



MCI Pre- and Post-Tests

Technical Guide

November 2009

Developed for:



Educators Publishing Service
www.epsbooks.com



MetaMetrics.
LINKING ASSESSMENT WITH INSTRUCTION

MetaMetrics | 1000 Park Forty Plaza Drive, Suite 120 Durham, North Carolina 27713 | www.Lexile.com

MetaMetrics®, the MetaMetrics logo and tagline, Lexile®, Lexile Framework®, Lexile Analyzer®, the Lexile logo, Quantile®, Quantile Framework® and the Quantile logo are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad. The trademarks and names of other companies and products mentioned herein are the property of their respective owners. Copyright © 2010 MetaMetrics, Inc. All rights reserved.

Table of Contents

	Page
List of Tables.....	ii
List of Figures	iii
Introduction	1
Background.....	1
Features of the <i>MCI Pre- and Post-Tests</i>	2
Purposes and Uses of the <i>MCI Pre- and Post-Tests</i>	3
Development Groups	4
A Word of Caution.....	4
The Lexile Framework for Reading.....	5
The Semantic Component	5
The Syntactic Component.....	6
Calibration of Text Difficulty	6
The Lexile Scale.....	7
Validity Evidence for the Lexile Framework for Reading.....	8
Forecasting Comprehension with the Lexile Framework	12
Description of <i>MCI Pre- and Post-Tests</i>	15
The <i>MCI Pre-Test</i>	15
The <i>MCI Post-Tests</i>	16
Interpreting and Using <i>MCI Pre- and Post-Test</i> Results	16
Development of the <i>MCI Pre- and Post-Tests</i>	23
Test Specifications	23
Passage Development.....	25
Item Development.....	29
Test Form Development.....	32
Scoring and Reporting	35
Reliability.....	37
Text Measure Error Associated with the Lexile Framework for Reading	37
Standard Error of Measurement.....	42
Validity	45
Content Validity	45
Construct Validity	45
References.....	47
Appendix	49

List of Tables

	Page
1. Results from linking studies conducted with The Lexile Framework for Reading	9
2. Correlations between theory-based calibrations produced by the Lexile equation and rank order of unit in basal readers	10
3. Correlations between theory-based calibrations produced by the Lexile equation and empirical difficulties	11
4. Comprehension rates for the same individual with materials of varying comprehension difficulty.....	13
5. Comprehension rates of different ability persons with the same material.....	14
6. Target Lexile measures and associated grade level norms for the pretest.....	16
7. Target Lexile measures and associated grade levels for the posttests	17
8. Lexile targets by reading level for each <i>MCI</i> test	23
9. Specifications for test passages.....	24
10. Specifications for test items	24
11. Passage Lexile distribution requirements for <i>MCI Pre-and Post-Tests</i>	25
12. Specification of native-Lexile items for the pretest and posttests	26
13. Specification of passages for passage-native Lexile items for the pretest and posttests ..	26
14. Technical information for the <i>MCI Pre-and Post-Test</i> forms	32
15. Standard errors for selected values of the length of the text	38
16. Analysis of 30 item ensembles providing an estimate of the theory misspecification error	40
17. Old method text readabilities, resampled SEMs, and new SEMs for selected books.....	42
18. Uncertainties for placement and baseline tests by Lexile range (25% - 75% correct)	43

List of Figures

	Page
1. Relationship between reader-text discrepancy and forecasted reading comprehension rate	13
2. Scatter plot between observed item difficulty and theoretical item difficulty	38
3a. Plot of observed ensemble means and theoretical calibrations (RMSE = 110L).	41
3b. Plot of simulated "true" ensemble means and theoretical calibrations (RMSE = 64L)	41

Introduction

Designed for struggling readers in grades 6 through 8 and above, the *MCI Pre- and Post-Tests* are an assessment series developed by EPS (Educators Publishing Service), a division of School Specialty, Inc., in conjunction with MetaMetrics, Inc. The *MCI Pre- and Post-Tests* were designed to be used with *MCI (Making Connections® Intervention)*, a direct, systematic, and interactive program developed primarily to teach literacy skills in the context of a strategic approach to reading. The *MCI Pre- and Post-Tests* add an assessment component to the instructional program that is designed to (1) identify the reading level of students and (2) monitor students' reading progress after they complete each level of the program.

The *MCI Pre- and Post-Tests* are designed for use by classroom teachers and reading specialists. The *MCI Pre- and Post-Tests* assess reading comprehension to determine a student's overall reading ability. The pre-test is used to determine a student's initial reading level as the student enters the program. The *MCI Post-Tests* for each instructional level assess student progress in reading as students move through the instructional levels. These assessments are designed to provide information to meet several goals:

- measure a student's reading ability upon entering the program,
- measure the growth of a student's reading ability, and
- target instruction based on measured student reading ability.

The *MCI Pre- and Post-Tests* ensure that struggling readers are monitored to determine their growth as they move through the *MCI* program. The comprehension components of *MCI* focus directly on three areas that have been found to improve reading comprehension for older struggling readers: explicit instruction in comprehension strategies, building vocabulary, and increasing fluency. In *MCI*, improvement in these three areas is monitored through informal formative assessments conducted throughout each unit as well as by formal formative assessments conducted at the end of each unit and at the middle and end of each level. The *MCI Pre- and Post-Tests* provide a way to determine how much growth a student has experienced in overall reading ability via the systematic focus on reading skills and strategies provided by the *MCI* program. The pre-test determines the student's reading ability prior to any instruction in the program. After completing a level of the instructional program, the student is given a post-test to measure reading growth.

Results from the *MCI Pre- and Post-Tests* are calculated using the Lexile® scale and The Lexile Framework® for Reading, a scientifically based scale of reading ability. The Lexile scale provides accurate feedback, helping to measure progress and forecast student development.

This technical guide provides users with a broad research foundation as well as descriptive information about the design and development of the *MCI Pre- and Post-Tests*. Such a base is essential when deciding how the *MCI Pre- and Post-Tests* should be used, and what kinds of inferences about readers are appropriate.

Background

In 2000, the National Reading Panel identified methods that consistently lead to success for most students. Their report recommended explicit, systematic teaching in these "five pillars": phonemic awareness, phonics, fluency, vocabulary, and comprehension. This report and the

Reading First Initiative established by No Child Left Behind led to emphasis on research-based instruction and a focus on the five pillars of reading instruction in the elementary grades.

On September 16, 2002, Dr. G. Reid Lyon of the National Institutes of Child Health and Human Development, a branch of the National Institutes of Health, spoke to a group of teachers and educators in Carroll County, Maryland. He noted that “37 percent of the nation’s fourth-graders read below basic level and the number climbs to 60 percent among minorities. About 75 percent of those who don’t learn to read by age 9 never learn” (Hare, 2002).

In 2004, *Reading Next: A Vision for Action and Research in Middle and High School Literacy*, a report to the Carnegie Corporation, found that almost 70 percent of students entering ninth grade read below grade level. The problem, according to *Reading Next*, is that these struggling readers do not comprehend what they read, for a variety of reasons. EPS developed the comprehension components of *MCI* to focus on the three areas *Reading Next* targeted for older readers: teaching comprehension strategies, building vocabulary, and increasing fluency.

MCI avoids putting struggling readers in situations where they fail because text is above their instructional level. The readability of student materials is targeted at a level three grades below the proficient levels for middle school students. The Lexile Framework for Reading was used to target the difficulty of the text to the appropriate levels.

Although *MCI* incorporates formative assessment throughout the instructional program to help instructors monitor skill growth and differentiate instruction, a measure of overall reading comprehension helps monitor how well students apply their reading skills to the general task of reading. EPS developed the *MCI Pre- and Post-Tests* to supplement their existing formative assessments and formal skill assessments with an overall measure of student reading ability. MetaMetrics, Inc. was invited to develop the *MCI Pre- and Post-Test* components, including the tests, items, scoring, and some ancillary materials.

Features of the *MCI Pre- and Post-Tests*

The *MCI Pre- and Post-Tests* are designed to measure how well readers read fiction and nonfiction materials. The assessments measure reading comprehension by focusing on skills readers use when reading fiction and nonfiction text. These skills include referring to details in the passage, making inferences, drawing conclusions, and making comparisons and generalizations. The assessments do not require prior knowledge of ideas outside of the passage or vocabulary taken out of context.

The *MCI Pre-Test* includes items at a range of difficulty appropriate for struggling readers in middle school or higher. If a student’s score on the pre-test falls within the Lexile range of one of the *MCI* levels—Aqua, Gold, or Crimson—that student will find this *MCI* level at his or her instructional level. For students whose Lexile score falls outside the range of a particular *MCI* level, teachers can adjust the instructional scaffolding. Either way, the pre-test can serve as a baseline score. Upon completion of an *MCI* level, the student is administered the appropriate *MCI Post-Test* to assess progress.

The *MCI Pre- and Post-Tests* are easy to use. After administering a pre-test at the beginning of the program, the instructor can monitor improvement in student reading comprehension through the post-tests administered at the end of each instructional level.

Several specific features of the program’s assessments are noteworthy.

- ◆ All of the tests in the *MCI Pre- and Post-Tests* can be easily administered and scored.
- ◆ The format of the tests supports quick administration in an un-timed, low-pressure atmosphere.
- ◆ No extensive or specialized preparation is needed to administer the program’s assessments, although proper interpretation and use of the results requires an understanding of the Lexile Framework for Reading.
- ◆ There are two versions of the *MCI Pre- and Post-Tests*. The booklet of reproducible tests supports rapid objective scoring and requires no computer software to administer or score. The online version features automatic scoring and reporting.
- ◆ Reading passages were developed to appeal to struggling readers in middle school and beyond. The passages contain fiction and nonfiction content, and they model authentic text that students will encounter in school and independently.
- ◆ The item formats used on the *MCI Pre- and Post-Tests* are the native-Lexile item type and the “embedded completion” item format. Both item types have been shown to measure the same core reading competency that is measured by norm-referenced, criterion-referenced, and individually administered reading tests (Stenner, Smith, Horiban, and Smith, 1987a).
- ◆ The *MCI Pre- and Post-Tests* are linked to the Lexile scale and, as such, the item calibrations used to convert a raw score (number correct) into the Lexile metric are provided by the Lexile Theory. The equation used to calibrate the program’s passages and test items is the same equation that is used to measure books/texts. Thus, readers and texts are placed on the same metric.
- ◆ More than a decade of research went into defining the rules for sampling text and writing passage native-Lexile and embedded-passage Lexile completion items. These rules were followed in developing the *MCI Pre- and Post-Tests* assessment items. A multi-stage review process was used to ensure conformance with the item writing specifications.

Purposes and Uses of the *MCI Pre- and Post-Tests*

The *MCI Pre- and Post-Tests* are designed to measure a reader’s ability to comprehend fiction and nonfiction texts of increasing difficulty. The tests can be used to measure where students are in the development of their reading ability and to determine how well the students are progressing toward meeting their reading goals.

One outcome of the *MCI Pre- and Post-Tests* is the location of the reader on the Lexile Map (Appendix A). Once a student’s reading ability is assessed, it is possible to forecast how well the reader will comprehend thousands of books and articles that have been measured in the Lexile metric. Readers and texts are similarly measured in the same Lexile metric, making it possible to directly compare a reader and text. When reader and text measures match, the Lexile Framework forecasts 75% comprehension. The operational definition of 75% comprehension is

that given 100 items from a text, the reader will be able to correctly answer 75 of them. When the text has a Lexile measure 250L higher than the reader measure, the Lexile Framework forecasts 50% comprehension. When the reader measure exceeds the text measure by 250L, the forecasted comprehension is 90%.

Each pre-test and post-test is comprised of unique items; there are no overlapping items. Each test is centered on a specific range of the Lexile scale. Students are generally well measured when they are administered a test that is centered near their true reading ability. When students take poorly targeted tests, there is considerable uncertainty about their location on the Lexile Map.

Development Groups

EPS provided the vision for the tests and guided the development of the content specifications of the *MCI Pre- and Post-Tests*. MetaMetrics, Inc. managed the overall development of the tests.

EPS approved final passage selection, approved final item sets, implemented the scoring and reporting of the tests, and implemented the production of *MCI Pre- and Post-Tests*.

MetaMetrics, Inc. designed the specifications for the tests, developed the passages and test items, coordinated the test development, and designed the scoring and reporting algorithms.

A Word of Caution

Instructional decisions are best made when using multiple sources of evidence about a student. Other sources include standardized test data, instructional group placement, lists of books read, courses completed, passing status, attendance, and, most importantly, teacher judgment. *One measure of student performance, taken on one day, is never sufficient to make high-stakes, student-level decisions such as retention.*

The Lexile Framework for Reading

A reader's comprehension of text is dependent on many factors—the purpose for reading, the ability of the reader, and the text that is being read. The reader can be asked to read a text for many purposes including entertainment (literary experience), to gain information, or to perform a task. Each reader brings to the reading experience a variety of important factors: reading ability, prior knowledge, interest level, and developmental readiness. For any text, there are three factors associated with the readability of the text: difficulty, support, and quality. All of these reader and text factors are important considerations when evaluating the appropriateness of a text for a reader. The Lexile Framework focuses primarily on two: reader ability and text difficulty.

Within the Lexile Framework, text difficulty is determined by examining the characteristics of word frequency and sentence length. All symbol systems share two features: a semantic component and a syntactic component. In language, the semantic units are words. Words are organized according to rules of syntax into thought units and sentences (Carver, 1974). In all cases, the semantic units vary in familiarity and the syntactic structures vary in complexity. The comprehensibility or difficulty of a message is dominated by the familiarity of the semantic units and by the complexity of the syntactic structures used in constructing the message. The Lexile Framework for Reading utilizes these two dominant features of language in measuring text difficulty. Text measures typically range from 0 Lexiles (L) to 1800 Lexiles (L), but can range from below 0L for beginning texts to above 1800L for more advanced texts. Within any one classroom there will be a range of reading materials to reflect the range of student reading abilities and interest in different topics and types of text.

The Semantic Component

Most operationalizations of semantic complexity are proxies for the probability that an individual will encounter a word in a familiar context and thus be able to infer its meaning (Bormuth, 1966). This is the basis of exposure theory, which explains the way receptive or hearing vocabulary develops (Miller and Gildea, 1987; Stenner, Smith, and Burdick, 1983). Klare (1963) hypothesized that the semantic component varied along a familiarity-to-rarity continuum. This concept was further developed by Carroll, Davies, and Richman (1971), whose word-frequency study examined the reoccurrence of words in a five-million-word corpus of running text. Knowing the frequency of words as they are used in written and oral communication provided the best means of inferring the likelihood that a word would be encountered by a reader and thus become a part of that individual's receptive vocabulary.

