	0.70	29
30	0.42	0.61	0.62	0.41	0.75	0.75	0.54	0.83	0.53	0.69	0.53	0.75	30
31	0.41	0.51	0.58	0.42	0.50	0.68	0.50	0.63	0.43	0.53	0.49	0.59	31
32	0.52	0.42	0.58	0.41	0.53	0.69	0.51	0.66	0.44	0.55	0.54	0.61	32
33	0.51	0.51	0.55	0.37	0.67	0.74	0.48	0.78	0.48	0.40	0.57	0.47	33
34	0.63	0.55	0.53	0.42	0.62	0.64	0.54	0.72	0.61	0.40	0.66	0.46	34
35	0.55	0.73	0.55	0.43	0.50	0.72	0.56	0.61	0.42	0.54	0.52	0.60	35
36	0.44	0.62	0.53	0.43	0.61	0.67	0.53	0.74	0.38	0.45	0.45	0.51	36
37	0.50	0.41	0.39	0.34	0.42	0.50	0.42	0.50	0.40	0.45	0.44	0.51	37
38	0.53	0.46	0.51	0.38	0.66	0.67	0.50	0.75	0.43	0.43	0.49	0.49	38
39	0.51	0.53	0.52	0.38	0.47	0.64	0.49	0.58	0.32	0.38	0.39	0.46	39
40	0.33		0.50	0.32		0.63	0.42		0.41	0.50	0.46	0.58	40
41	0.48		0.56	0.41		0.67	0.49		0.40	0.55	0.47	0.63	41
42	0.35		0.47	0.31		0.60	0.46		0.41	0.29	0.47	0.34	42
43	0.40		0.57	0.34		0.70	0.42		0.34	0.41	0.39	0.49	43
44									0.30	0.53	0.38	0.61	44
45									0.23	0.43	0.29	0.49	45
46										0.26		0.32	46
47										0.48		0.55	47
48										0.41		0.48	48

[a] Word Decoding
[b] Word Knowledge

Table 33. Item Difficulties (*p*-values)
Levels 4, 5, and 6, Form S

Item No.	Level 4, Grade 4				Level 5, Grade 5				Level 6, Grade 6				Item No.
	Fall		Spring		Fall		Spring		Fall		Spring		
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.93	0.74	0.96	0.79	0.91	0.84	0.94	0.86	0.93	0.90	0.93	0.91	1
2	0.88	0.88	0.92	0.90	0.85	0.74	0.87	0.79	0.84	0.65	0.85	0.71	2
3	0.88	0.66	0.88	0.71	0.79	0.61	0.81	0.65	0.85	0.77	0.85	0.78	3
4	0.85	0.62	0.86	0.68	0.79	0.87	0.81	0.90	0.79	0.69	0.81	0.71	4
5	0.77	0.63	0.83	0.64	0.78	0.39	0.80	0.42	0.67	0.83	0.69	0.84	5
6	0.85	0.62	0.89	0.68	0.75	0.86	0.77	0.89	0.79	0.76	0.80	0.77	6
7	0.78	0.45	0.82	0.52	0.57	0.48	0.63	0.51	0.72	0.30	0.75	0.33	7
8	0.70	0.72	0.72	0.78	0.67	0.70	0.70	0.73	0.70	0.89	0.71	0.90	8
9	0.74	0.64	0.79	0.68	0.71	0.72	0.74	0.76	0.61	0.57	0.61	0.62	9
10	0.69	0.77	0.73	0.83	0.65	0.70	0.67	0.72	0.71	0.76	0.73	0.77	10
11	0.76	0.75	0.76	0.82	0.71	0.69	0.73	0.72	0.69	0.65	0.69	0.68	11
12	0.71	0.71	0.75	0.78	0.67	0.56	0.72	0.58	0.69	0.83	0.69	0.84	12
13	0.67	0.82	0.75	0.86	0.62	0.70	0.69	0.73	0.73	0.73	0.78	0.76	13
14	0.70	0.51	0.79	0.58	0.68	0.69	0.70	0.74	0.71	0.82	0.75	0.83	14
15	0.63	0.79	0.71	0.84	0.59	0.70	0.62	0.74	0.66	0.79	0.70	0.80	15
16	0.44	0.78	0.45	0.82	0.45	0.35	0.47	0.39	0.63	0.72	0.70	0.73	16
17	0.71	0.85	0.75	0.88	0.73	0.52	0.75	0.54	0.62	0.45	0.65	0.45	17
18	0.74	0.68	0.75	0.74	0.61	0.44	0.64	0.46	0.56	0.23	0.61	0.27	18
19	0.63	0.47	0.67	0.55	0.58	0.31	0.65	0.35	0.52	0.64	0.56	0.68	19
20	0.64	0.54	0.68	0.60	0.65	0.36	0.67	0.38	0.65	0.70	0.68	0.71	20
21	0.68	0.56	0.73	0.62	0.67	0.80	0.72	0.82	0.66	0.83	0.68	0.84	21
22	0.63	0.73	0.64	0.78	0.52	0.46	0.56	0.49	0.56	0.76	0.58	0.78	22
23	0.38	0.39	0.50	0.43	0.56	0.68	0.60	0.73	0.55	0.80	0.60	0.82	23
24	0.41	0.45	0.44	0.46	0.58	0.63	0.60	0.66	0.52	0.63	0.58	0.65	24
25	0.54	0.75	0.59	0.78	0.54	0.84	0.58	0.86	0.47	0.74	0.53	0.74	25
26	0.53	0.69	0.55	0.74	0.58	0.31	0.64	0.37	0.45	0.58	0.45	0.60	26
27	0.68	0.48	0.71	0.55	0.49	0.67	0.52	0.73	0.45	0.66	0.45	0.66	27
28	0.53	0.39	0.53	0.44	0.50	0.57	0.50	0.60	0.41	0.35	0.47	0.35	28
29	0.49	0.53	0.56	0.58	0.34	0.56	0.42	0.61	0.52	0.47	0.61	0.48	29
30	0.36	0.66	0.47	0.69	0.42	0.54	0.44	0.58	0.61	0.41	0.61	0.41	30
31	0.59	0.43	0.62	0.48	0.44	0.73	0.50	0.76	0.40	0.51	0.49	0.56	31
32	0.35	0.64	0.38	0.68	0.41	0.49	0.51	0.54	0.31	0.50	0.34	0.54	32
33	0.45	0.64	0.50	0.68	0.46	0.70	0.51	0.73	0.41	0.50	0.42	0.51	33
34	0.35	0.41	0.41	0.47	0.43	0.62	0.48	0.67	0.26	0.37	0.28	0.42	34
35	0.44	0.36	0.47	0.43	0.44	0.59	0.46	0.65	0.47	0.54	0.49	0.57	35
36	0.55	0.41	0.59	0.48	0.38	0.47	0.40	0.50	0.46	0.49	0.47	0.55	36
37	0.53	0.54	0.62	0.58	0.44	0.40	0.48	0.43	0.40	0.58	0.50	0.61	37
38	0.38	0.45	0.45	0.52	0.36	0.50	0.41	0.55	0.35	0.75	0.39	0.77	38
39	0.34	0.41	0.40	0.48	0.30	0.62	0.36	0.67	0.50	0.61	0.50	0.67	39
40	0.42	0.42	0.49	0.48	0.32	0.42	0.34	0.49	0.29	0.55	0.31	0.60	40
41	0.40	0.43	0.48	0.51	0.32	0.44	0.35	0.49	0.43	0.42	0.44	0.48	41
42	0.27	0.51	0.32	0.58	0.29	0.63	0.34	0.68	0.33	0.45	0.38	0.52	42
43	0.42	0.40	0.48	0.47	0.27	0.60	0.30	0.66	0.26	0.58	0.31	0.64	43
44	0.25	0.50	0.35	0.57	0.27	0.54	0.32	0.58	0.39	0.34	0.46	0.38	44
45	0.26	0.34	0.34	0.41	0.18	0.42	0.21	0.47	0.22	0.33	0.30	0.38	45
46		0.55		0.62		0.30		0.33		0.35		0.43	46
47		0.33		0.41		0.33		0.41		0.28		0.33	47
48		0.39		0.45		0.28		0.31		0.38		0.46	48