Variables such as the average number of letters or syllables per word have been observed to be proxies for word frequency. There is a high negative correlation between the length of words and the frequency of word usage. Polysyllabic words are used less frequently than monosyllabic words, making word length a good proxy for the likelihood that an individual will be exposed to a word.

In a study examining receptive vocabulary, Stenner, Smith, and Burdick (1983) analyzed more than 50 semantic variables in order to identify those elements that contributed to the difficulty of the 350 vocabulary items on Forms L and M of the *Peabody Picture Vocabulary Test—Revised* (Dunn and Dunn, 1981). Variables included part of speech, number of letters, number of syllables, the modal grade at which the word appeared in school materials, content classification

of the word, the frequency of the word from two different word counts, and various algebraic transformations of these measures. The word frequency measure used was the raw count of how often a given word appeared in a corpus of 5,088,721 words sampled from a broad range of school materials (Carroll, Davies, and Richman, 1971). A “word family” included: (1) the stimulus word; (2) all plurals (adding “-s” or changing “-y” to “-ies”); (3) adverbial forms; (4) comparatives and superlatives; (5) verb forms (“-s,” “-d,” “-ed,” and “-ing”); (6) past participles; and (7) adjective forms. Correlations were computed between algebraic transformations of these means and the rank order of the test items. Since the items were ordered according to increasing difficulty, the rank order was used as the observed item difficulty. The mean log word frequency provided the highest correlation with item rank order ($r = -0.779$) for the items on the combined form.

The Lexile Framework uses word frequency as the semantic component of the text difficulty measure. Currently, a 600-million-word corpus is employed when examining the semantic component of text. This corpus was assembled from the thousands of texts publishers have measured.

The Syntactic Component

Klare (1963) provided a possible interpretation for how sentence length works in predicting passage difficulty. He speculated that the syntactic component varied with the load placed on short-term memory. Crain and Shankweiler (1988), Shankweiler and Crain (1986), and Liberman, Mann, Shankweiler, and Westelman (1982) have also supported this explanation. The work of these individuals has provided evidence that sentence length is a good proxy for the demand that structural complexity places upon verbal short-term memory.

While sentence length has been shown to be a powerful proxy for the syntactic complexity of a passage, an important caveat is that sentence length is not the underlying causal influence (Chall, 1988). Researchers sometimes incorrectly assume that manipulation of sentence length will have a predictable effect on passage difficulty. Davidson and Kantor (1982), for example, illustrated rather clearly that sentence length can be reduced and difficulty increased and vice versa.

Based on previous research, it was decided to use sentence length as a proxy for the syntactic component of reading difficulty in the Lexile Framework.

Calibration of Text Difficulty

A research study on semantic units conducted by Stenner, Smith, and Burdick (1983) was extended to examine the relationship of word frequency and sentence length to reading comprehension. In 1987(a), Stenner, Smith, Horabin, and Smith performed exploratory regression analyses to test the explanatory power of these variables. This analysis involved calculating the mean word frequency and the log of the mean sentence length for each of the 66 reading comprehension passages on the *Peabody Individual Achievement Test*. The observed difficulty of each passage was the mean difficulty of the items associated with the passage (provided by the publisher) converted to the logit scale. A regression analysis based on the word-frequency and sentence-length measures produced a regression equation that explained most of the variance found in the set of reading comprehension tasks. The resulting correlation between the observed logit difficulties and the theoretical calibrations was 0.97 after correction for range restriction and measurement error. The regression equation was further refined based

on its use in predicting the observed difficulty of the reading comprehension passages on eight other standardized tests. The resulting correlation between the observed logit difficulties and the theoretical calibrations when the nine tests were combined into one was 0.93 after correction for range restriction and measurement error.

Once a regression equation was established linking the syntactic and semantic features of text to the difficulty of text, and then the equation was used to calibrate test items and text.

The Lexile Scale

In developing the Lexile scale, the Rasch item response theory model (Wright and Stone, 1979) was used to estimate the difficulties of items and the abilities of persons on the logit scale.

The calibrations of the items from the Rasch model are objective in the sense that the relative difficulties of the items will remain the same across different samples of persons (specific objectivity). When two items are administered to the same person it can be determined which item is harder and which one is easier. This ordering is likely to hold when the same two items are administered to a second person. If two different items are administered to the second person, there is no way to know which set of items is harder and which set is easier. The problem is that the location of the scale is not known. General objectivity requires that scores obtained from different test administrations be tied to a common zero—absolute location must be sample independent (Stenner, 1990). To achieve general objectivity, the theoretical logit difficulties must be transformed to a scale where the ambiguity regarding the location of zero is resolved.

The first step in developing a scale with a fixed zero was to identify two anchor points for the scale. The following criteria were used to select the two anchor points: they should be intuitive, easily reproduced, and widely recognized. For example, with most thermometers the anchor points are the freezing and boiling points of water. For the Lexile scale, the anchor points are text from seven basal primers for the low end and text from *The Electronic Encyclopedia* (Grolier, Inc., 1986) for the high end. These points generally correspond to the middle of first grade text and the midpoint of workplace text.

The next step was to determine the unit size for the scale. For the Celsius thermometer, the unit size (a degree) is 1/100th of the difference between freezing (0 degrees) and boiling (100 degrees) water. For the Lexile scale the unit size was defined as 1/1000th of the difference between the mean difficulty of the primer material and the mean difficulty of the encyclopedia samples. Therefore, a Lexile by definition equals 1/1000th of the difference between the comprehensibility of the primers and the comprehensibility of the encyclopedia.

The third step was to assign a value to the lower anchor point. The low-end anchor on the Lexile scale was assigned a value of 200.

Finally, a linear equation of the form

$$[(\text{Logit} + \text{constant}) \times \text{CF}] + 200 = \text{Lexile text measure} \quad (\text{Equation 1})$$

was developed to convert logit difficulties to Lexile calibrations. The values of the conversion factor (CF) and the constant were determined by substituting in the anchor points and then solving the system of equations.

Validity Evidence for the Lexile Framework for Reading

The 1999 *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education) state that “validity refers to the degree to which evidence and theory support the interpretations of test scores entailed in the uses of tests” (p. 9). In applying this definition to the Lexile Framework for Reading, the question that should be asked is: What evidence supports the use of the Lexile Framework to describe text difficulty and reader ability? Because the Lexile Framework addresses reading comprehension, an important aspect of validity evidence that should be brought to bear is evidence showing that the construct being addressed is indeed reading comprehension. This type of validity evidence has traditionally been called construct validity. One source of construct validity evidence for The Lexile Framework for Reading can be evaluated by examining how well Lexile measures relate to other measures of reading comprehension.

Lexile Framework Linked to other Measures of Reading Comprehension. The Lexile Framework for Reading has been linked to several standardized tests of reading comprehension. When assessment scales are linked, a common frame of reference can be used to interpret the test results. This frame of reference can be “used to convey additional normative information, test-content information, and information that is jointly normative and content-based. For many test uses, ... [this frame of reference] conveys information that is more crucial than the information conveyed by the primary score scale” (Petersen, Kolen, and Hoover, 1989, p. 222). Linking the Lexile Framework to other measures of reading comprehension produces a common frame of reference: Lexile measure.

Table 1 presents the results from linking studies conducted with The Lexile Framework for Reading. For each of the tests listed, student reading comprehension scores can also be reported as Lexile measures. This dual reporting provides a rich, criterion-related frame of reference for interpreting the standardized test scores. When a student takes one of the standardized tests, in addition to receiving his norm-referenced test results, he can receive a reading list that is targeted to his specific reading level.

Table 1. Results from linking studies conducted with The Lexile Framework for Reading.

Standardized Test	Grades in Study	N	Correlation Between Test Score and Lexile measure
Stanford Achievement Tests (Ninth Edition)	4, 6, 8, 10	1,167	0.92
Stanford Diagnostic Reading Test (Version 4)	4, 6, 8, 10	1, 169	0.91
North Carolina End-of-Grade Tests (Reading Comprehension)	3, 4, 5, 8	956	0.90
TerraNova (CTBS/5)	2, 4, 6, 8	2,713	0.92
Texas Assessment of Academic Skills (TAAS)	3–8	3,623	0.73 to 0.78*
Metropolitan Achievement Test (Eighth Edition)	2, 4, 6, 8, and 10	2,382	0.93
Gates-MacGinitie Reading Test (Version 4)	2, 4, 6, 8, and 10	4,644	0.92
Utah Core Assessments	3–6	1,551	0.73
Texas Assessment of Knowledge and Skills (TAKS)	3, 5, and 8	1,960	0.60 to 0.73*
The Iowa Tests (Iowa Tests of Basic Skills and Iowa Tests of Educational Development)	3, 5, 7, 9, and 11	4,666	0.88
Stanford Achievement Test (Tenth Edition)	2, 4, 6, 8, and 10	3,064	0.93
Oregon Reading/Literature Knowledge and Skills Test	3, 5, 8, and 10	3,180	0.89
Mississippi Curriculum Test (MCT)	2, 4, 6, and 8	7,045	0.90
Georgia Criterion Referenced Competency Test (CRCT and GHSGT)	1 – 8, and 11	16,363	0.72 to 0.88*
Wyoming Performance Assessment for Wyoming Students (PAWS)	3, 5, 7, and 11	3,871	0.91
Arizona Instrument to Measure Progress (AIMS)	3, 5, 7, and 10	7,735	0.89
South Carolina Palmetto Achievement Challenge Tests (PACT)	3 – 8	15,559	0.87 to 0.88*

Notes: Results are based on final samples used with each linking study.

*TAAS, TAKS, CRCT/GHSGT, and PACT were not vertically equated; separate linking equations were derived for each grade.

Lexile Framework and the Difficulty of Basal Readers. The Lexile measures are organized in a sequential manner, so a lower Lexile measure for a text means that the text is less difficult than texts with higher Lexile measures. Validity evidence for the internal structure (the sequential structure) of the Lexile Framework was obtained through a study that examined the relationship of basal reader sequencing to the structure of the Lexile Framework. In a study conducted by Stenner, Smith, Horabin, and Smith (1987b), Lexile calibrations were obtained for units in 11 basal series. It was presumed that each basal series was sequenced by difficulty. So, for example, the latter portion of a third-grade reader is presumably more difficult than the first portion of the same book. Likewise, a fourth-grade reader is presumed to be more difficult than a third-grade reader is. Observed difficulties for each unit in a basal series were estimated by the rank order of the unit in the series. Thus, the first unit in the first book of the first-grade reader was assigned a rank order of one and the last unit of the eighth-grade reader was assigned the highest rank order number.

Correlations were computed between the rank order and the Lexile calibration of each unit in each series. After correction for range restriction and measurement error, the average disattenuated correlation between the Lexile calibration of text comprehensibility and the rank order of the basal units was 0.995 (see *Table 2*).

Table 2. Correlations between theory-based calibrations produced by the Lexile equation and rank order of unit in basal readers.

Basal Series	Number of Units	r_{OT}	R_{OT}	R'_{OT}
Ginn Rainbow Series (1985)	53	.93	.98	1.00
HBJ Eagle Series (1983)	70	.93	.98	1.00
Scott Foresman Focus Series (1985)	92	.84	.99	1.00
Riverside Reading Series (1986)	67	.87	.97	1.00
Houghton-Mifflin Reading Series (1983)	33	.88	.96	.99
Economy Reading Series (1986)	67	.86	.96	.99
Scott Foresman American Tradition (1987)	88	.85	.97	.99
HBJ Odyssey Series (1986)	38	.79	.97	.99
Holt Basic Reading Series (1986)	54	.87	.96	.98
Houghton-Mifflin Reading Series (1986)	46	.81	.95	.98
Open Court Headway Program (1985)	52	.54	.94	.97
Total/Means	660	.839	.965	.995

r_{OT} = raw correlation between observed difficulties (*O*) and theory-based calibrations (*T*).

R_{OT} = correlation between observed difficulties (*O*) and theory-based calibrations (*T*) corrected for range restriction.

R'_{OT} = correlation between observed difficulties (*O*) and theory-based calibrations (*T*) corrected for range restriction and measurement error.

*Mean correlations are the weighted averages of the respective correlations.

Based on the consistency of the results in *Table 2*, the Lexile Theory was able to account for the unit rank ordering of the 11 basal series even with numerous differences in the series—prose selections, developmental range addressed, types of prose introduced (i.e., narrative versus expository), and purported skills and objectives emphasized.

Lexile Framework and the Difficulty of Reading Test Items. Additional construct validity evidence was obtained by exploring the relationship between Lexile calibrations of item difficulties and actual item difficulties of reading comprehension tests. In a study conducted by Stenner, Smith, Horabin, and Smith (1987a), 1,780 reading comprehension test items appearing on nine nationally-normed tests were analyzed. The study correlated empirical item difficulties provided by the publisher with the Lexile calibrations specified by the computer analysis of the text of each item. The empirical difficulties were obtained in one of three ways. Three of the tests included observed logit difficulties from either a Rasch or three-parameter analysis (e.g., NAEP). For four of the tests, logit difficulties were estimated from item p-values and raw score means and standard deviations (Poznansky, 1990; Stenner, Wright, and Linacre, 1994). Two of the tests provided no item parameters, but in each case items were ordered on the test in terms of difficulty (e.g., PIAT). For these two tests, the empirical difficulties were approximated by the difficulty rank order of the items. In those cases where multiple questions were asked about a single passage, empirical item difficulties were averaged to yield a single observed difficulty for the passage.

Once theory-specified calibrations and empirical item difficulties were computed, the two arrays were correlated and plotted separately for each test. The plots were checked for unusual residual distributions and curvature, and it was discovered that the equation did not fit poetry items or non-continuous prose items (e.g., recipes, menus, or shopping lists). This indicated that the universe to which the Lexile equation could be generalized was limited to continuous prose. The poetry and non-continuous prose items were removed and correlations were recalculated. *Table 3* contains the results of this analysis.

Table 3. Correlations between theory-based calibrations produced by the Lexile equation and empirical item difficulties.

Test	Number of Questions	Number of Passages	Mean	SD	Range	Min	Max	r_{OT}	R_{OT}	R'_{OT}
SRA	235	46	644	353	1303	33	1336	.95	.97	1.00
CAT-E	418	74	789	258	1339	212	1551	.91	.95	.98
Lexile	262	262	771	463	1910	-304	1606	.93	.95	.97
PIAT	66	66	939	451	1515	242	1757	.93	.94	.97
CAT-C	253	43	744	238	810	314	1124	.83	.93	.96
CTBS	246	50	703	271	1133	173	1306	.74	.92	.95
NAEP	189	70	833	263	1162	169	1331	.65	.92	.94
Battery	26	26	491	560	2186	-702	1484	.88	.84	.87
Mastery	85	85	593	488	2135	-586	1549	.74	.75	.77
Total/ Mean	1780	722	767	343	1441	50	1491	.84	.91	.93

r_{OT} = raw correlation between observed difficulties (O) and theory-based calibrations (T).