Table 34. Item Difficulties (*p*-values)
Level 7/9, Form S

Item No.	Level 7/9, Grade 7 Fall		Level 7/9, Grade 7 Spring		Level 7/9, Grade 8 Fall		Level 7/9, Grade 8 Spring		Level 7/9, Grade 9 Fall		Level 7/9, Grade 9 Spring		Item No.
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.89	0.57	0.89	0.60	0.89	0.63	0.89	0.65	0.91	0.68	0.91	0.70	1
2	0.81	0.52	0.82	0.56	0.83	0.60	0.84	0.63	0.85	0.66	0.86	0.68	2
3	0.86	0.90	0.86	0.90	0.87	0.90	0.88	0.91	0.89	0.93	0.89	0.95	3
4	0.76	0.78	0.78	0.79	0.78	0.82	0.78	0.83	0.83	0.86	0.84	0.88	4
5	0.79	0.73	0.80	0.73	0.81	0.75	0.81	0.77	0.83	0.79	0.83	0.84	5
6	0.47	0.80	0.49	0.81	0.55	0.87	0.58	0.86	0.58	0.87	0.60	0.90	6
7	0.55	0.80	0.55	0.81	0.61	0.83	0.61	0.84	0.66	0.85	0.67	0.87	7
8	0.68	0.64	0.70	0.64	0.75	0.66	0.75	0.67	0.78	0.69	0.80	0.73	8
9	0.67	0.47	0.69	0.48	0.72	0.52	0.74	0.53	0.76	0.56	0.77	0.59	9
10	0.47	0.56	0.49	0.57	0.53	0.58	0.55	0.59	0.59	0.62	0.60	0.65	10
11	0.59	0.58	0.61	0.58	0.68	0.59	0.72	0.60	0.77	0.61	0.77	0.63	11
12	0.64	0.68	0.70	0.68	0.71	0.73	0.72	0.74	0.74	0.76	0.76	0.80	12
13	0.65	0.49	0.71	0.51	0.74	0.54	0.76	0.57	0.81	0.62	0.83	0.64	13
14	0.67	0.69	0.68	0.69	0.68	0.72	0.69	0.74	0.70	0.78	0.70	0.80	14
15	0.55	0.40	0.55	0.42	0.55	0.46	0.56	0.47	0.56	0.50	0.57	0.53	15
16	0.51	0.65	0.57	0.65	0.58	0.66	0.61	0.68	0.63	0.69	0.65	0.72	16
17	0.57	0.61	0.59	0.63	0.64	0.67	0.66	0.70	0.71	0.71	0.73	0.74	17
18	0.83	0.61	0.83	0.61	0.85	0.62	0.85	0.63	0.88	0.69	0.88	0.69	18
19	0.58	0.75	0.58	0.78	0.65	0.78	0.65	0.80	0.68	0.84	0.70	0.84	19
20	0.59	0.83	0.62	0.84	0.61	0.85	0.65	0.86	0.68	0.89	0.70	0.89	20
21	0.51	0.80	0.57	0.80	0.66	0.80	0.70	0.81	0.78	0.86	0.80	0.86	21
22	0.37	0.67	0.40	0.71	0.43	0.73	0.43	0.76	0.47	0.79	0.47	0.80	22
23	0.69	0.70	0.75	0.73	0.78	0.75	0.80	0.78	0.82	0.80	0.84	0.83	23
24	0.35	0.67	0.40	0.69	0.45	0.71	0.49	0.72	0.52	0.75	0.54	0.77	24
25	0.70	0.70	0.74	0.73	0.74	0.76	0.76	0.78	0.77	0.82	0.78	0.82	25
26	0.35	0.62	0.38	0.62	0.44	0.65	0.49	0.67	0.49	0.70	0.51	0.70	26
27	0.53	0.56	0.58	0.57	0.63	0.61	0.69	0.62	0.71	0.68	0.71	0.68	27
28	0.47	0.33	0.51	0.37	0.51	0.40	0.52	0.42	0.53	0.48	0.54	0.48	28
29	0.30	0.37	0.32	0.40	0.36	0.46	0.36	0.48	0.41	0.49	0.43	0.54	29
30	0.47	0.57	0.49	0.57	0.56	0.60	0.56	0.60	0.62	0.64	0.62	0.64	30
31	0.44	0.45	0.51	0.45	0.56	0.50	0.60	0.51	0.67	0.53	0.70	0.55	31
32	0.29	0.59	0.31	0.64	0.33	0.67	0.36	0.72	0.41	0.73	0.43	0.75	32
33	0.45	0.40	0.48	0.44	0.48	0.48	0.52	0.51	0.52	0.54	0.57	0.56	33
34	0.39	0.64	0.47	0.66	0.48	0.69	0.52	0.72	0.52	0.73	0.55	0.75	34
35	0.37	0.67	0.40	0.69	0.43	0.72	0.43	0.74	0.47	0.75	0.47	0.78	35
36	0.40	0.65	0.43	0.65	0.45	0.69	0.48	0.71	0.50	0.72	0.55	0.75	36
37	0.42	0.51	0.43	0.51	0.46	0.54	0.47	0.55	0.49	0.56	0.56	0.57	37
38	0.35	0.48	0.37	0.52	0.38	0.57	0.41	0.62	0.41	0.63	0.41	0.67	38
39	0.22	0.65	0.23	0.70	0.30	0.73	0.32	0.75	0.33	0.76	0.37	0.78	39
40	0.40	0.56	0.43	0.62	0.43	0.66	0.47	0.68	0.49	0.69	0.49	0.72	40
41	0.42	0.47	0.43	0.52	0.50	0.56	0.50	0.60	0.52	0.61	0.52	0.63	41
42	0.45	0.52	0.50	0.58	0.50	0.58	0.51	0.62	0.51	0.63	0.55	0.67	42
43	0.32	0.53	0.34	0.61	0.39	0.63	0.40	0.66	0.43	0.67	0.46	0.70	43
44	0.29	0.38	0.30	0.39	0.35	0.45	0.35	0.47	0.35	0.48	0.38	0.54	44
45	0.30	0.34	0.35	0.39	0.35	0.43	0.38	0.47	0.38	0.47	0.38	0.52	45
46		0.25		0.25		0.31		0.33		0.35		0.37	46
47		0.35		0.37		0.39		0.45		0.46		0.50	47
48		0.35		0.37		0.42		0.45		0.47		0.51	48

Table 35. Item Difficulties (*p*-values)
Level 10/12, Form S

Item No.	Level 10/12, Grade 10 Fall		Level 10/12, Grade 10 Spring		Level 10/12, Grade 11 Fall		Level 10/12, Grade 11 Spring		Level 10/12, Grade 12 Fall		Level 10/12, Grade 12 Spring		Item No.
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.86	0.87	0.86	0.88	0.87	0.88	0.87	0.89	0.88	0.91	0.89	0.95	1
2	0.87	0.56	0.87	0.57	0.87	0.59	0.87	0.61	0.87	0.63	0.87	0.67	2
3	0.83	0.58	0.83	0.62	0.83	0.63	0.83	0.64	0.83	0.64	0.83	0.72	3
4	0.67	0.80	0.69	0.84	0.73	0.84	0.78	0.85	0.78	0.86	0.79	0.90	4
5	0.78	0.71	0.79	0.73	0.79	0.75	0.81	0.77	0.81	0.77	0.82	0.82	5
6	0.78	0.61	0.78	0.62	0.79	0.64	0.79	0.66	0.80	0.67	0.81	0.68	6
7	0.66	0.75	0.66	0.78	0.67	0.78	0.69	0.80	0.69	0.81	0.69	0.85	7
8	0.78	0.30	0.78	0.33	0.79	0.33	0.79	0.35	0.80	0.35	0.80	0.44	8
9	0.67	0.80	0.68	0.81	0.71	0.81	0.73	0.84	0.73	0.85	0.73	0.86	9
10	0.47	0.63	0.49	0.64	0.50	0.66	0.51	0.69	0.51	0.70	0.52	0.71	10
11	0.63	0.80	0.65	0.81	0.66	0.81	0.67	0.84	0.68	0.85	0.69	0.88	11
12	0.52	0.51	0.52	0.54	0.53	0.54	0.54	0.56	0.56	0.58	0.58	0.59	12
13	0.52	0.54	0.52	0.60	0.55	0.61	0.58	0.62	0.60	0.63	0.60	0.68	13
14	0.57	0.43	0.57	0.44	0.59	0.47	0.59	0.47	0.59	0.50	0.59	0.52	14
15	0.63	0.65	0.65	0.66	0.66	0.66	0.68	0.70	0.68	0.71	0.72	0.71	15
16	0.52	0.58	0.52	0.60	0.54	0.62	0.56	0.64	0.56	0.64	0.56	0.69	16
17	0.54	0.53	0.55	0.57	0.56	0.57	0.59	0.59	0.59	0.60	0.59	0.67	17
18	0.42	0.40	0.42	0.44	0.44	0.45	0.46	0.46	0.48	0.49	0.50	0.52	18
19	0.43	0.62	0.43	0.65	0.46	0.66	0.48	0.69	0.48	0.69	0.48	0.74	19
20	0.57	0.47	0.58	0.47	0.58	0.47	0.60	0.50	0.61	0.50	0.64	0.55	20
21	0.57	0.64	0.59	0.67	0.60	0.68	0.62	0.68	0.62	0.69	0.65	0.74	21
22	0.52	0.39	0.55	0.45	0.56	0.46	0.58	0.49	0.59	0.49	0.64	0.53	22
23	0.48	0.54	0.52	0.55	0.52	0.61	0.52	0.61	0.52	0.62	0.52	0.65	23
24	0.51	0.49	0.51	0.51	0.52	0.55	0.52	0.57	0.53	0.60	0.53	0.63	24
25	0.44	0.60	0.46	0.63	0.47	0.65	0.49	0.65	0.50	0.65	0.55	0.65	25
26	0.55	0.42	0.56	0.45	0.57	0.46	0.58	0.48	0.60	0.48	0.60	0.53	26
27	0.51	0.51	0.53	0.55	0.54	0.56	0.59	0.58	0.59	0.58	0.60	0.62	27
28	0.42	0.61	0.44	0.62	0.45	0.63	0.48	0.63	0.49	0.64	0.50	0.66	28
29	0.41	0.51	0.41	0.51	0.42	0.51	0.42	0.52	0.43	0.54	0.43	0.59	29
30	0.42	0.41	0.42	0.41	0.43	0.46	0.44	0.46	0.45	0.47	0.45	0.48	30
31	0.46	0.26	0.47	0.26	0.48	0.28	0.49	0.29	0.51	0.29	0.53	0.36	31
32	0.40	0.41	0.43	0.43	0.44	0.44	0.46	0.44	0.46	0.44	0.47	0.47	32
33	0.48	0.43	0.51	0.43	0.52	0.47	0.54	0.48	0.58	0.48	0.60	0.52	33
34	0.34	0.68	0.37	0.68	0.39	0.71	0.43	0.71	0.45	0.73	0.45	0.73	34
35	0.45	0.53	0.47	0.53	0.49	0.57	0.50	0.57	0.51	0.58	0.55	0.59	35
36	0.43	0.40	0.43	0.42	0.43	0.43	0.43	0.43	0.47	0.46	0.47	0.51	36
37	0.40	0.64	0.41	0.64	0.43	0.64	0.45	0.65	0.46	0.67	0.46	0.69	37
38	0.33	0.40	0.33	0.40	0.34	0.46	0.35	0.46	0.38	0.48	0.38	0.50	38
39	0.26	0.51	0.26	0.52	0.27	0.52	0.28	0.55	0.29	0.57	0.29	0.59	39
40	0.24	0.53	0.24	0.54	0.25	0.57	0.26	0.57	0.28	0.58	0.28	0.59	40
41	0.39	0.44	0.39	0.45	0.39	0.46	0.39	0.46	0.40	0.50	0.45	0.52	41
42	0.32	0.50	0.33	0.51	0.33	0.51	0.34	0.53	0.34	0.54	0.36	0.61	42
43	0.39	0.20	0.42	0.25	0.42	0.25	0.44	0.25	0.44	0.31	0.49	0.33	43
44	0.19	0.56	0.20	0.57	0.20	0.58	0.21	0.59	0.22	0.61	0.26	0.64	44
45	0.23	0.45	0.25	0.48	0.26	0.48	0.27	0.52	0.31	0.53	0.36	0.57	45
46		0.32		0.37		0.37		0.38		0.39		0.41	46
47		0.46		0.48		0.52		0.52		0.53		0.53	47
48		0.37		0.38		0.38		0.41		0.41		0.45	48