R_{OT} = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction.

R'_{OT} = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction and measurement error.

*Means are computed on Fisher Z transformed correlations.

The last three columns in *Table 3* show the raw correlation between observed (O) item difficulties and theoretical (T) item calibrations, with the correlations corrected for restriction in range and measurement error. The Fisher Z mean of the raw correlations (r_{OT}) is 0.84. When

corrections are made for range restriction and measurement error, the Fisher Z mean disattenuated correlation between theory-based calibrations and empirical difficulties in an unrestricted group of reading comprehension items (R'_{OT}) is 0.93.

These results suggest that most attempts to measure reading comprehension, no matter what the item form, type of skill or objectives assessed, or response requirement used, measure a common comprehension factor specified by the Lexile Theory.

Forecasting Comprehension with the Lexile Framework

A reader with a measure of 600L who is given a text measured at 600L is expected to have a 75-percent comprehension rate. This 75-percent comprehension rate is the basis for selecting text that is targeted to a reader's reading ability, but what exactly does it mean? And what would the comprehension rate be if this same reader were given a text measured at 350L or one at 850L?

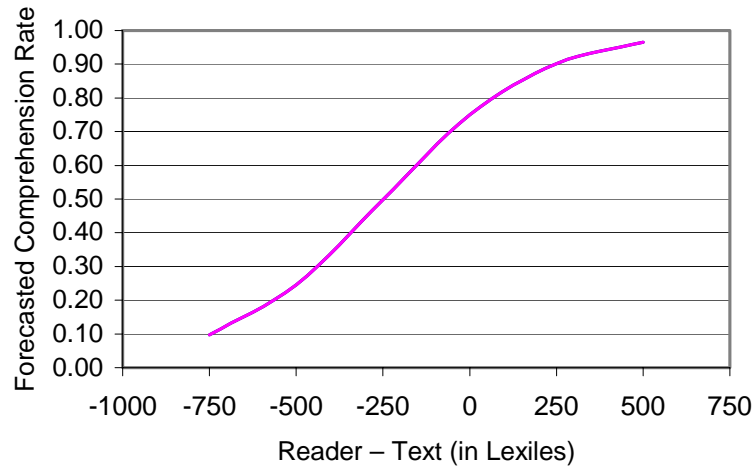
The 75-percent comprehension rate for a reader-text pairing can be given an operational meaning by imagining the text to be carved into item-sized slices of approximately 125-140 words with a question embedded in each slice. A reader who answers three-fourths of the questions correctly has a 75-percent comprehension rate.

Suppose instead that the text and reader measures are not the same. It is the difference in Lexiles between reader and text that governs comprehension. If the text measure is less than the reader measure, the comprehension rate will exceed 75 percent. If not, it will be less. The question is "By how much?" What is the expected comprehension rate when a 600L reader reads a 350L text?

If all the item-sized slices in the 350L text had the same calibration, the 250L difference between the 600L reader and the 350L text could be determined using the Rasch model equation (Equation 1 on page 7). This equation describes the relationship between the measure of a student's level of reading comprehension and the calibration of the items. Unfortunately, comprehension rates calculated by this procedure would be biased because the calibrations of the slices in ordinary prose are not all the same. The average difficulty level of the slices and their variability both affect the comprehension rate.

Figure 1 shows the general relationship between reader-text discrepancy and forecasted comprehension rate. When the reader measure and the text measure are the same (difference of 0L on the x-axis) then the forecasted comprehension rate is 75%. In the example in the preceding paragraph, the difference between the reader measure of 600L and the text measure of 350L is 250L. Referring to *Figure 1* and using +250L (reader minus text), the forecasted comprehension rate for this reader-text combination would be 90%.

Figure 1. Relationship between reader-text discrepancy and forecasted reading comprehension rate.



Tables 4 and 5 show comprehension rates calculated for various combinations of reader measures and text measures.

Table 4. Comprehension rates for the same individual with materials of varying comprehension difficulty.

Person Measure	Text Calibration	Sample Titles	Forecast Comprehension
1000L	500L	<i>Tornado</i> (Byars)	96%
1000L	750L	<i>The Martian Chronicles</i> (Bradbury)	90%
1000L	1000L	<i>Reader's Digest</i>	75%
1000L	1250L	<i>The Call of the Wild</i> (London)	50%
1000L	1500L	<i>On the Equality Among Mankind</i> (Rousseau)	25%

Table 5. Comprehension rates of different ability persons with the same material.

Person Measure	Calibration for <i>Sports Illustrated</i>	Forecast Comprehension Rate
500L	1000L	25%
750L	1000L	50%
1000L	1000L	75%
1250L	1000L	90%
1500L	1000L	96%

The subjective experience of 50%, 75%, and 90% comprehension as reported by readers varies greatly. A 1000L reader reading 1000L text (75% comprehension) reports confidence and competence. Teachers listening to such a reader report that the reader can sustain the meaning thread of the text and can read with motivation and appropriate emotion and emphasis. In short, such readers sound like they comprehend what they are reading. A 1000L reader reading 1250L text (50% comprehension) encounters so much unfamiliar vocabulary and difficult syntactic structures that the meaning thread is frequently lost. Such readers report frustration and seldom choose to read independently at this level of comprehension difficulty. Finally, a 1000L reader reading 750L text (90% comprehension) reports total control of the text, reads with speed, and experiences automaticity during the reading process.

The primary utility of the Lexile Framework is its ability to forecast what happens when readers confront text. With every application by teacher, student, librarian, or parent there is a test of the Framework's accuracy. The Framework makes a point prediction every time a text is chosen for a reader. Anecdotal evidence suggests that the Lexile Framework predicts as intended. That is not to say that there is an absence of error in forecasted comprehension. There is error in text measures, reader measures, and their difference modeled as forecasted comprehension. However, the error is sufficiently small that the judgments about readers, texts, and comprehension rates are useful.

Description of the *MCI Pre- and Post-Tests*

The *MCI Pre- and Post-Tests* are designed to measure reading comprehension and monitor student growth as students move through the *MCI* instructional program. The test results indicate how much information the student understands while reading. Because student reading comprehension is assessed progressively as students move through the program levels, reading comprehension improvement can be identified, and the scaffolded support can be adjusted appropriately.

The *MCI Pre- and Post-Tests* include one pre-test and three post-tests – one post-test for each of the three *MCI* levels. Before beginning the *MCI* program, the student is administered a pre-test, which can aid in placing the student in the appropriate *MCI* level and determining the amount of scaffolded support the student needs. The pre-test provides an estimate of the student's reading ability prior to any instruction in the assigned level, so it can serve as a starting point for monitoring reading improvement. After the student has completed an instructional level (Aqua, Gold, or Crimson), a post-test is administered. The post-tests provide information about how much the student's reading ability has grown over the course of the instructional period. The pre-test and the post-tests are each 40 items in length.

The *MCI Pre-Test*

The *MCI Pre-Test* identifies the student's initial reading level and also helps to determine which level of *MCI* is appropriate for the student. Appropriate placement helps to maximize the effectiveness of reading instruction because the student will be reading instructional materials targeted to his/her reading level. While *MCI* is designed for struggling readers, even students who score above a specified Lexile measure on the pre-test can benefit from the skill and strategy instruction in *MCI*. They will require the lowest level of scaffolded support. Additionally, students who score well below the targeted reading range of the program may benefit from the additional decoding and vocabulary support provided by the *MCI Word Study* component.

Many students using the *MCI* program are reading up to three grade levels below their current grade level, so the *MCI Pre-Test* was designed to target students at this level. Because Lexile measures are not bounded by grade level, it is possible to interpret difficulty in terms of various grade level norms. The normative information for Lexiles was used to determine the range for the pre-tests and post-tests and is shown in *Tables 6 and 7*.

The easiest items on the test were targeted to the 5th percentile reader in 6th grade, which is approximately the same level as the 30th percentile reader in 3rd grade. The mean item difficulty was targeted to the 50th percentile of the 6th grade, which is approximately the same level as the 75th percentile of 4th grade. Because students in 7th grade, 8th grade, or higher may take the *MCI Pre-Test*, the highest level item was targeted to approximately the 90th percentile of 8th grade students. Including these relatively high-level items also enables the pre-test to screen for students whose reading ability may be too high to benefit from instruction in *MCI*.

Table 6. Target Lexile measures and associated grade level norms for the pre-test.

	Lexile Measure	Approximate Grade Level Norm
Minimum Lexile Target	450L	5 th Percentile, Grade 6 30 nd Percentile, Grade 3
Mean Lexile Target	870L	50 th Percentile, Grade 6 75 th percentile, Grade 4
Maximum Lexile Target	1290L	90 th Percentile, Grade 8

The MCI Post-Tests

The *MCI Post-Tests* provide a Lexile measure for a student's reading ability after the student has completed a level of *MCI*. The post-tests are designed to target the ability of the readers placed into the specified *MCI* levels, and provide a more accurate estimate of the reader's end-of-level ability than if the test were not well-targeted to the reader's ability level. The specifications for the post-tests called for items at a reading difficulty approximately three grade levels below the actual grade level of the students in the program. The targeting information for the post-tests is shown in *Table 7*.

Interpreting and Using MCI Pre- and Post-Test Results

The Lexile Framework for Reading provides teachers with tools to help them link assessment results with subsequent instruction. Assessments such as the *MCI Pre- and Post-Tests* that are linked to the Lexile scale provide tools for monitoring the progress of students as they move through *MCI*.

The Lexile Framework for Reading is also a tool that helps determine the reading level of written material—from a book, to a test item, to a magazine article, to a textbook. After test results are converted into Lexile measures, readers can be matched with materials on their own level.

A Lexile measure is the specific number assigned to any text. A computer program called the Lexile Analyzer[®] computes the Lexile measure for a text. The Analyzer carefully examines the complete text to measure such characteristics as sentence length and word frequency—characteristics that are highly related to overall reading comprehension. The Analyzer then reports a Lexile measure for the text. More than 100,000 books, 80 million periodical articles, and many newspapers have been given Lexile measures using this tool. Noting the Lexile measure of a text can assist in choosing reading materials that present an appropriate level of challenge for a reader.

Table 7. Target Lexile measures and associated grade levels for the post-tests.

MCI Connections Intervention Level	Target Level	Lexile Target	Approximate Grade Level Norm
Aqua Post-Test	Minimum	450L	5 th Percentile, Grade 6 30 nd Percentile, Grade 3
	Mean	750L	30 th Percentile, Grade 6
	Maximum	1050L	75 th Percentile, Grade 6 98 nd Percentile, Grade 3
Gold Post-Test	Minimum	550L	5 th Percentile, Grade 7 30 th percentile, Grade 4
	Mean	800L	30 th Percentile, Grade 7
	Maximum	1150L	80 th Percentile, Grade 7 98 th percentile, Grade 4
Crimson Post-Test	Minimum	600L	5 th Percentile, Grade 8 30 th percentile, Grade 5
	Mean	880L	30 th Percentile, Grade 8
	Maximum	1200L	80 th Percentile, Grade 7 98 th percentile, Grade 4

A Lexile measure can also be used to identify the reading ability of a particular reader. Tests that are linked to the Lexile Framework or tests such as *MCI Pre- and Post-Tests* that are specifically developed to report Lexile scores can provide a Lexile measure for a reader. By using the Lexile measure for both reader and text as a tool to help target reading at the optimal, 75-percent comprehension range, reading development can be maximized.

The Lexile scale. Results on all *MCI Pre- and Post-Tests* are reported as a Lexile measure. The Lexile measure places the reader on the Lexile scale. There are many reasons to use scale scores rather than raw scores to report test results. Scale scores overcome the disadvantage of many other types of scores (e.g., percentiles, raw scores, or grade equivalents), in that equal differences between scale score points represent equal differences in ability. Each question on a test has a unique level of difficulty; therefore, answering 23 questions correctly on one form of a test may require a slightly different level of ability than answering 23 items correctly on another form of the test. In contrast, receiving a scale score (Lexile measure) of 875L on one form of a test represents a similar level of reading ability as receiving a scale score (Lexile measure) of 875L on another form of the test.

Individual *MCI Pre- and Post-Test* reading scores are calculated by first summing the number of correct responses (omitted items and multiple responses are counted as incorrect). The number correct is then converted to a scaled Lexile measure. The typical range of the Lexile scale is from 200 to 1700 Lexiles, although actual Lexile measures can range from below zero to above

2000 Lexiles. Reader Lexile measures are reported in 5-unit intervals. For *MCI* student reporting purposes, the lowest Lexile measure reported is BR for scores at or below 0L. This is a typical convention for reporting Lexile reader measures.

The Lexile Framework for Reading provides teachers with tools to help them link assessment results with subsequent instruction. The *MCI Pre- and Post-Tests* have been developed using the Lexile Framework for Reading, and all results are reported on the Lexile scale. These test results provide tools for monitoring the progress of students at any time during their course of instruction.

A Lexile reader measure can be used to forecast how well a student will comprehend a text with a specific Lexile measure. When the reader and text have the same Lexile measure, the reader is “targeted.” This means, for example, that a student whose reading ability has been measured at 500L is expected to read with 75-percent comprehension a book that is also measured at 500L. A targeted reader reports confidence, competence, and control over the text. When a text measure is 250L above the reader’s measure, comprehension is predicted to drop to 50 percent and the reader will likely experience frustration and inadequacy. Conversely, when a text measure is 250L below the reader’s measure, comprehension is predicted to go up to 90% and the reader is expected to experience control and fluency. When reading a book within his or her Lexile range (50L above his or her Lexile measure to 100L below), the reader is forecasted to comprehend enough of the text to make sense of it, while still being challenged enough to maintain interest and learning. Matching readers to appropriate text is a powerful use of the Lexile measure.

The Lexile Framework reporting scale is not bounded by grade level and Lexile measures are not reported as grade equivalents, although typical Lexile measure ranges have been identified for students in specific grades. Students within any given grade will have a wide range of Lexile measures, but a struggling reader is not stigmatized by a grade-based score below his or her actual grade level. Because the Lexile Framework scale is not bounded by grade level, it allows for nonjudgmental targeting of students with text at the appropriate difficulty level. See the Lexile Framework Map for fiction and nonfiction titles, leveled reading samples, and approximate grade ranges (Appendix A).