Table 36. Item Difficulties (*p*-values)
Levels 2 and 3, Form T

Item No.	Level 2, Grade 2						Level 3, Grade 3				Item No.
	Fall			Spring			Fall		Spring		
	WD[a]	WK[b]	Comp	WD[a]	WK[b]	Comp	Voc	Comp	Voc	Comp	
1	0.81	0.82	0.78	0.92	0.93	0.88	0.87	0.73	0.93	0.80	1
2	0.80	0.75	0.81	0.91	0.86	0.91	0.84	0.59	0.90	0.66	2
3	0.79	0.78	0.79	0.90	0.89	0.89	0.81	0.71	0.87	0.78	3
4	0.79	0.79	0.77	0.90	0.90	0.87	0.77	0.77	0.83	0.85	4
5	0.78	0.66	0.63	0.89	0.77	0.73	0.79	0.69	0.85	0.76	5
6	0.72	0.63	0.69	0.82	0.74	0.79	0.75	0.65	0.81	0.72	6
7	0.74	0.69	0.67	0.85	0.80	0.77	0.75	0.69	0.81	0.76	7
8	0.75	0.72	0.73	0.86	0.83	0.84	0.60	0.65	0.66	0.72	8
9	0.75	0.64	0.77	0.86	0.75	0.87	0.70	0.64	0.76	0.71	9
10	0.71	0.69	0.72	0.81	0.80	0.83	0.74	0.62	0.80	0.69	10
11	0.71	0.76	0.72	0.82	0.87	0.83	0.81	0.57	0.87	0.64	11
12	0.59	0.58	0.67	0.69	0.69	0.77	0.64	0.65	0.70	0.72	12
13	0.71	0.70	0.69	0.82	0.81	0.79	0.69	0.68	0.75	0.75	13
14	0.74	0.60	0.67	0.85	0.71	0.77	0.67	0.62	0.73	0.69	14
15	0.71	0.61	0.56	0.81	0.72	0.66	0.68	0.64	0.74	0.71	15
16	0.69	0.51	0.77	0.79	0.61	0.88	0.75	0.67	0.81	0.74	16
17	0.73	0.51	0.75	0.84	0.62	0.86	0.65	0.65	0.71	0.72	17
18	0.71	0.58	0.30	0.82	0.69	0.40	0.63	0.54	0.69	0.61	18
19	0.67	0.59	0.69	0.77	0.70	0.79	0.64	0.62	0.70	0.69	19
20	0.71	0.60	0.63	0.81	0.71	0.73	0.48	0.60	0.54	0.67	20
21	0.63	0.63	0.60	0.73	0.74	0.70	0.73	0.76	0.79	0.83	21
22	0.65	0.57	0.71	0.75	0.68	0.82	0.65	0.58	0.71	0.65	22
23	0.64	0.47	0.65	0.74	0.58	0.75	0.67	0.62	0.73	0.69	23
24	0.67	0.51	0.60	0.77	0.62	0.70	0.64	0.42	0.70	0.50	24
25	0.65	0.58	0.66	0.75	0.69	0.76	0.52	0.61	0.58	0.68	25
26	0.66	0.50	0.59	0.76	0.61	0.69	0.64	0.49	0.70	0.57	26
27	0.66	0.49	0.63	0.76	0.60	0.73	0.51	0.48	0.57	0.56	27
28	0.56	0.54	0.59	0.66	0.65	0.69	0.54	0.39	0.60	0.47	28
29	0.62	0.49	0.41	0.72	0.60	0.51	0.49	0.57	0.55	0.64	29
30	0.62	0.49	0.38	0.72	0.60	0.48	0.50	0.38	0.56	0.46	30
31	0.57	0.42	0.47	0.67	0.53	0.57	0.60	0.29	0.66	0.37	31
32	0.62	0.54	0.69	0.72	0.65	0.79	0.47	0.61	0.53	0.68	32
33	0.57	0.37	0.70	0.67	0.48	0.80	0.58	0.54	0.64	0.61	33
34	0.55	0.45	0.62	0.65	0.56	0.72	0.46	0.25	0.52	0.33	34
35	0.56	0.46	0.66	0.66	0.57	0.76	0.47	0.41	0.53	0.49	35
36	0.48	0.36	0.47	0.59	0.47	0.57	0.40	0.54	0.46	0.61	36
37	0.55	0.42	0.57	0.65	0.53	0.67	0.43	0.52	0.49	0.60	37
38	0.53	0.37	0.48	0.63	0.48	0.58	0.39	0.37	0.45	0.45	38
39	0.60	0.40	0.64	0.70	0.51	0.74	0.37	0.52	0.44	0.60	39
40	0.56	0.34		0.66	0.45		0.33	0.53	0.40	0.60	40
41	0.51	0.31		0.62	0.42		0.37	0.36	0.44	0.44	41
42	0.54	0.32		0.64	0.43		0.36	0.61	0.43	0.68	42
43	0.52	0.16		0.63	0.26		0.29	0.54	0.36	0.61	43
44							0.34	0.32	0.41	0.40	44
45							0.32	0.39	0.39	0.47	45
46								0.50		0.58	46
47								0.49		0.57	47
48								0.50		0.58	48