Using the Lexile Framework to Select Books. Teachers, parents, and students can use the tools provided by the Lexile Framework to plan instruction. When teachers provide parents and students with lists of titles that match the students’ Lexile measures, they can then work together to choose appropriate titles that also match the students’ interests and background knowledge. *The Lexile Framework does not prescribe a reading program, but it gives educators more control over the variables involved when they design reading instruction.* The Lexile Framework yields multiple opportunities for use in a variety of instructional activities. After becoming familiar with the Lexile Framework, teachers are likely to think of a variety of additional creative ways to use this tool to match students with books that students find challenging, but not frustrating.

The Find a Book with Lexiles website (www.lexile.com/findabook) helps readers select books in their chosen interest area that are matched to the reader’s Lexile range. The Find a Book website also connects readers to the WorldCat library database so they can find the nearest public library with the selected book in its catalog.

Remember, there are many factors that affect the relationship between a reader and a book. These factors include text content, age of the reader, interests of the reader, suitability of the

text, and text difficulty. The Lexile measure of a text, a measure of text difficulty, is a good starting point in the selection process, but other factors also must be considered. The Lexile measure should never be the only piece of information used when selecting a text for a reader.

Help Students Set Appropriate Learning Goals. Students' Lexile measures can be used to identify reading materials that they are likely to comprehend with 75% accuracy. Students can set goals of improving their reading comprehension and plan clear strategies for reaching those goals using literature from the appropriate Lexile ranges. Progress assessments throughout the course can help to monitor students' progress toward their goals.

Monitor Reading Program Goals. As a student's Lexile measure increases, his reading comprehension ability increases and the set of reading materials he can comprehend at 75% accuracy changes. Schools often write grant applications in which they are required to state how they will monitor progress of the intervention funded by the grant. Schools that receive funds targeted to assist students improve their reading skills can use the Lexile Framework for evaluation purposes. Schools can use student-level and school-level Lexile information to monitor and evaluate interventions designed to improve reading skills.

Measurable goals can be clearly stated in terms of Lexile measures. Examples of measurable goals and clearly related strategies for reading intervention programs might include:

Goal: At least half of the students will improve reading comprehension abilities by 100L after one year of use of an intervention.

Goal: Students' attitudes about reading will improve after reading 10 books at their 75% comprehension level.

These examples of goals emphasize the fact that the Lexile Framework is not an intervention, but a tool to help educators plan instruction and measure the success of the reading program.

Communicate With Parents Meaningfully to Include Them in the Educational Process. Teachers can make statements to parents such as, "Your child will be able to read with at least 75 comprehension these kinds of materials which are at the next grade level." Or, "Your child will need to be able to increase his/her Lexile measure by 400-500 Lexiles in the next few years to be prepared for college reading demands. Here is a list of appropriate titles your child can choose from for reading this summer."

Improve Students' Reading Fluency. Educational researchers have found that students who spend a minimum of three hours a week reading at their own level for their own purposes develop reading fluency that leads to improved mastery. Not surprisingly, researchers have also found that students who read age-appropriate materials with a high level of comprehension also learn to enjoy reading.

Teach Learning Strategies by Controlling Comprehension Match. The Lexile Framework permits the teacher to target readers with challenging text and to systematically off-target students when the teacher wants fluency and automaticity (i.e., reader measure is well above text measure) or wants to teach strategies for attacking "hard" text (i.e., reader measure is well below text measure). For example, metacognitive ability has been well-documented to play an important role in reading comprehension performance. Once teachers know the kinds of texts that would be challenging for a group of readers, they can systematically target instruction or provide scaffolding that will allow students to encounter difficult text in a controlled fashion, as is done in

MCI. The teacher models appropriate learning strategies for students, such as rereading or applying appropriate vocabulary strategies, so that students can then learn what to do when comprehension breaks down. Then, students practice these metacognitive strategies on selected text while the teacher monitors their progress.

Teachers can use Lexiles to guide a struggling student toward texts at the lower end of the student's Lexile range (100L to 50L below his or her Lexile measure). Similarly, advanced students can be adequately challenged by reading texts at the midpoint of their Lexile range, or slightly above. Challenging new topics or genres may be approached in the same way.

Reader-focused adjustment of the reading experience also relates to the student's motivation and purpose. If a student is highly motivated for a particular reading task (e.g., self-selected free reading), the teacher may suggest books higher in the student's Lexile range. If the student is less motivated or intimidated by a reading task, material at the lower end of his or her Lexile range can provide the basic comprehension support to keep the student from feeling overwhelmed.

Target Instruction to Students' Abilities. To encourage optimal progress with the use of any reading materials, teachers need to be aware of the difficulty level of the text relative to a student's reading level. A text that is too difficult may serve to undermine a student's confidence and diminish learning. Frequent use of text that is too easy may foster poor work habits and unrealistic expectations that will undermine the later success of the best students.

When students confront new kinds of texts, the introduction can be softened and made less intimidating by guiding the student to easier reading. On the other hand, students who are comfortable with a particular genre or format can be challenged with more difficult readability levels, which will prevent boredom and promote the greatest rate of development of vocabulary and comprehension skills.

To become better readers, students need to be challenged continually—they need to be exposed to less frequent and more difficult vocabulary in meaningful contexts. A 75% comprehension level provides an appropriate level of challenge, but is not too challenging. If text is too difficult for a reader, the result is frustration and potentially a growing dislike for reading. If text is too easy, the result is often boredom. Targeting reading levels promotes growth and literacy by providing the optimal balance.

Apply Lexiles Across the Curriculum. Over 450 publishers Lexile their trade books and textbooks, enabling educators to link all of the different components of the curriculum to more effectively target instruction. With a student's Lexile measure, teachers can connect him or her with tens of thousands of books (www.Lexile.com) and tens of thousands of newspaper and magazine articles (through periodical databases) that also have Lexile measures.

Use Lexiles in the Classroom

- Develop individualized reading lists that are tailored to provide appropriately challenging reading.
- Enhance thematic teaching by building a bank of titles at varying levels that not only support the theme, but also provide a way for all students to successfully learn about and participate in discussions about the theme. Many database service providers have had their collection of articles measured and have Lexile search capabilities. Users can

search for topics of interest at specific Lexile levels. This is a tremendous resource for content-area teachers.

- Use Lexiles as an additional organizing tool when sequencing materials. For example, an instructor might choose one article a week for use as a read-aloud. In addition to considering the topic, the instructor could increase the difficulty of the articles throughout the course. This approach is also useful when utilizing a core program or textbook that is set up in anthology format. (The order of the anthologies may need to be rearranged to best meet student needs.)
- Develop a reading folder that goes home with students and comes back for weekly review. The folder can contain a reading list of books within the student's Lexile range, reports of recent assessments, and a form to record reading that occurs at home.
- Choose texts lower in the student's Lexile range when factors make the reading situation more challenging, threatening or unfamiliar. Select texts at or above the student's range to stimulate growth when a topic is of extreme interest to a student, or when adding additional support such as background teaching or discussion.
- Use the free Lexile Book Database and Find a Book website (at www.Lexile.com/booksearch) to support book selection and create booklists within a student's Lexile range to help the student make informed choices when selecting texts.
- Use the free Lexile Calculator (at www.Lexile.com) to gauge expected reading comprehension at different Lexile measures for readers and texts.

Use Lexiles in the Library

- Make the Lexile measures of books available to students to better enable them to find books of interest at their appropriate reading level.
- Compare student Lexile levels with the Lexile levels of the books and periodicals in the library to analyze and develop the collection to more fully meet the needs of all students.
- Use the database resources to search for articles at specific Lexile levels to support classroom instruction and independent student research. A list of the database service providers that have had their collections measured can be found at www.Lexile.com/resources. Click on Library Resources to find the list.
- Use the free Lexile Book Database (at www.Lexile.com) to support book selection and create booklists within a student's Lexile range to help educators guide student reading selections.

Use Lexiles at Home

- Ensure that your child gets plenty of reading practice, concentrating on material within his or her Lexile range. Ask your child's teacher or school librarian to print a list of books in your child's range, or search the Lexile Book Database or Find a Book website (at www.Lexile.com/booksearch).
- Communicate with your child's teacher and school librarian about his or her reading needs and accomplishments. They can use the Lexile scale to let you know their assessment of your child's reading ability.
- When a reading assignment proves too challenging for your child, use activities to help. For example, review the words and definitions from the glossary, and the review questions at the end of a chapter before your child reads the text. Afterward, be sure to return to the glossary and review questions to make certain your child understood the material.
- Celebrate your child's reading accomplishments. One of the great things about the Lexile Framework is that it provides an easy way for readers to keep track of their own

growth and progress. You and your child can set goals for reading—sticking to a reading schedule, reading a book at a higher Lexile measure, trying new kinds of books and articles, or reading a certain number of pages per week. When your child hits the goal, make an occasion out of it!

Limitations of the Lexile Framework. Just as variables other than temperature affect comfort, variables other than semantic and syntactic complexity affect reading comprehension. A student's personal interests and background knowledge are known to affect comprehension. However, although temperature alone does not fully identify the comfort level of an environment, we do not dismiss the importance of the information communicated by temperature. Similarly, the information communicated by the Lexile Framework is valuable, even though other information also enhances instructional decisions. In fact, the meaningful communication that is possible when test results are linked to instruction provides the opportunity for parents and students to give input regarding interests and background knowledge.

Results of the MCI Pre- and Post-Tests and Grade Levels. Lexile measures do not translate specifically to grade levels. Within any grade there will be students who vary widely in their reading ability and text that varies in its difficulty. For example, in a sixth-grade classroom there may be readers who are far ahead of their classmates (about 250L above the typical reader) and there will be some readers who are far below (about 250L below the typical reader). To say that some books are “just right” for sixth graders assumes that all sixth graders are reading at the same level. The Lexile Framework can be used to match readers with texts at whatever level the reader is reading.

Simply because a student is an excellent reader, it should not be assumed that the student will necessarily comprehend a text typically found at a higher grade level. Without adequate background knowledge, the words may not have sufficient meaning to the student. A high Lexile measure for a grade indicates that the student can read grade-appropriate materials at a higher comprehension level (90% or above, for example).

The real power of the Lexile Framework is in examining the growth of readers—wherever the reader may be in the development of his or her reading skills. Readers can be matched with texts that they are forecasted to read with 75% comprehension. As a reader grows, he or she can be matched with more demanding texts. And as the texts become more demanding the reader grows.

Development of the *MCI Pre- and Post-Tests*

The *MCI Pre- and Post-Tests* are designed to measure student reading ability. EPS identified the criteria for the development of the assessment:

- Simplified test administration and scoring that could be accomplished through a paper-based environment and online.
- Development of test forms that could target struggling readers at multiple grade levels.
- Development of a sufficient number of test forms to monitor student growth in reading for each *MCI* instructional level.
- Minimum number of items per test form and minimum administration time while still ensuring minimal measurement error when determining each student's reading ability.

Development of the test specifications for the assessments began in the fall of 2008. The passages were commissioned during the winter of 2009 with item development and review following closely behind during late winter and early spring. Test development and final test evaluation occurred in late spring. The following sections describe the specifications for the final forms.

Test Specifications

The specifications for the assessments included one pre-test and three post-tests. The pre-test is designed to assess a wide range of reading abilities because it will be administered to students at middle school and beyond. The pre-test is designed to help educators determine an initial starting reading ability for students entering the *MCI* program. The specifications for the pre-test called for 40 items with a range of difficulty from the 5th percentile of 6th grade student ability (450L) to the 90th percentile of 8th grade student ability (1290L). The target mean was set to approximately the 50th percentile of 6th grade (870L).

Each post-test also contains 40 items. The specifications for the pre-test and the post-tests called for approximately 30% fiction passages and items and approximately 70% nonfiction passages and items on each form. The three post-test forms were designed to have a parallel structure, and differ only in the level of difficulty of the reading demand. The Aqua Post-Test is centered at approximately the 30th percentile of 6th grade (750L), the Gold Post-Test is centered at approximately the 30th percentile of 7th grade (800L), and the Crimson Post-Test is centered at approximately the 30th percentile of Grade 8 (880L). Reading levels are presented in *Table 8*.

Table 8. Lexile targets by reading level for each *MCI* test.

Test	Center	Reading Range
Pre-Test	870L	450L to 1290L
Aqua Post-Test	750L	450L to 1050L
Gold Post-Test	800L	550L to 1150L
Crimson Post-Test	880L	600L to 1200L

The pre-test was designed to contain only native-Lexile items (short passages). Each native-Lexile item consists of one short passage and the item. The post-tests were designed to have a combination of native-Lexile items (short passages and item) and passage-native Lexile items (long passages and multiple items). Each passage-native Lexile item is part of a group of items attached to a long passage. The long passages developed for the *MCI Pre- and Post-Tests* contained five embedded items. The distributions of passages and items for each test are shown in *Tables 9 and 10*.

Table 9. Specifications for test passages.

	Native-Lexile Passages (short)	Passage-Native Passages (long)
Pre-Test	40	0
Aqua Post-Test	15	5
Gold Post-Test	15	5
Crimson Post-Test	15	5
Total Passages Needed	85	15

Table 10. Specifications for test items.

	Native-Lexile Items (short passages)	Passage-Native Lexile Items (long passages)	Total Items Needed
Pre-Test	40	0	40
Aqua Post-Test	15	25	40
Gold Post-Test	15	25	40
Crimson Post-Test	15	25	40
Total Number of Items Needed	85	75	160

Passage Development

All of the passages required for development of the *MCI Pre- and Post-Tests* were either written by staff at MetaMetrics or commissioned by MetaMetrics. The long passages developed for the passage-native Lexile items were reviewed and approved by EPS prior to item development. After the passages were approved, the passage-native Lexile items were written and sent to EPS for review along with their associated passages. The passages for the short, native-Lexile items were developed in conjunction with their associated items. The passages and items together were sent to EPS for review.

Each reading passage was measured using the Lexile Analyzer to determine the Lexile measure of the passage. The design specifications for the *MCI Pre- and Post-Tests* required that the passages had Lexile measures within specific ranges. To achieve the required number of passages and items, 20 long passages and 95 short passages were developed and submitted for review. From the submitted passages, 15 long and 85 short passages were used to complete the final test forms. *Table 11* contains the Lexile distribution requirements of the passages. *Tables 12* and *13* contain the specifications for the pre-test and the post-tests.

Table 11. Passage Lexile distribution requirements for MCI Pre- and Post-Tests.

Lexile Range	Long Passages (Passage-native Lexile)	Short Passages (Native-Lexile)
400L to 490L	1	2
500L to 590L	1	7
600L to 690L	3	13
700L to 790L	2	13
800L to 890L	3	16
900L to 990L	1	12
1000L to 1090L	2	12
1100L to 1190L	2	6
1200L to 1290L	0	4
Total	15	85

Table 12. Specification of native-Lexile items for the pre-test and post-tests.