[a] Word Decoding
[b] Word Knowledge

Table 37. Item Difficulties (*p*-values)
Levels 4, 5, and 6, Form T

Item No.	Level 4, Grade 4				Level 5, Grade 5				Level 6, Grade 6				Item No.
	Fall		Spring		Fall		Spring		Fall		Spring		
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.91	0.55	0.96	0.60	0.89	0.73	0.93	0.77	0.88	0.84	0.91	0.87	1
2	0.87	0.73	0.92	0.79	0.87	0.54	0.91	0.58	0.81	0.80	0.84	0.83	2
3	0.84	0.59	0.89	0.64	0.78	0.83	0.82	0.87	0.82	0.82	0.85	0.85	3
4	0.79	0.72	0.83	0.78	0.76	0.76	0.80	0.80	0.83	0.86	0.86	0.89	4
5	0.81	0.83	0.85	0.88	0.63	0.79	0.67	0.83	0.75	0.82	0.78	0.85	5
6	0.80	0.71	0.85	0.77	0.60	0.54	0.64	0.58	0.75	0.37	0.78	0.40	6
7	0.65	0.73	0.70	0.79	0.70	0.81	0.74	0.85	0.71	0.62	0.74	0.65	7
8	0.73	0.79	0.77	0.84	0.66	0.74	0.70	0.78	0.73	0.52	0.76	0.55	8
9	0.69	0.63	0.73	0.68	0.68	0.69	0.72	0.73	0.70	0.63	0.73	0.66	9
10	0.78	0.79	0.82	0.84	0.60	0.73	0.64	0.77	0.71	0.64	0.74	0.67	10
11	0.75	0.85	0.79	0.90	0.61	0.50	0.65	0.54	0.77	0.84	0.80	0.88	11
12	0.55	0.74	0.60	0.80	0.75	0.67	0.79	0.71	0.68	0.61	0.71	0.64	12
13	0.64	0.53	0.69	0.58	0.67	0.57	0.71	0.62	0.70	0.52	0.73	0.55	13
14	0.61	0.71	0.66	0.76	0.75	0.67	0.79	0.71	0.70	0.51	0.73	0.54	14
15	0.64	0.71	0.69	0.76	0.74	0.44	0.78	0.48	0.57	0.80	0.60	0.83	15
16	0.51	0.75	0.56	0.81	0.75	0.77	0.79	0.81	0.68	0.60	0.71	0.63	16
17	0.65	0.24	0.70	0.29	0.59	0.74	0.63	0.78	0.69	0.71	0.72	0.75	17
18	0.69	0.73	0.73	0.79	0.53	0.62	0.57	0.67	0.63	0.57	0.66	0.60	18
19	0.60	0.55	0.65	0.60	0.60	0.55	0.64	0.60	0.60	0.50	0.63	0.53	19
20	0.55	0.71	0.60	0.77	0.37	0.72	0.40	0.76	0.56	0.50	0.59	0.53	20
21	0.62	0.58	0.67	0.63	0.51	0.66	0.55	0.70	0.61	0.43	0.64	0.46	21
22	0.60	0.67	0.65	0.72	0.54	0.58	0.58	0.63	0.46	0.68	0.49	0.71	22
23	0.72	0.71	0.76	0.77	0.63	0.58	0.67	0.63	0.49	0.27	0.52	0.30	23
24	0.76	0.63	0.80	0.68	0.46	0.59	0.50	0.64	0.59	0.68	0.62	0.72	24
25	0.53	0.77	0.58	0.83	0.60	0.48	0.64	0.52	0.55	0.58	0.58	0.61	25
26	0.55	0.73	0.60	0.79	0.50	0.53	0.54	0.57	0.46	0.37	0.49	0.40	26
27	0.60	0.57	0.65	0.62	0.66	0.74	0.70	0.78	0.61	0.74	0.64	0.78	27
28	0.65	0.58	0.70	0.63	0.56	0.73	0.60	0.77	0.40	0.44	0.42	0.47	28
29	0.51	0.63	0.56	0.68	0.37	0.74	0.41	0.78	0.39	0.54	0.42	0.57	29
30	0.50	0.47	0.55	0.52	0.34	0.68	0.37	0.72	0.44	0.44	0.47	0.47	30
31	0.46	0.56	0.51	0.61	0.47	0.44	0.51	0.48	0.55	0.45	0.58	0.48	31
32	0.37	0.80	0.42	0.86	0.44	0.56	0.48	0.61	0.56	0.44	0.59	0.47	32
33	0.49	0.35	0.54	0.40	0.52	0.47	0.56	0.51	0.45	0.57	0.48	0.60	33
34	0.43	0.57	0.48	0.62	0.37	0.35	0.41	0.39	0.35	0.71	0.37	0.75	34
35	0.38	0.41	0.43	0.46	0.35	0.41	0.39	0.45	0.37	0.59	0.39	0.62	35
36	0.49	0.51	0.54	0.56	0.32	0.51	0.35	0.55	0.42	0.44	0.45	0.47	36
37	0.42	0.30	0.47	0.35	0.44	0.27	0.48	0.31	0.42	0.32	0.45	0.35	37
38	0.30	0.39	0.35	0.44	0.43	0.41	0.47	0.45	0.39	0.51	0.42	0.54	38
39	0.47	0.44	0.52	0.49	0.35	0.52	0.38	0.56	0.32	0.62	0.35	0.65	39
40	0.37	0.47	0.42	0.52	0.27	0.33	0.30	0.37	0.39	0.56	0.41	0.59	40
41	0.24	0.55	0.29	0.60	0.38	0.52	0.42	0.56	0.28	0.63	0.31	0.66	41
42	0.34	0.45	0.39	0.50	0.27	0.54	0.30	0.59	0.30	0.49	0.33	0.52	42
43	0.42	0.56	0.47	0.61	0.27	0.54	0.30	0.58	0.29	0.52	0.32	0.55	43
44	0.37	0.53	0.42	0.58	0.35	0.40	0.38	0.44	0.31	0.39	0.34	0.42	44
45	0.38	0.42	0.43	0.47	0.35	0.39	0.39	0.43	0.21	0.49	0.24	0.52	45
46		0.58		0.63		0.38		0.42		0.43		0.46	46
47		0.59		0.64		0.30		0.34		0.47		0.50	47
48		0.44		0.49		0.28		0.32		0.61		0.64	48