Lexile Range	Pre-Test	Aqua Post-Test	Gold Post-Test	Crimson Post-Test
400L to 490L	1	1		
500L to 590L	4	2	1	
600L to 690L	4	3	3	3
700L to 790L	6	3	2	2
800L to 890L	6	2	4	4
900L to 990L	6	2	2	2
1000L to 1090L	6	2	2	2
1100L to 1190L	4		1	1
1200L to 1290L	3			1
Mean	870L	750L	800L	880L

Table 13. Specification of passages for passage-native Lexile items for the pre-test and post-tests.

Lexile Range	Aqua Post-Test	Gold Post-Test	Crimson Post-Test
400L to 490L	1	1	
500L to 590L			
600L to 690L	1	1	1
700L to 790L	1	1	1
800L to 890L	1		1
900L to 990L		1	
1000L to 1090L	1		1
1100L to 1190L		1	1
1200L to 1290L			
Mean	750L	800L	880L

Because struggling readers may have difficulty reading nonfiction text, and because many curriculum standards have begun to increase the proportion of nonfiction text students are expected to read, the passage specifications for the *MCI Pre- and Post-Tests* were set at 30% fiction and 70% nonfiction.

Passages used for the passage-native Lexile items (five items per passage) were approximately 300 words in length. Passages used with the native-Lexile items (one item per passage) were no longer 125 words in length. A frame was written for each passage-native Lexile passage to introduce the passage and to direct the reader. The convention for the frame was approximately 1 to 2 sentences in length.

The following criteria were used as guidelines in the development and review of passages for the *MCI Pre- and Post-Tests*.

- Passages should be developmentally appropriate for students in middle school and beyond. The selected passages should have appropriate content for a reading assessment.
- Passages selected should be approximately 300 words in length. Passages in the lower Lexile ranges may have fewer words and passages in the higher Lexile ranges may have more words.
- Passages should be interesting for a typical struggling reader at the middle school or high school level.
- Passages should stand alone as a uniform piece of text that can be understood without additional explanatory text other than an introductory frame.
- Passages and items should be free of bias based on race, gender, age, ethnicity, religion, disability, sexual orientation, or socioeconomic status. No group should have an advantage over another because of values, vocabulary, phrasing, or assumptions in a passage.
- Passages should be free of stereotypes of any group (e.g., race/ethnicity, gender).
- To the degree possible, prior knowledge should not be required for the examinee to understand or appreciate the passage. References to events, people, and places should be explained within the passage unless considered common knowledge. The introductory frame can be used to explain necessary information not provided in the passage.
- Passages should avoid topics that may be offensive to, or induce an emotional reaction from, an examinee, parent, or citizen group (e.g., violence, abuse, terminal illness, and poverty).
- Passages and items should be free of registered trademarks and brand names. Common business names should also be avoided.
- Generally, contact information should not be given in a passage. However, when necessary to include fictional contact information (e.g., a customer service phone number in an appliance manual), it should be modeled after real-world contact information. To ensure that the contact information is not real, follow these guidelines:

Phone numbers

Local phone numbers should appear as 555 + (0100–0199), (e.g., 555–0185).

Long distance numbers should appear as (live area code [except 555] + 555 + 0100–0199), (e.g., 319–555–0177).

Toll-free numbers should appear as 1 + 888 + (100–199) + four digits (e.g., 1–888–190–4455).

“Vanity” numbers may be used following the above guidelines (e.g., 1–888–100–1–FUN).

Addresses

Complete addresses should use a mismatched city and zip code.

To the extent possible, verify that the named business or individual cannot be reached in the named city.

URLs and E-Mail Addresses

URLs should be acceptable to EPS and not proprietary (e.g., www.cityzoo.pqrf.com).

E-mail addresses should end in @pqrf.com or another address that is acceptable to EPS and not proprietary (e.g., abc@pqrf.com).

- All source material used for the development of passages must be documented, and copies of book, magazine, or web pages, and any other source material should be sent back to MetaMetrics with the passage. Some helpful websites (for content and style) include:

<http://news.nationalgeographic.com/kids/>

<http://www.timeforkids.com/TFK/news>

<http://www.factmonster.com/>

Although the content of the texts used in the passages could be altered if necessary, it was important to develop passages that were free of sensitive issues. The following guidelines were used to help ensure the creation of non-offensive and bias-free passages. These guidelines were assembled from the results of MetaMetrics' collaboration with various partners in textbook and test publishing.

1. Violence/crime: Avoid weapons, fights, arrests, illegal activities, abuse, and murders.
2. Depressing situations or death: Avoid sickness, death, and other negative situations.
3. Offensive language: Avoid use of curse words or words used to cover up a harsher curse; avoid oaths such as "Oh God!", words that belittle others, or other insulting words such as "ignorant," or "ugly."
4. Drugs/alcohol/tobacco: Avoid any mention of drugs, alcohol, tobacco, and anything associated with these topics such as rehab, bars, etc.
5. Sex/attraction: Avoid issues that call for a discussion of sex, sexual orientation, or relationships of either a romantic or sexual nature.
6. Race: Avoid racial slurs, belittling words, stereotypes and unbalanced representations of a race.
7. Class: Avoid mentioning economic and social differences and avoid stereotypes.
8. Gender: Use gender free language (e.g., firefighter instead of fireman); avoid using male pronouns to refer to both sexes; show both genders in a variety of roles; avoid stereotypical portrayals of men or women.
9. Religion: Avoid selections that promote or demean a religious belief; avoid the assumption that people share a common belief; avoid mention of a reference to any holidays of a religious nature (e.g., Christmas, Halloween).
10. Supernatural/magic: Avoid mention of witches, goblins, wizards, and other supernatural beings; avoid magic in general.
11. Parents/family: Avoid selections that question parents, authority, or judgment; avoid negative relationships within the family; avoid raising the issue of alternative families.
12. Politics: Avoid controversial issues (e.g., unions, strikes) and selections, which portray political bias.
13. Animals/environment: Avoid hunting and cruelty to animals (e.g., fur coats, trapping animals) and be sensitive to environmental issues and animal rights.
14. Brand names/junk food: Avoid mentioning either.

Item Development

The pre-test and post-test items were developed concurrently. One native-Lexile item was developed for each native passage. Five passage-native Lexile items were developed for each long passage.

The passage-native and native completion format is similar to the fill-in-the-blank format. When properly written, this format directly assesses the reader's ability to draw inferences and establish logical connections between the ideas in the passage. In the passage-native Lexile format, the reader is presented with text of approximately 300 words in length. In the native-Lexile format, the reader is presented with text of approximately 115 words in length. The passage is then response illustrated: a statement is added at the end of the passage with a missing word or phrase followed by four options. From the four presented options, the reader is asked to select the "best" option that completes the statement. With this format, all options are syntactically and semantically appropriate completions of the sentence, but one option is unambiguously the "best" option when considered in the context of the passage. This format is "well-suited for testing a student's ability to evaluate" (Haladyna, 1994, p. 62). In addition, this format is also useful as an instructional tool.

Below is a sample native-Lexile passage and item provided as an example of the type of item used in the *MCI* tests. The correct answer is A: mended.

Have you ever noticed that a hem on your shirt or pants is coming undone? You don't have to get rid of that piece of clothing. Instead you can fix a fraying hem. First you want to fold the cloth over itself so that the fabric cannot fray anymore. Then you use a needle and thread to sew the fabric down. A sewing machine, which usually works best for larger sewing projects, can also be used. A hem can also be fixed using a product called hem repair tape. The tape is placed inside the hem. The fabric is ironed, causing the tape to melt, and then the fabric is stuck together.

A hem can be _____.

- A. mended**
- B. released**
- C. delicate**
- D. invisible**

On the next page is a sample passage with introductory frame and passage-native Lexile items provided as an example of the type of passage-native Lexile passage used in the EPS assessments. The correct answer is followed by an asterisk. This selection has a Lexile measure of 740L.

Read the following passage about the Millennium Ecosystem Assessment and complete the statements. *Choose the best answer based on the information in the passage.*

For four years, the scientists worked together on a report about Earth. It was called the *Millennium Ecosystem Assessment*. They examined the changes in the environment. They studied plants and animals in their habitats. They studied satellite pictures. They analyzed statistics. They ran tests. And, after all that work, they shared their results. They had an important message: All people must work together to reduce damage to the Earth! **People must __1__.**

- 1.
- A. cooperate*
- B. demonstrate
- C. migrate
- D. relax

The director of the group who created the *Millennium Ecosystem Assessment* said that humans are damaging Earth's environment. And we are doing it faster than ever. We need to act now. **The damage is __2__.** If we work together, everyone can achieve more. If we work together, we can reduce the damage we humans cause to the environment. Everyone can do something, and together we can make big changes.

- 2.
- A. unknown
- B. routine
- C. humorous
- D. severe*

The average American throws away 7.5 pounds of garbage every day. A lot of the garbage can be recycled. Many people don't recycle cans, paper, bottles, or glass. About 36% of our garbage is paper and cardboard. For every ton of recycled paper, seven trees are saved. Imagine how many trees we might save if we recycled more! **Recycling __3__ excessive garbage.**

- 3.
- A. produces
- B. reflects
- C. prevents*
- D. manufactures

Another thing we can all do is use less energy. We all use energy for transportation, heat, and light, but there are ways to cut our energy use. **We can all be __4__ of ways to cut energy use.** To save energy, people can ride bikes, take a bus, or walk instead of going in a car. Turning off the lights when leaving a room not only uses less energy, but can also cut down electricity bills. Turning off the TV when no one is watching also helps. Turning down the heat when no one is home can cut down on energy use. These are just a few ways that we can all reduce our everyday energy usage. **Following these __5__ can benefit everyone.**

- 4.
- A. conscious*
- B. tired
- C. cautious
- D. fond

- 5.
- A. resolutions
- B. triumphs
- C. markings
- D. suggestions*

There are two main advantages to using the passage-native Lexile and native-Lexile item formats on the *MCI Pre- and Post-Tests*. The first is that the level of reading of the statement and the four answer options is controlled to ensure that their difficulty level is similar to the difficulty level of the passage. The second advantage is that the statement is crafted to be as short as or shorter than the typical sentence in the passage. These two advantages help ensure that the statement is easier than the accompanying passage.

The statement portion of the completion item can assess a variety of skills related to reading comprehension: paraphrase information in the passage, draw a logical conclusion based on the information in the passage, make an inference, identify a supporting detail, or make a generalization based on the information in the passage. The statement is written to ensure that by reading and comprehending the passage the reader is able to select the correct option. When the passage-native completion statement is read by itself, each of the four options is plausible.

The following criteria are used to develop the statement portion of the passage-native and native completion items.

The statement should:

1. Require the student to draw an unambiguous conclusion or inference from the passage.
2. Not overlap ideas with other questions in the passage.
3. Be clear as to what or whom the sentence question is about.
4. Not use wording that is the same as or closely similar to wording that appears elsewhere in the passage.
5. Attempt to avoid the use of negatives.

The answer choices should:

1. Be reasonably Lexile targeted (100L below to 100L above the passage as a general guideline).
2. Logically complete the statement to force passage dependence for answering correctly. All foils should make sense in the context of the sentence, but only the correct choice should make sense in the context of the paragraph.
3. Be one word or a short phrase.
4. Not be homonyms, as this may merely confuse the reader. Avoid using antonyms; if two choices are opposite there is a high probability that one is correct.
5. Be balanced; if correct choice is a word or phrase containing a positive connotation, at least one other choice should be positive so the correct choice does not stand out. However, with higher-level texts it is best to try and make all of the words positive or negative.
6. Not use negative sentence structure.
7. Be selected in accordance to sensitivity guidelines.

Item writers were MetaMetrics staff who were experienced in item development and who had experience with the everyday reading ability of students at various levels.

Item writers were provided vocabulary lists to use during statement and option development. The vocabulary lists were compiled by MetaMetrics based on research to determine the Lexile measures of words (i.e., their difficulty). The Lexile Vocabulary Analyzer (LVA) determines the Lexile measure of a word using a set of features related to the source text and the word's

prevalence in the MetaMetrics corpus (MetaMetrics, 2006). The rationale used to compile the vocabulary lists was that the words should be part of a reader’s “working” vocabulary if they had likely been encountered in easier text (those with lower Lexile measures).

All items went through a two-stage review process prior to initial delivery to EPS. First, items were reviewed and edited by an editor according to the item development criteria and for sensitivity issues. Items were then reviewed and edited by a group of specialists that represent various perspectives, including test developers and editors. These individuals examined each item for sensitivity issues and for the quality of the response options. During this second stage of the item review process, additional edits were incorporated.

The reviewed items were sent to EPS for additional edits. At this review stage, items were either “approved as presented,” “approved with edits,” or “deleted.” Deleted items were removed from the item pool and additional passages/items were developed to replace them. Items that were “approved with edits” were edited and subsequently approved or deleted by EPS. Only approved items were retained in the item pool.

Test Form Development

The pre-test and three post-test forms were constructed according to the specifications previously described. Statistics for the final forms are shown in *Table 14*.

Table 14. Technical information for the *MCI Pre- and Post-Test* forms.

	Target Lexile Range	Mean Lexile	Standard Dev	Lexile Minimum	Lexile Maximum
Pre-Test	450L – 1290L	875L	212	470L	1250L
Aqua Post-Test	450L – 1050L	751L	182	480L	1070L
Gold Post-Test	550 L – 1150L	813L	186	550L	1140L
Crimson Post-Test	600L – 1200L	883L	182	620L	1200L

Prior to finalizing the test forms, each form went through a three-stage review process. First, the test and passage specifications were reviewed: Lexile measures of items, mean and standard deviation of test form, word counts across the forms, proportion of fiction and nonfiction passages, and distribution of content and distribution of correct responses. Next, the test was taken to verify the answer keys and review the foils in relation to the passages and items. Finally, the overall tests were reviewed for flow and consistency. The following criteria were used to evaluate each set:

Curricular Perspective

1. Do the topics of the passages in a form flow well?
2. Is there a variety of passages on each form and no repeated content (e.g., two passages on extreme sports)?

Psychometric Perspective

1. Do the final forms have the same approximate mean and standard deviation of Lexile measures as the target specifications?
2. Is the distribution of the placement of correct answers within a form approximately equal (about 25% for each response position)?
3. Are runs of the same correct response position avoided? (e.g., more than 3 of any response positions in a row would be undesirable.)
4. Is the use of the same word as the correct response for more than one item on a form avoided?

Where necessary, passages and/or items were revised or replaced and test forms were reconstituted to more closely reflect the test specifications. Because the Lexile Theory was used to estimate the difficulties of the passages and items, it was necessary for the test specifications to be adhered to as closely as possible.