Table 38. Item Difficulties (*p*-values)
Level 7/9, Form T

Item No.	Level 7/9, Grade 7				Level 7/9, Grade 8				Level 7/9, Grade 9				Item No.
	Fall		Spring		Fall		Spring		Fall		Spring		
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.85	0.64	0.86	0.66	0.86	0.71	0.87	0.73	0.88	0.74	0.90	0.76	1
2	0.73	0.85	0.76	0.86	0.76	0.86	0.78	0.88	0.81	0.89	0.83	0.91	2
3	0.79	0.75	0.81	0.77	0.81	0.79	0.82	0.81	0.83	0.84	0.85	0.86	3
4	0.72	0.80	0.74	0.82	0.74	0.83	0.76	0.85	0.78	0.87	0.80	0.89	4
5	0.66	0.64	0.69	0.66	0.70	0.69	0.72	0.71	0.78	0.77	0.80	0.79	5
6	0.53	0.61	0.56	0.64	0.56	0.64	0.56	0.66	0.57	0.68	0.57	0.70	6
7	0.59	0.46	0.62	0.48	0.69	0.54	0.71	0.56	0.76	0.59	0.78	0.62	7
8	0.80	0.73	0.83	0.75	0.83	0.76	0.84	0.78	0.85	0.80	0.86	0.82	8
9	0.78	0.78	0.81	0.80	0.81	0.80	0.83	0.82	0.86	0.84	0.88	0.86	9
10	0.58	0.50	0.61	0.52	0.61	0.54	0.63	0.57	0.67	0.60	0.69	0.63	10
11	0.71	0.71	0.74	0.73	0.76	0.73	0.78	0.75	0.78	0.79	0.80	0.81	11
12	0.56	0.56	0.59	0.59	0.61	0.60	0.63	0.62	0.64	0.66	0.66	0.69	12
13	0.68	0.77	0.71	0.79	0.72	0.80	0.74	0.82	0.78	0.84	0.80	0.86	13
14	0.67	0.54	0.69	0.57	0.69	0.61	0.71	0.64	0.76	0.68	0.78	0.70	14
15	0.39	0.50	0.41	0.52	0.53	0.56	0.56	0.59	0.66	0.59	0.68	0.61	15
16	0.65	0.46	0.68	0.48	0.72	0.54	0.73	0.57	0.73	0.61	0.75	0.64	16
17	0.74	0.74	0.77	0.76	0.77	0.78	0.77	0.80	0.77	0.82	0.79	0.84	17
18	0.51	0.55	0.54	0.58	0.57	0.58	0.58	0.59	0.61	0.62	0.62	0.65	18
19	0.53	0.48	0.56	0.50	0.57	0.50	0.60	0.52	0.60	0.56	0.61	0.58	19
20	0.54	0.60	0.57	0.63	0.58	0.70	0.60	0.72	0.61	0.74	0.62	0.76	20
21	0.59	0.65	0.62	0.67	0.66	0.72	0.68	0.74	0.69	0.75	0.71	0.77	21
22	0.44	0.51	0.47	0.53	0.48	0.59	0.50	0.61	0.55	0.64	0.56	0.67	22
23	0.67	0.67	0.69	0.69	0.69	0.74	0.71	0.76	0.73	0.78	0.75	0.80	23
24	0.52	0.65	0.55	0.67	0.60	0.69	0.63	0.73	0.63	0.73	0.65	0.74	24
25	0.57	0.67	0.60	0.69	0.63	0.72	0.65	0.74	0.69	0.75	0.73	0.77	25
26	0.44	0.45	0.47	0.47	0.49	0.49	0.50	0.51	0.57	0.55	0.58	0.57	26
27	0.45	0.65	0.48	0.67	0.51	0.73	0.52	0.75	0.61	0.77	0.62	0.79	27
28	0.45	0.54	0.48	0.57	0.56	0.61	0.58	0.64	0.60	0.66	0.61	0.69	28
29	0.52	0.50	0.55	0.52	0.58	0.58	0.61	0.60	0.63	0.61	0.64	0.64	29
30	0.34	0.55	0.36	0.58	0.42	0.60	0.44	0.63	0.50	0.67	0.51	0.70	30
31	0.35	0.73	0.37	0.75	0.46	0.76	0.48	0.78	0.54	0.79	0.55	0.81	31
32	0.42	0.44	0.45	0.46	0.56	0.49	0.58	0.51	0.58	0.59	0.59	0.61	32
33	0.67	0.57	0.68	0.60	0.68	0.66	0.69	0.68	0.71	0.70	0.73	0.72	33
34	0.31	0.39	0.33	0.41	0.37	0.44	0.39	0.46	0.40	0.47	0.41	0.49	34
35	0.47	0.63	0.50	0.64	0.53	0.64	0.55	0.67	0.56	0.67	0.57	0.70	35
36	0.26	0.25	0.28	0.27	0.35	0.32	0.37	0.34	0.37	0.38	0.38	0.40	36
37	0.45	0.63	0.48	0.65	0.50	0.67	0.52	0.69	0.55	0.69	0.56	0.70	37
38	0.40	0.52	0.43	0.54	0.46	0.56	0.48	0.59	0.50	0.61	0.51	0.64	38
39	0.34	0.31	0.36	0.33	0.37	0.37	0.39	0.39	0.43	0.44	0.44	0.46	39
40	0.35	0.52	0.37	0.55	0.42	0.60	0.43	0.63	0.44	0.65	0.45	0.68	40
41	0.20	0.54	0.22	0.57	0.27	0.60	0.29	0.62	0.31	0.62	0.32	0.65	41
42	0.31	0.51	0.33	0.53	0.37	0.53	0.38	0.55	0.41	0.60	0.42	0.63	42
43	0.37	0.57	0.40	0.60	0.46	0.63	0.48	0.66	0.51	0.66	0.52	0.69	43
44	0.20	0.55	0.22	0.58	0.24	0.64	0.26	0.67	0.26	0.67	0.27	0.70	44
45	0.20	0.46	0.23	0.48	0.26	0.53	0.28	0.55	0.32	0.55	0.33	0.57	45
46		0.42		0.44		0.48		0.50		0.50		0.52	46
47		0.44		0.46		0.54		0.56		0.58		0.61	47
48		0.20		0.22		0.28		0.30		0.35		0.37	48