Scoring and Reporting

The *MCI Pre- and Post-Test* scores are reported on the Lexile scale. The typical range of the Lexile scale is from 200 to 1700 Lexiles, although actual Lexile measures can range from below zero to above 2000 Lexiles. Reader Lexile measures are reported in 5-unit intervals. For reporting purposes, the lowest Lexile measure reported with the *MCI Pre- and Post-Tests* is “BR” for “Beginning Reader” (scores at or below 0L).

Individual scores are calculated by first summing the number of correct responses (omitted items and multiple responses are counted as incorrect). The number correct is then converted to a scaled Lexile measure.

There are many reasons to use scale scores rather than raw scores to report test results. Scale scores overcome the disadvantage of many other types of scores (e.g., percentiles and raw scores), in that equal differences between scale score points represent equal differences in ability. Each question on a test has a unique level of difficulty; therefore, answering 23 questions correctly on one form of a test requires a slightly different level of ability from answering 23 items correctly on another form of the test. Receiving a scale score (Lexile measure) of 675L on one form of a test, however, represents a similar level of reading ability as receiving a scale score (Lexile measure) of 675L on another form of the test.

The scale score associated with the number of items correct for each test form was generated through the use of item difficulty modeling and likelihood analysis. The difficulty of each item was calculated based on the theoretical difficulty of the passage (Lexile measure) to which it is associated. The difficulty of a passage determines the mean of the item difficulties developed for that text. Given these modeled item difficulties, the scale score is then computed by performing a likelihood analysis that results in a Lexile measure being associated with each raw score.

Correspondence tables were provided for each test form based upon the difficulties of the passages and the items on the form. The Lexile uncertainties (standard errors) for each score point were also provided.

Conventions for Reporting. Lexile measures are reported as a number followed by a capital “L” for “Lexile.” There is no space between the measure and the “L” and measures of 1,000 or greater are reported without a comma (e.g., 1050L).

The measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is accountability (at the student, school, or district level), then actual measures should be reported at all score points. If the purpose is instructional, then the scores should be capped at the upper bounds of measurement error (90th percentile point based on prior research by MetaMetrics, Inc. with the Lexile Framework). In an instructional environment where the purpose of the Lexile measure is to appropriately match readers with books, all scores at or below 0L should be reported as “BR” (Beginning Reader); no student should receive a negative Lexile measure.

Test Use Guidelines. Students should not be administered a specific test form more than once within any one year. When a student takes the same assessment form a second time within the span of one year, we are unsure as to how to interpret change in Lexiles: (1) because the

student's reading ability has improved/grown, or (2) because the student remembers some of the items and has experience with the testing environment.

Assessment practices should be in accordance with the generally accepted ethical standards of the education profession. Accordingly, any practice that increases students' scores should simultaneously represent an increase in students' mastery (i.e., increasing students' abilities to perform skills or demonstrate knowledge in real world situations) of the content domains tested. For more information, refer to *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999).

Reliability

If use is to be made of some piece of information, then the information should be reliable—stable, consistent, and dependable. In reality, all test scores have some error (or level of uncertainty). This uncertainty in the measurement process is related to three factors: (1) the statistical model that was used to compute the score, (2) the questions that were used to determine the score, and (3) the condition of the reader when the questions used to determine the score were collected. Once the level of uncertainty in a test score is known, then it can be taken into account when using the test results.

Reliability, or the consistency of scores obtained from an assessment, is a major consideration in evaluating any assessment procedure. Two sources of uncertainty have been examined with *MCI Pre- and Post-Tests*—text error and reader error.

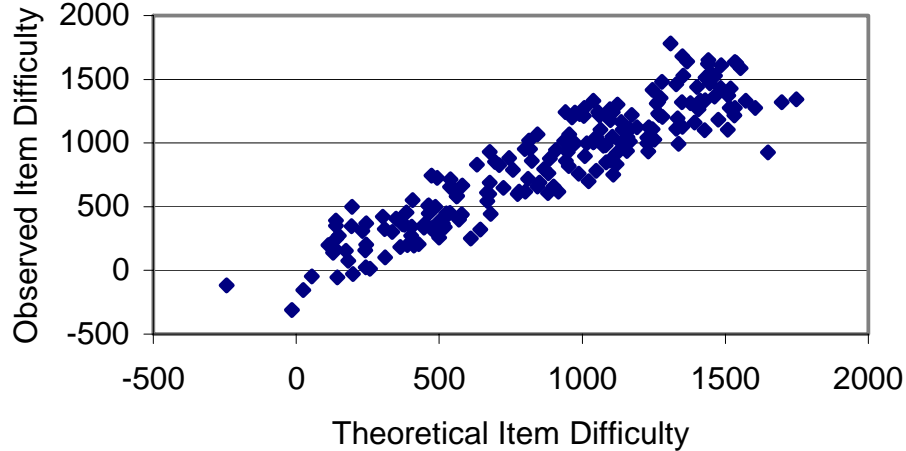
Text Measure Error Associated with The Lexile Framework for Reading

To determine a Lexile measure for a text, the standard procedure is to process the entire text. All pages in the work are concatenated into an electronic file that is processed by a software package called the Lexile Analyzer (developed by MetaMetrics, Inc.). The analyzer “slices” the text file into as many 125-word passages as possible, analyzes the set of slices, and then calibrates each slice in terms of the logit metric. That set of calibrations is then processed to determine the Lexile measure corresponding to a 75% comprehension rate. The analyzer uses the slice calibrations as test item calibrations and then solves for the measure corresponding to a raw score of 75% (e.g., 30 out of 40 correct, as if the slices were test items). Obviously, the measure corresponding to a raw score of 75% on *Goodnight Moon* (300L) slices would be lower than the measure corresponding to a comparable raw score on *USA Today* (1200L) slices. The Lexile Analyzer automates this process, but what “certainty” can be attached to each text measure?

Using the bootstrap procedure to examine error due to the text samples, the above analysis could be repeated. The result would be an identical text measure to the first because there is no sampling error when a complete text is calibrated.

Study 1. There is, however, another source of error that increases the uncertainty about where a text is located on the Lexile Map. The Lexile Theory is imperfect in its calibration of the difficulty of individual text slices. To examine this source of error, 200 items that had been previously calibrated and shown to fit the model were administered to 3,026 students in grades 2 through 12 in a large urban school district. For each item, the observed item difficulty calibrated from the Rasch model was compared with the theoretical item difficulty calibrated from the regression equation used to calibrate texts. A scatter plot of the data is presented in *Figure 2*.

Figure 2. Scatter plot between observed item difficulty and theoretical item difficulty.



The correlation between the observed and the theoretical calibrations for the 200 items was 0.92 and the root mean square error was 178L. Therefore, for an individual slice of text the measurement error is 178L.

The standard error of measurement associated with a text is a function of the error associated with one slice of text (178L) and the number of slices that are calibrated from a text. Very short books have larger uncertainties than longer books. A book with only four slices would have an uncertainty of 89 Lexiles whereas a longer book such as *War and Peace* (4,082 slices of text) would only have an uncertainty of 3 Lexiles (*Table 15*).

Table 15. Standard errors for selected values of the length of the text.

Title	Number of Slices	Lexile Text Measure	Standard Error of Text
<i>The Stories Julian Tells</i>	46	520L	26
<i>Bunicula</i>	102	710L	18
<i>The Pizza Mystery</i>	137	620L	15
<i>Meditations of First Philosophy</i>	206	1720L	12
<i>Metaphysics of Morals</i>	209	1620L	12
<i>Adventures of Pinocchio</i>	294	780L	10
<i>Red Badge of Courage</i>	348	900L	10
<i>Scarlet Letter</i>	597	1420L	7
<i>Pride and Prejudice</i>	904	1100L	6
<i>Decameron</i>	2431	1510L	4
<i>War and Peace</i>	4082	1200L	3

A typical grade 3 reading test has appropriately 2,000 words in the passages. To calibrate this text, it would be sliced into 16 125-word passages. The error associated with this text measure

would be 45L. A typical grade 7 reading test has approximately 3,000 words in the passages and the error associated with the text measure would be 36L. A typical grade 10 reading test has approximately 4,000 words in the passages and the error associated with the text measure would be 30L.

Study 2. A second study was conducted by Stenner, Burdick, Sanford, and Burdick (2006) during 2002 to examine ensemble differences across items. An ensemble consists of the all of the items that could be developed from a selected piece of text. The Lexile measure of a piece of text is the mean difficulty.

Participants. Participants in this study were students from four school districts in a large southwestern state. These students were participating in a larger study that was designed to assess reading comprehension with the Lexile scale. The total sample included 1,186 grade 3 students, 893 grade 5 students, and 1,531 grade 8 students. The mean tested abilities of the three samples were similar to the mean tested abilities of all students in each grade on the state reading assessment. Though 3,610 students participated in the linking study, the data records for only 2,867 of these students were used for determining the ensemble item difficulties presented in this paper. The students were administered one of four forms at each grade level. The reduction in sample size occurred because one of the four forms was created using the same ensemble items as another form. For consistency of sample sizes across forms, the data records from this fourth form were not included in the ensemble study.

Instrument. Thirty text passages were response-illustrated by three different item writing teams resulting in three items nested within each of 30 passages for a total of 90 items. All three teams employed a similar item-writing protocol. The ensemble items were spiraled into test forms at the grade level (3, 5, or 8) that most closely corresponded with the item's theoretical calibration.

Winsteps (Wright & Linacre, 2003) was used to estimate item difficulties for the 90 ensemble study items. Of primary interest in this study was the correspondence between theoretical text calibrations and the 30 ensemble means and the consequences that theory misspecification holds for text measure standard errors.

Results. Table 16 presents the ensemble study data in which three independent teams wrote one item for each of thirty passages for ninety items. Observed ensemble means taken over the three ensemble item difficulties for each passage are given along with an estimate of the within ensemble standard deviation for each passage.

Table 16. Analysis of 30 item ensembles providing an estimate of the theory misspecification error.

Item Number	Theory (T)	Team A	Team B	Team C	Mean ^a (O)	SD ^b	Within Ensemble Variance	T-O
1	400L	456	553	303	437	126	15,909	-37
2	430L	269	632	704	535	234	54,523	-105
3	460L	306	407	483	399	88	7,832	61
4	490L	553	508	670	577	84	6,993	-87
11	510L	267	602	468	446	169	28,413	64
5	540L	747	825	654	742	86	7,332	-202
6	569L	909	657	582	716	172	29,424	-147
7	580L	594	683	807	695	107	11,386	-115
8	620L	897	805	497	733	209	43,808	-113
9	720L	584	850	731	722	133	17,811	-2
12	720L	953	587	774	771	183	33,386	-51
13	745L	791	972	490	751	244	59,354	-6
14	770L	855	1017	958	944	82	6,717	-174
16	770L	1077	1095	893	1022	112	12,446	-252
15	790L	866	557	553	659	180	32,327	131
21	812L	902	1133	715	917	209	43,753	-105
10	820L	967	740	675	794	153	23,445	26
17	850L	747	864	674	762	96	9,257	88
22	866L	819	809	780	803	20	419	63
18	870L	974	1197	870	1014	167	28,007	-144
19	880L	1093	733	692	839	221	48,739	41
23	940L	945	1057	965	989	60	3,546	-49
24	960L	1124	1205	1170	1166	41	1,653	-206
25	1010L	926	1172	899	999	151	22,733	11
20	1020L	888	1372	863	1041	287	82,429	-21
26	1020L	1260	987	881	1043	196	38,397	-23
27	1040L	1503	1361	1239	1368	132	17,536	-328
28	1060L	1109	1091	981	1061	69	4,785	-1
29	1150L	1014	1104	1055	1058	45	2,029	92
30	1210L	1275	1291	1014	1193	156	24,204	17

Total MSE = Average of $(T-O)^2 = 12022$; Pooled within variance for ensembles = 7984; Remaining between ensemble variance = 4038; Theory misspecification error = 64L.

Barlett's test for homogeneity of variance produced an approximate chi-square statistic of 24.6 with 29 degrees of freedom and sustained the null hypothesis that the variances are equal across ensembles.

Note. All data are reported in Lexiles.

a. Mean (O) is the observed ensemble mean.

b. SD is the standard deviation within ensemble.

The difference between passage text calibration and observed ensemble mean is provided in the last column. The RMSE from regressing observed ensemble means on text calibrations is 110L. *Figures 3a* and *3b* show plots of observed ensemble means against theoretical text calibrations.

Figure 3a. Plot of observed ensemble means and theoretical calibrations (RMSE = 110L).

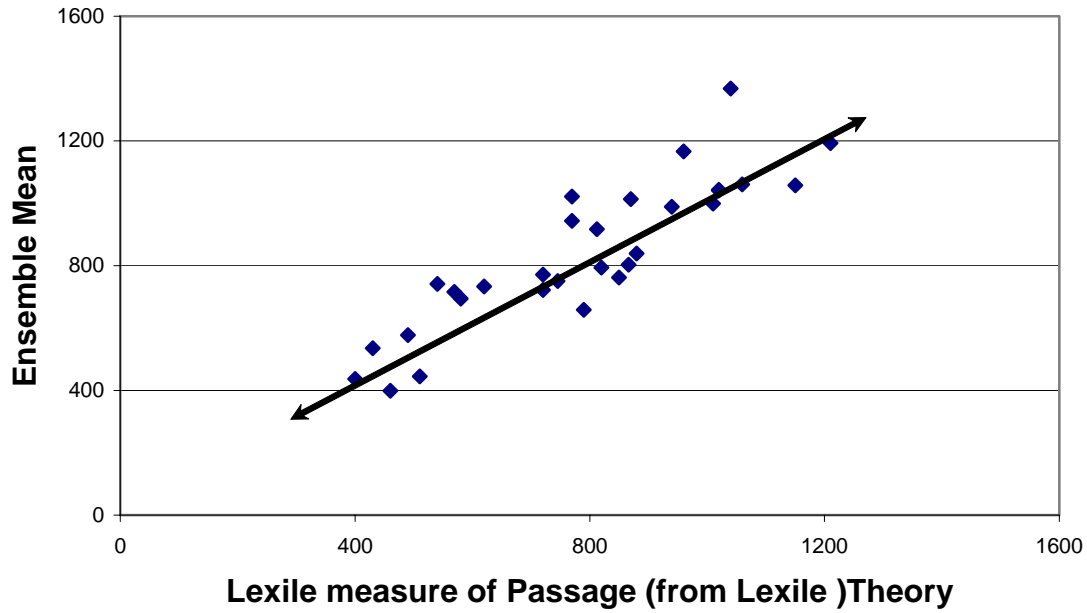
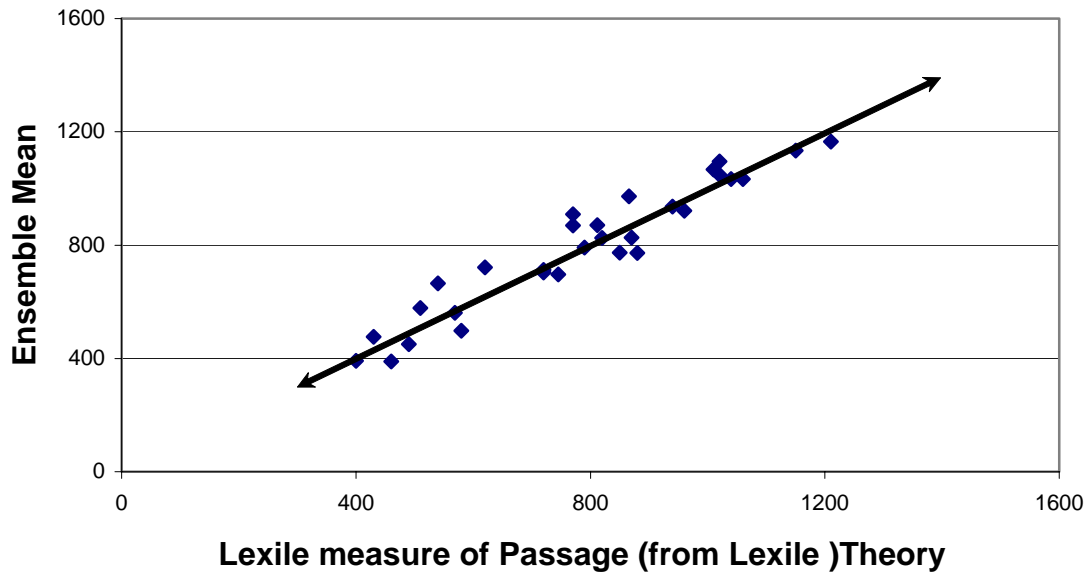


Figure 3b. Plot of simulated “true” ensemble means and theoretical calibrations (RMSE = 64L).



Note that some of the deviations about the identity line are because ensemble means are poorly estimated given that each mean is based on only three items. *Figure 3b* depicts simulated data when an error term [distributed $\sim N(0, \sigma = 64L)$] is added to each theoretical value. Contrasting the two plots in *Figures 3a* and *3b* provide a visual depiction of the difference between regressing observed ensemble means on theory and regressing “true” ensemble means on theory. An estimate of the RMSE when “true” ensemble means are regressed on the Lexile Theory is $64L (\sqrt{110^2 - 89^2} = \sqrt{4,038} = 63.54)$. This is the average error at the passage level when predicting “true” ensemble means from the Lexile Theory.

Since the RMSE equal to $64L$ applies to the expected error at the passage/slice level, a text made up of n_i slices would have an expected error of $64 \div \sqrt{n_i}$. Thus, a short periodical article of 500 words ($n_i = 4$) would have a SEM of $32L (64 \div \sqrt{4})$, whereas a much longer text like the novel *Harry Potter: Chamber of Secrets* (880L, Rowling, 2001) would have a SEM of $2L (64 \div \sqrt{900})$. *Table 17* contrasts the SEMs computed using the old method with SEMs computed using the Lexile Framework for several books across a broad range of Lexile measures.

Table 17: Old method text readabilities, resampled SEMs, and new SEMs for selected books.

Book	Number of Slices	Lexile Measure	Resampled Old SEM ^a	New SEM
The Boy Who Drank Too Much	257	447L	102	4
Leroy and the Old Man	309	647L	119	4
Angela and the Broken Heart	157	555L	118	5
The Horse of Her Dreams	277	768L	126	4
Little House by Boston Bay	235	852L	126	4
Marsh Cat	235	954L	125	4
The Riddle of the Rosetta Stone	49	1063L	70	9
John Tyler	223	1151L	89	4
A Clockwork Orange	419	1260L	268	3
Geometry and the Visual Arts	481	1369L	140	3
The Patriot Chiefs	790	1446L	139	2
Traitors	895	1533L	140	2

Notes: (a) Three slices selected for each replicate: one slice from the first third of the book, one from the middle third and one from the last third. Resampled 1,000 times. SEM = SD of the resampled distribution.

Standard Error of Measurement

Uncertainty and Standard Error of Measurement. Because of the presence of measurement error associated with test unreliability, there is always some uncertainty about a student’s true score. This uncertainty is known as the standard error of measurement (SEM). The magnitude of the SEM of an individual student’s score depends on the following characteristics of the test:

- the number of test items—smaller standard errors are associated with longer tests,

- the quality of the test items—in general, smaller standard errors are associated with highly discriminating items for which correct answers cannot be obtained by guessing, and
- the match between item difficulty and student ability—smaller standard errors are associated with tests comprised of items with difficulties approximately equal to the ability of the student (targeted tests).

Whenever a model is used to explain the relationship between parameters, some of the differences between observed and theoretical measures cannot be explained. The *MCI Pre- and Post-Tests* were developed using the Rasch one-parameter item response theory model to relate a reader's ability and the difficulty of the items. There is a unique amount of measurement error due to model misspecification (violation of model assumptions) associated with each score on the assessment. *Table 18* describes the uncertainties due to model misspecification for the *MCI Pre- and Post-Tests*. The Lexile ranges shown in the table indicate reader measures associated with scores of approximately 25% to approximately 75% correct. Complete correspondence tables for all possible scores on each assessment have been provided to EPS.

Table 18. Uncertainties for MCI Pre- and Post-Tests by Lexile range (25% - 75% correct).

	Pre-Test	Aqua Post-Test	Gold Post-Test	Crimson Post-Test
100L to 195L				
200L to 295L		74		
300L to 395L		70	72	
400L to 495L	72	66	68	71
500L to 595L	68	66	66	67
600L to 695L	65	67	66	66
700L to 795L	66	71	69	67
800L to 895L	68		74	70
900L to 995L	68			
1000L to 1095L				
Average	68	68	69	68

Validity

The 1999 *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education) state that “validity refers to the degree to which evidence and theory support the interpretations of test scores entailed in the uses of tests” (p. 9). Validity evidence provides information about how well a test will fulfill its intended function. “The process of ascribing meaning to scores produced by a measurement procedure is generally recognized as the most important task in developing an educational or psychological measure, be it an achievement test, interest inventory, or personality scale” (Stenner, Smith, and Burdick, 1983). Because a test score from an *MCI Pre- or Post-Test* will be used as a measure of the reading ability of a student and will be used to target reading instruction, validity evidence should primarily focus on the degree to which the *MCI Pre- and Post-Tests* measure reading comprehension of appropriate reading material. For convenience, the various sources of validity evidence—content, criterion-related, and construct validity evidence—will be described as if they are unique, independent components rather than interrelated parts. At this time the primary source of validity evidence comes from examination of the content of the *MCI Pre- and Post-Tests* and the degree to which the assessments can be said to measure reading comprehension (construct validity evidence). As more data are collected and more studies are completed, additional validity evidence will be described.

Content Validity Evidence

Validity evidence for the content of a test relates to the degree to which the test content is supportive of the intended interpretations of the test scores. The *MCI Pre- and Post-Tests* have been designed to measure reading comprehension ability of fiction and nonfiction texts. To this end, both fiction and nonfiction texts have been included in the tests. In addition, the text difficulty of the reading passages was analyzed using the Lexile Analyzer to ensure that the difficulty of the text was appropriate for the students taking the pre-test and the post-tests. The difficulty of the item vocabulary was also matched to the difficulty of the passage. All passages were designed to reflect authentic material, and students are asked to respond to the text in ways that are appropriate for the genre. The passages and items were thoroughly reviewed prior to placement on a test.

Construct Validity Evidence

Evidence for construct validity of the *MCI Pre- and Post-Tests* is provided by the extensive body of research supporting the Lexile Framework for Reading. The development of the *MCI Pre- and Post-Tests* utilized tools for text measurement such as the Lexile Analyzer and procedures for item development that have been shown to result in effective measures of reading comprehension. All of the items on the *MCI Pre- and Post-Tests* are native items or passage-native items in the family of items upon which the research on the Lexile Framework was based. The section in this technical report entitled *The Lexile Framework for Reading* provides a detailed description of the framework and evidence to support that tests based upon the framework measure reading comprehension.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bormuth, J.R. (1966). Readability: New approach. *Reading Research Quarterly*, 7, 79-132.
- Bormuth, J.R. (1967). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, February 1967, 292-299.
- Bormuth, J.R. (1968). Cloze test readability: Criterion reference scores. *Journal of Educational Measurement*, 3(3), 189-196.
- Bormuth, J.R. (1970). *On the theory of achievement test items*. Chicago: The University of Chicago Press.
- Carroll, J.B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.
- Carver, R.P. (1974). Measuring the primary effect of reading: Reading storage technique, understanding judgments and cloze. *Journal of Reading Behavior*, 6, 249-274.
- Chall, J.S. (1988). "The beginning years." In B.L. Zakaluk and S.J. Samuels (Eds.), *Readability: Its past, present, and future*. Newark, DE: International Reading Association.
- Crain, S. & Shankweiler, D. (1988). "Syntactic complexity and reading acquisition." In A. Davidson and G.M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Erlbaum Associates.
- Davidson, A. & Kantor, R.N. (1982). On the failure of readability formulas to define readable text: A case study from adaptations. *Reading Research Quarterly*, 17, 187- 209.
- Dunn, L.M. & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test-Revised*, Forms L and M. Circle Pines, MN: American Guidance Service.
- Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hare, M.G. (2002, September 17). 'Reading Czar' has talk with educators. *The Baltimore Sun*. Retrieved September 21, 2002, from <http://www.sunspot.net>.
- Klare, G.R. (1963). *The measurement of readability*. Ames, IA: Iowa State University Press.
- Lieberman, I.Y., Mann, V.A., Shankweiler, D., & Westelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. *Cortex*, 18, 367-375.
- MetaMetrics, Inc. (2006, August). *Lexile Vocabulary Analyzer: Technical report*. Durham, NC: Author.

- Miller, G.A. & Gildea, P.M. (1987). How children learn words. *Scientific American*, 257, 94-99.
- Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). "Scaling, Norming, and Equating." In R.L. Linn (Ed.), *Educational Measurement* (Third Edition) (pp. 221-262). New York: American Council on Education and Macmillan Publishing Company.
- Poznanski, J.B. (1990). A meta-analytic approach to the estimation of item difficulties. Unpublished doctoral dissertation, Duke University, Durham, NC.
- Salvia, J. & Ysseldyke, J.E. (1998). *Assessment* (Seventh Edition). Boston: Houghton Mifflin Company.
- Shankweiler, D. & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 14, 139-168.
- Stenner, A.J. (1990). Objectivity: Specific and general. *Rasch Measurement Transactions*, 4, 111.
- Stenner, A.J., Burdick, H., Sanford, E.E., & Burdick, D.S. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement*, 7(3), 307-322.
- Stenner, A.J., Smith, M., & Burdick, D.S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement*, 20(4), 305-315.
- Stenner, A.J., Smith, D.R., Horiban, I., & Smith, M. (1987a). Fit of the Lexile Theory to item difficulties on fourteen standardized reading comprehension tests. Durham, NC: MetaMetrics, Inc.
- Stenner, A.J., Smith, D.R., Horiban, I., & Smith, M. (1987b). Fit of the Lexile Theory to sequenced units from eleven basal series. Durham, NC: MetaMetrics, Inc.
- Stenner, A.J., Wright, B.D. & Linacre, J.M. (1994, August). *The Rasch model as a foundation for the Lexile Framework*. Unpublished manuscript.
- Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press

Appendices

Appendix A: The Lexile Framework for Reading Map A-1

The Lexile Framework[®]