Table 39. Item Difficulties (*p*-values)
Level 10/12, Form T

Item No.	Level 10/12, Grade 10 Fall		Level 10/12, Grade 10 Spring		Level 10/12, Grade 11 Fall		Level 10/12, Grade 11 Spring		Level 10/12, Grade 12 Fall		Level 10/12, Grade 12 Spring		Item No.
	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	Voc	Comp	
1	0.80	0.75	0.81	0.77	0.82	0.77	0.84	0.78	0.86	0.78	0.88	0.82	1
2	0.60	0.70	0.61	0.72	0.61	0.75	0.62	0.77	0.63	0.82	0.64	0.86	2
3	0.76	0.73	0.77	0.74	0.77	0.75	0.78	0.76	0.80	0.77	0.86	0.78	3
4	0.68	0.53	0.69	0.55	0.71	0.61	0.75	0.62	0.77	0.63	0.82	0.64	4
5	0.63	0.58	0.64	0.61	0.66	0.62	0.70	0.63	0.73	0.70	0.80	0.73	5
6	0.63	0.49	0.64	0.51	0.66	0.56	0.71	0.57	0.73	0.64	0.74	0.67	6
7	0.60	0.54	0.61	0.54	0.65	0.55	0.69	0.56	0.71	0.58	0.77	0.61	7
8	0.71	0.52	0.72	0.54	0.73	0.54	0.75	0.55	0.80	0.64	0.86	0.67	8
9	0.63	0.71	0.63	0.73	0.64	0.74	0.64	0.75	0.64	0.76	0.64	0.80	9
10	0.59	0.53	0.60	0.55	0.62	0.63	0.63	0.64	0.64	0.68	0.67	0.71	10
11	0.59	0.26	0.60	0.28	0.62	0.29	0.66	0.30	0.69	0.45	0.70	0.49	11
12	0.57	0.72	0.57	0.74	0.58	0.76	0.59	0.77	0.59	0.80	0.60	0.84	12
13	0.56	0.66	0.57	0.68	0.57	0.68	0.58	0.69	0.58	0.71	0.59	0.74	13
14	0.53	0.50	0.54	0.51	0.55	0.56	0.56	0.57	0.57	0.58	0.58	0.59	14
15	0.52	0.63	0.54	0.65	0.54	0.74	0.55	0.76	0.55	0.76	0.55	0.77	15
16	0.57	0.36	0.58	0.38	0.60	0.44	0.61	0.46	0.70	0.49	0.71	0.53	16
17	0.59	0.23	0.60	0.25	0.61	0.34	0.61	0.36	0.62	0.37	0.62	0.38	17
18	0.54	0.48	0.55	0.50	0.55	0.50	0.57	0.51	0.57	0.51	0.58	0.53	18
19	0.51	0.51	0.53	0.51	0.53	0.51	0.53	0.52	0.54	0.52	0.55	0.52	19
20	0.47	0.58	0.49	0.60	0.49	0.60	0.50	0.60	0.50	0.64	0.51	0.70	20
21	0.38	0.46	0.38	0.48	0.39	0.55	0.40	0.56	0.40	0.57	0.40	0.57	21
22	0.57	0.51	0.58	0.53	0.60	0.54	0.62	0.55	0.63	0.56	0.63	0.57	22
23	0.47	0.55	0.47	0.58	0.48	0.62	0.48	0.63	0.49	0.64	0.49	0.65	23
24	0.33	0.38	0.34	0.40	0.46	0.41	0.46	0.42	0.47	0.43	0.47	0.47	24
25	0.53	0.58	0.53	0.60	0.54	0.61	0.54	0.61	0.54	0.62	0.54	0.62	25
26	0.31	0.61	0.31	0.61	0.31	0.62	0.31	0.62	0.31	0.63	0.33	0.64	26
27	0.37	0.51	0.38	0.53	0.38	0.53	0.40	0.53	0.40	0.53	0.42	0.54	27
28	0.42	0.59	0.44	0.61	0.50	0.61	0.51	0.64	0.56	0.64	0.57	0.65	28
29	0.41	0.54	0.42	0.56	0.43	0.63	0.44	0.64	0.44	0.64	0.44	0.65	29
30	0.29	0.49	0.30	0.51	0.32	0.57	0.32	0.58	0.33	0.58	0.33	0.61	30
31	0.62	0.52	0.63	0.54	0.63	0.55	0.63	0.56	0.66	0.58	0.67	0.61	31
32	0.38	0.42	0.39	0.44	0.39	0.45	0.40	0.46	0.54	0.47	0.55	0.48	32