F O R R E A D I N G

Text Level	Literature Titles	Benchmarks	Tests/Textbooks
1700L	DISCOURSE ON THE METHOD AND MEDITATIONS ON FIRST		
1690	Concerning Civil Government	To such a class of things pertains corporeal nature in general, and its extension, the figure of extended things, their quantity or magnitude and number, as also the place in which they are, the time which measures their duration, and so on. That is possibly why our reasoning is not unjust when we conclude from this that Physics, Astronomy, Medicine and all other sciences which have as their end the consideration of composite things, are very dubious and uncertain; but that Arithmetic, Geometry and other sciences of that kind which only treat of things that are very simple and very general, without taking great trouble to ascertain whether they are actually existent or not, contain some measure of certainty and an element of the indubitable. (René Descartes, author)	1670 The Principles of Scientific Management; Dover Publications
1680	Critique of Judgment		1630 The American Constitution: Cases, comments, questions, 7th ed.; West Publishing
1660	On Abraham Lincoln		1610 The Condition of Postmodernity; Blackwell Publishers
1660	On the Law Which Has Regulated the Introduction of New Species		
1600L	FUNDAMENTAL PRINCIPLES OF THE METAPHYSICS OF MORALS		
1570	Aeropagiitica	In fact, it is absolutely impossible to make out by experience with complete certainty a single case in which the maxim of an action, however right in itself, rested simply on moral grounds and on the conception of duty. Sometimes it happens that with the sharpest self-examination we can find nothing beside the moral principle of duty which could have been powerful enough to move us to this or that action and to so great a sacrifice; yet we cannot from this infer with certainty that it was not really some secret impulse of self-love, under the false appearance of duty, that was the actual determining cause of the will. (Immanuel Kant, author)	1550 Culture/Power/History: A Reader in Contemporary Social Theory; Princeton University Press
1550	God, Idea of the Ancients		1530 On Injuries of the Head; Project Gutenberg
1530	Plutarch's Lives		1510 On Human Nature; Howard University Press
1520	A Modest Proposal		1500 On Liberty; Hackett Publishing
1510	On Human Nature		1500 The Making of Memory: From Molecules to Mind; Doubleday
1500	The Decameron		
1500L	ON ANCIENT MEDICINE		
1490	Ring of Bright Water	And as to him who had been accustomed to dinner, since, as soon as the body required food, and when the former meal was consumed, and he wanted refreshment, no new supply was furnished to it, he wastes and is consumed from want of food. For all the symptoms which I describe as falling to this man I refer to want of food. And I also say that all men who, when in a state of health, remain for two or three days without food, experience the same unpleasant symptoms as those which I described in the case of him who had omitted to take dinner. (Hippocrates, author)	1450 Philosophical Essays; Hackett Publishing
1470	Utilitarianism		1440 Graduate Management Admission Test
1450	The Confessions of Nat Turner		1430 Certified Public Accountant Examination
1440	The Legend of Sleepy Hollow		1430 Criminal Justice Today; Prentice Hall Science and Education;
1420	Master Humphrey's Clock		1410 The Citadel Press
1410	Profiles in Courage		1400 Test of English as a Foreign Language
1400L	THE SCARLET LETTER		
1380	Life in a Medieval Castle	But the point which drew all eyes, and, as it were, transfigured the wearer—so that both men and women who had been familiarly acquainted with Hester Prynne were now impressed as if they beheld her for the first time—was that SCARLET LETTER, so fantastically embroidered and illuminated upon her bosom. It had the effect of a spell, taking her out of the ordinary relations with humanity, and enclosing her in a sphere by herself. "She hath good skill at her needle, that's certain," remarked one of her female spectators; "but did ever a woman, before this brazen hussy, contrive such a way of showing it? Why, gossips, what is it but to laugh in the faces of our godly magistrates, and make a pride out of what they, worthy gentlemen, meant for a punishment?" (Nathaniel Hawthorne, author)	1390 Graduate Record Examination
1350	The Guns of August		1380 College Board Achievement Test in English
1340	The Hunchback of Notre		1380 Law School Admission Test
1330	Dame		1330 Scholastic Aptitude Test
1320	The Snow Leopard		1320 Medical College Admission Test
1300	The Metamorphosis		Psychology: An Introduction; Prentice Hall
1300L	BROWN v. BOARD OF EDUCATION; 1954		
1280	House of the Spirits	Under that doctrine, equality of treatment is accorded when the races are provided substantially equal facilities, even though these facilities be separate. In the Delaware case, the Supreme Court of Delaware adhered to that doctrine, but ordered that the plaintiffs be admitted to the white schools because of their superiority to the Negro schools. The plaintiffs contend that segregated public schools are not "equal" and cannot be made "equal," and that hence they are deprived of the equal protection of the laws. Because of the obvious importance of the question presented, the Court took jurisdiction. Argument was heard in the 1952 Term, and reargument was heard this Term on certain questions propounded by the Court. (347 US 483, 98 L ed 873, 74 S Ct 686)	1290 Understanding Sociology; Glencoe/McGraw-Hill
1270	Chronicle of a Death Foretold		1290 Speech Science Primer; Williams & Wilkins
1240	The Midwife's Apprentice		1240 Business; Prentice Hall
1240	Dragon Seed		1230 Armed Services Vocational Aptitude Battery
1210	Cold Mountain		1210 ASVAB
1200	The Trumpeter of Krakow		American College Testing Program
1200L	WAR AND PEACE		
1190	Hiroshima	Pierre had been educated abroad, and this reception at Anna Pavlovna's was the first he had attended in Russia. He knew that all the intellectual lights of Petersburg were gathered there and, like a child in a toyshop, did not know which way to look, afraid of missing any clever conversation that was to be heard. Seeing the self-confident and refined expression on the faces of those present he was always expecting to hear something very profound. At last he came up to Morio. Here the conversation seemed interesting and he stood waiting for an opportunity to express his own views, as young people are fond of doing. (Leo Tolstoy, author)	1170 Scholastic Reading Inventory
1160	The Pickwick Papers		1160 SRI-Level 18
1130	The Great Fire		1150 History of a Free Nation; Glencoe/McGraw-Hill
1130	Abigail Adams: Witness to a Revolution		1130
1120	Gutsy Girls: Young Women Who Dare		1100 NAEP Text
1100	Eleanor Roosevelt: A Life of Discovery		1100 Modern Biology; Holt, Reinhart & Winston
1100L	PRIDE AND PREJUDICE		
1090	Amos Fortune, Free Man	Occupied in observing Mr. Bingley's attentions to her sister, Elizabeth was far from suspecting that she was herself becoming an object of some interest in the eyes of his friend. Mr. Darcy had at first scarcely allowed her to be pretty; he had looked at her without admiration at the ball; and when they next met, he looked at her only to criticise. But no sooner had he made it clear to himself and his friends that she had hardly a good feature in her face, than he began to find it was rendered uncommonly intelligent by the beautiful expression of her dark eyes. (Jane Austen, author)	1060 Test of General Educational Development
1070	All Things Bright and Beautiful		1050 GED
1030	Now is Your Time!		1040 Test of Adult Basic Education, General Form
1030	Adam of the Road		1040 Scholastic Reading Inventory
1000	Island of the Blue Dolphins		1020 SRI-Level 17
1000	Parrot in the Oven: Mi Vida		Writing & Grammar: Gold Level; Prentice Hall
1000L	BLACK BEAUTY		
970	Leon's Story	One day, when there was a good deal of kicking, my mother whinnied to me to come to her, and then she said: "I wish you to pay attention to what I am going to say to you. The colts who live here are very good colts, but they are cart-horse colts, and of course they have not learned manners. You have been well-bred and well-born; your father has a great name in these parts, and your grandfather won the cup two years at the Newmarket races; your grandmother had the sweetest temper of any horse I ever knew, and I think you have never seen me kick or bite. I hope you will grow up gentle and good, and never learn bad ways; do your work with a good will, lift your feet up well when you trot, and never bite or kick even in play." (Anna Sewell, author)	990 NAEP Text
960	The Samurai's Tale		950 Scholastic Reading Inventory
950	Bud, Not Buddy		940 SRI-Level 16
940	All the Pretty Horses		930 World Cultures: A Global Mosaic; Prentice Hall
930	The Golden Compass		910 Stanford Achievement Test
920	Talking with Artists		900 Test of Adult Basic Education
900L	TOM SWIFT IN THE LAND OF WONDERS		
880	Her Stories	Just what Tom's thoughts were, Ned, of course, could not guess. But by the flush that showed under the tan of his chum's cheeks the young financial secretary felt pretty certain that Tom was a bit apprehensive of the outcome of Professor Beecher's call on Mary Nestor. "So he is going to see her about 'something important, Ned?' " "That's what some members of his party called it." "And they're waiting here for him to join them?" "Yes. And it means waiting a week for another steamer. It must be something pretty important, don't you think, to cause Beecher to risk that delay in starting after the idol of gold?" "Important? Yes, I suppose so," assented Tom. (Victor Appleton, author)	870 Word 97; Glencoe/McGraw-Hill
870	The View from Saturday		860 Scholastic Reading Inventory
860	Julie of the Wolves		850 SRI-Level 15
820	Maniac Magee		820 Stanford Achievement Test
800	Homeless Bird		810 NAEP Text
800	Scouter		800 Stanford Achievement Test
800L	THE ADVENTURES OF PINOCCHIO		
790	Flour Babies	"Great soul!" said Pinocchio, fondly embracing his friend. Five months passed and the boys continued playing and enjoying themselves from morn till night, without ever seeing a book, or a desk, or a school. But, my children, there came a morning when Pinocchio awoke and found a great surprise awaiting him, a surprise which made him feel very unhappy, as you shall see. Everyone, at one time or another, has found some surprise awaiting him. Of the kind which Pinocchio had on that eventful morning of his life, there are but few. What was it? I will tell you, my dear little readers. On awakening, Pinocchio put his hand up to his head and there he found—Guess! He found that, during the night, his ears had grown at least ten full inches! (Carlo Collodi, author)	780 World Explorer: The U.S. & Canada; Prentice Hall
770	The Giver		770 World Explorer: Latin America; Prentice Hall
760	Walk Two Moons		760 Scholastic Reading Inventory
730	The Apprentice		760 SRI-Level 14
720	Some of the Kinder Planets		730 Stanford Achievement Test
710	The Friends		720 Test of Adult Basic Education
700L	BUNNICULA: A RABBIT TALE OF MYSTERY		
670	The Girl Who Loved Wild Horses	"Of course he bites vegetables. All rabbits bite vegetables." "He bites them, Harold, but he does not eat them. That tomato was all white. What does that mean?" "It means that he paints vegetables?" I ventured. "It means he bites vegetables to make a hole in them, and then he sucks out all the juices." "But what about all the lettuce and carrots that Toby has been feeding him in his cage?" "Ah ha. What indeed!" Chester said. "Look at this!" Whereupon, he stuck his paw under the chair cushion and brought out with a flourish an assortment of strange white objects. Some of them looked like unironed handkerchiefs, and the others well, the others didn't look like anything I'd ever seen before. (Deborah and James Howe, authors) © 1979 by James Howe. Reprinted by permission of Simon & Schuster Children's Publishing Division. All rights reserved.	680 One Nation Many People, Volume One; Globe Fearon
670	The Number Stars		670 Science; Addison-Wesley
660	Holes		660 Understanding Technology; Goodheart-Wilcox
650	The Robber and Me		650 Scholastic Reading Inventory
620	M.C. Higgins, the Great		610 SRI-Level 13
610	Beat the Story-Drum, Pum-Pum		600 Stanford Achievement Test
600L	A BABY SISTER FOR FRANCES		
570	The Whipping Boy	"Did you forget that I like raisins?" "No, I did not forget," said Mother, "but you finished up the raisins yesterday and I have not been out shopping yet." "Well," said Frances, "things are not very good around here anymore. No clothes to wear. No raisins for the oatmeal. I think maybe I'll run away." "Finish your breakfast," said Mother. "It is almost time for the school bus." "What time will dinner be tonight?" said Frances. "Half past six," said Mother. "Then I will have plenty of time to run away after dinner," said Frances, and she kissed her mother good-bye and went to school. After dinner that evening Frances packed her little knapsack very carefully. She put in her tiny special blanket and her alligator doll. (Russell Hoban, author) © 1964 by Russell Hoban. Reprinted by permission of HarperCollins Publishers, Inc. All rights reserved.	550 Communities; Harcourt Brace Jovanovich
560	Sarah, Plain and Tall		540 People and Places; Silver Burdett Ginn
540	The Adventures of Sparrowboy		510 Team Spirit; Scholastic Inc.
530	It's All Greek to Me		510 Scholastic Reading Inventory
520	John Henry: An American Legend		500 Stanford Achievement Test
510	Karen's Chain Letter		9-Primary 2
500L	THE MAGIC SCHOOL BUS INSIDE THE EARTH		
490	Harold and the Purple	But suddenly, the bus began to spin like a top. That sort of thing doesn't happen on most class trips. When the spinning finally stopped, some things had changed. We all had on new clothes. The bus had turned into a steam shovel. And there were shovels and picks for every kid in the class. "Start digging!" yelled Ms. Frizzle. And we began making a huge hole right in the middle of the field. Before long CLUNK we hit rock. The Friz handed out jackhammers. We began to break through the hard rock. "Hey, these rocks have stripes," said a kid. Ms. Frizzle explained that each stripe was a different kind of rock. We chipped off pieces of the rocks for our class rock collection. "These rocks are called sedimentary rocks, class," said Ms. Frizzle. (Joanna Cole, author) THE MAGIC SCHOOL BUS is a registered trademark of Scholastic Inc. © 1987 by Joanna Cole. Reprinted by permission of Scholastic Inc. All rights reserved.	480 Once Upon a Hippo; Scott Foresman
440	Crayon		470 Bears Don't Go to School; Houghton Mifflin
420	All Tutus Should Be Pink		440 Imagine That; Scholastic Inc.
420	Michael Bird-Boy		440 Traveling Star; SRA/McGraw Hill
410	Angel Child, Dragon Child		400 We Are All Alike; Benchmark Education
400	Sam the Minuteman		
400L	FROG AND TOAD ARE FRIENDS		
370	The Drinking Gourd	"That button is thin. My button was thick." Toad put the thin button in his pocket. He was very angry. He jumped up and down and screamed, "The whole world is covered with buttons, and not one of them is mine!" Toad ran home and slammed the door. There, on the floor, he saw his white, four-holed, big, round, thick button. "Oh," said Toad. "It was here all the time. What a lot of trouble I have made for Frog." Toad took all of the buttons out of his pocket. He took his sewing box down from the shelf. Toad sewed the buttons all over his jacket. The next day Toad gave his jacket to Frog. Frog thought it was beautiful. He put it on and jumped for joy. (Arnold Lobel, author) © 1970 by Arnold Lobel. Reprinted by permission of HarperCollins Publishers, Inc. All rights reserved.	390 Discover Science; Scott Foresman
370	A My Name Is Alice		390 Carousels; Houghton Mifflin
370	Owl at Home		360 Scholastic Reading Inventory
360	The Best Way to Play		350 SRI-Level 11
330	Clifford, the Small Red Puppy		340 My World; Harcourt Brace
320	Miss Nelson Is Back		330 Stanford Achievement Test
300L	CLIFFORD'S MANNERS		
290	Sarah's Unicorn	Clifford loves to go visiting. When he visits his sister in the country, he always calls ahead. Clifford always arrives on time. Don't be late. Knock before you walk in. He knocks on the door before he enters. He wipes his feet first. Clifford kisses his sister. He shakes hands with her friend. Shake hands. Wash up before you eat. Clifford's sister has dinner ready. Clifford washes his hands before he eats. Clifford chews his food with his mouth closed. He never talks with his mouth full. Don't talk with your mouth full. Help clean up. Clifford helps with the clean-up. Say good-bye. Then he says thank you and good-bye to his sister and to his friend. Everyone loves Clifford's manners. (Norman Bridwell, author) © 1972 by Norman Bridwell. Reprinted by permission of Scholastic Inc. All rights reserved.	280 Too Big; Houghton Mifflin
270	Baseball Ballerina		270 Test of Adult Basic Education
270	In the Forest		270 Parades; Houghton Mifflin
260	At the Crossroads		250 My Family, Your Family; Silver Burdett Ginn
230	The Boy Who Cried Wolf		240 My Pet Pup; Benchmark Education
220	Play Ball, Amelia Bedelia		
200L			

About The Lexile Framework for Reading

The Lexile Framework[®] for Reading provides a common scale for matching reader ability and text difficulty, allowing easy monitoring of progress. Lexile[®] measures give teachers and parents the confidence to choose materials that will improve student reading skills across the curriculum and at home. A student's Lexile measure is the position on the Lexile scale where a reader can expect to have 75 percent comprehension – difficult enough to be challenging without undue frustration. Other factors (such as interest, age-appropriateness and prior knowledge) are also important when selecting a text. Listed titles are examples only.

Tens of thousands of books and tens of millions of articles have Lexile measures, and all major standardized tests can report student reading scores in Lexiles. As the most widely adopted reading measure in use today, Lexiles are part of reading and testing programs at district, state and federal levels. The Lexile Framework was developed by MetaMetrics[®] Inc., an independent education company based in Durham, N.C., after 15 years of research funded by the National Institutes of Health. For more information about The Lexile Framework for Reading, call 1.888.LEXILES or visit www.Lexile.com.

MetaMetrics, Lexile, the Lexile symbol, Lexile Framework, Lexile Analyzer, Lingos, PowerV and Power Vocabulary are service marks, trademarks or U.S. registered trademarks of MetaMetrics, Inc. The names of other companies and products mentioned herein may be the trademarks of their respective owners. © 2004 MetaMetrics, Inc.



Lexile: Matching readers to text