33	0.41	0.37	0.42	0.39	0.42	0.42	0.42	0.44	0.47	0.54	0.48	0.59	33
34	0.40	0.46	0.41	0.47	0.42	0.47	0.42	0.48	0.42	0.49	0.43	0.53	34
35	0.30	0.58	0.30	0.60	0.30	0.60	0.30	0.61	0.30	0.61	0.31	0.62	35
36	0.38	0.30	0.39	0.32	0.39	0.39	0.39	0.42	0.40	0.49	0.41	0.54	36
37	0.24	0.56	0.24	0.57	0.25	0.57	0.25	0.57	0.25	0.58	0.27	0.59	37
38	0.28	0.63	0.28	0.65	0.29	0.66	0.29	0.66	0.30	0.67	0.30	0.67	38
39	0.41	0.34	0.42	0.36	0.47	0.43	0.48	0.45	0.53	0.46	0.54	0.47	39
40	0.25	0.50	0.25	0.52	0.26	0.54	0.26	0.55	0.35	0.62	0.37	0.65	40
41	0.23	0.41	0.23	0.43	0.23	0.45	0.23	0.46	0.28	0.48	0.30	0.52	41
42	0.21	0.32	0.21	0.34	0.25	0.34	0.25	0.34	0.38	0.40	0.40	0.44	42
43	0.23	0.28	0.23	0.30	0.30	0.30	0.30	0.32	0.31	0.35	0.33	0.35	43
44	0.20	0.54	0.20	0.57	0.23	0.60	0.25	0.61	0.25	0.62	0.25	0.65	44
45	0.22	0.51	0.22	0.53	0.23	0.57	0.23	0.57	0.24	0.58	0.24	0.58	45
46		0.27		0.29		0.30		0.31		0.42		0.46	46
47		0.40		0.42		0.47		0.48		0.49		0.53	47
48		0.51		0.53		0.53		0.53		0.54		0.55	48

Table 40. Item Difficulties (*p*-values)
Level AR, Forms S and T

Item No.	Fall, Form S		Fall, Form T		Item No.
	Voc	Comp	Voc	Comp	
1	0.89	0.89	0.93	0.93	1
2	0.92	0.93	0.89	0.65	2
3	0.86	0.93	0.88	0.91	3
4	0.86	0.91	0.91	0.86	4
5	0.87	0.92	0.82	0.85	5
6	0.81	0.87	0.81	0.93	6
7	0.73	0.77	0.84	0.72	7
8	0.74	0.90	0.80	0.78	8
9	0.92	0.74	0.82	0.86	9
10	0.64	0.64	0.77	0.77	10
11	0.83	0.82	0.86	0.79	11
12	0.83	0.80	0.77	0.79	12
13	0.78	0.89	0.79	0.74	13
14	0.53	0.81	0.79	0.91	14
15	0.75	0.86	0.79	0.68	15
16	0.88	0.81	0.75	0.89	16
17	0.78	0.63	0.76	0.35	17
18	0.68	0.77	0.77	0.87	18
19	0.85	0.66	0.75	0.86	19
20	0.79	0.84	0.77	0.90	20
21	0.62	0.63	0.58	0.81	21
22	0.51	0.61	0.69	0.76	22
23	0.54	0.58	0.66	0.51	23
24	0.76	0.74	0.61	0.64	24
25	0.65	0.84	0.61	0.88	25
26	0.61	0.44	0.58	0.58	26
27	0.66	0.79	0.57	0.63	27
28	0.61	0.47	0.71	0.60	28
29	0.38	0.71	0.37	0.69	29
30	0.47	0.68	0.62	0.52	30
31	0.49	0.58	0.61	0.71	31
32	0.60	0.69	0.57	0.69	32
33	0.56	0.52	0.59	0.69	33
34	0.65	0.82	0.52	0.47	34
35	0.52	0.85	0.38	0.61	35
36	0.44	0.70	0.47	0.63	36
37	0.59	0.54	0.51	0.61	37
38	0.57	0.75	0.50	0.52	38
39	0.49	0.76	0.46	0.54	39
40	0.47	0.63	0.44	0.48	40
41	0.52	0.67	0.36	0.54	41
42	0.38	0.57	0.56	0.52	42
43	0.41	0.35	0.50	0.74	43
44	0.34	0.39	0.55	0.52	44
45	0.24	0.63	0.34	0.67	45
46		0.52		0.43	46
47		0.64		0.41	47
48		0.39		0.63	48

GMRT®
Forms S and T

9-40364