LIGHTSAIL ASSESSMENT COMPONENTS

Development and Technical Guide

LightSail
READ. LEARN. GROW.

MetaMetrics
LINKING ASSESSMENT WITH INSTRUCTION

Learn about the validity and reliability of LightSail's diagnostic and progress monitoring assessments in this guide created by MetaMetrics®, a leading educational research organization, recognized worldwide for its distinct value in differentiating instruction and personalizing learning. In it, MetaMetrics® reviews psychometric data that describes the accuracy of the assessments that drive LightSail's recommendation algorithm.

This guide is designed to support educators and district administrators with the research needed to understand if LightSail is right for their districts.

LightSail Assessment Components

Development and Technical Guide

Prepared by MetaMetrics, Inc. for LightSail Inc. (www. lightsailed.com) under Contract to LightSail Inc. (Contract dated June 10, 2013, Amendment No. 1 dated June 26, 2015).

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

September 2015

Table of Contents

List of Tables

List of Figures

Introduction

LightSail® software (http://LightSailed.com/) consists of a robust library of digital books across a variety of topics, genres, and levels with in-text embedded assessments providing students and teachers with real-time, actionable data. The premise of the software (LightSail, 2015) is that the solution appeals to students through several key motivators:

- Choice of titles and authors that they actually want to read
- In-app social network allows chat with peers and teachers.
- Digital solution brings schoolwork into the modern era.
- Personalized dashboard encourages achievement with graphics and badges.

A student begins working in the software by taking the Lightsail Power Challenge™ where the result is an initial Lexile measure to describe his or her reading ability. A personalized library is developed for each student and is based on the student's reading level. Books that are contained in a student's personalized library are called "power" texts and have a Lexile text complexity measure within 100L of the student's reading ability. Completion of in-text embedded assessments in the "power" texts produces repeated measures of students' reading abilities as they learn. LightSail employs a Bayesian scoring algorithm within the software to provide continually updated measures that monitor progress in reading development. The Bayesian approach uses prior scores to refine each new estimate of achievement to improve the accuracy of measurement as students learn. In this way, LightSail uses multiple measures over time to improve the assessment of reading ability, which in turn improves the ability to match students with appropriate texts.

During the spring of 2013, LightSail Inc. met with MetaMetrics to discuss ways that an assessment could be developed for use within the LightSail software to assess initial reading level and to monitor reading ability development. The result was the development of an online reading system employing assessment components developed by MetaMetrics, Inc. for use in Grades 1 through 12. Initially, the assessment system consisted of only in-text embedded assessments using a cloze item format (Lexile Cloze Generation Engine). The texts are part of the digital library within the software and consist of a large variety of fiction and nonfiction titles for students in K-12. Students would read and complete an initial "burn-in" phase of in-text embedded assessments prior to the reporting of his or her initial Lexile measure with a Bayesian scoring component. A Bayesian scoring algorithm provides continually updated measures that monitor progress in reading development. During spring 2014, LightSail Inc. and MetaMetrics discussed that a more immediate and precise estimate of each student's reading ability was needed. The result was the development of the Power Challenge that is administered at the beginning of a student's experience with the LightSail software.

The LightSail assessment components are reported on the Lexile® scale, a scientifically based scale of reading ability. The Lexile scale is applied to both readers and texts, making it possible to match readers with texts of appropriate difficulty to facilitate reading improvement. Importantly, the Lexile scale provides accurate feedback on a students' developing reading ability, helping measure progress. All measures within LightSail – Power Challenge scores and

in-text embedded assessment scores – are calculated using the Lexile Analyzer and the Lexile scale developed by MetaMetrics. With these tools, the LightSail software provides accurate information to help students and teachers measure progress in reading development.

This technical guide should provide users with a broad research foundation of the features of the LightSail assessment components. Such a base is essential when deciding if and how the LightSail assessment results should be used and what kinds of inferences about readers are permissible.

Background

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). Partially in response to startling statistics like these, Congress passed the No Child Left Behind (NCLB) Act of 2001, a reauthorization of the Elementary and Secondary Education Act. This act required states to administer annual assessments to all students in grades 3 through 8 by the end of the 2005-2006 school year. Under the legislation, states may select and design tests of their choosing, but the tests must be aligned with the respective state's reading and language arts standards. This legislation requires states to:

1. Create statewide proficiency standards for student achievement in reading and mathematics in grades 3-8.
2. Define these standards according to student performance on statewide outcome assessments.
3. Measure and monitor student progress (aggregated at the school level) toward achieving these proficiency goals, i.e., toward achieving Adequate Yearly Progress (AYP). Student performance is aggregated at the school level and then disaggregated into 11 specific demographic categories specified in the legislation. In order to demonstrate AYP, schools must show that all students are on a trajectory to achieve grade-level proficiency by the end of grade 12.

Schools, districts and states that fail to demonstrate AYP face serious consequences, ranging from school reorganizations and takeovers to a loss of federal funding.

Although many states have made gains in reading achievement since the NCLB Act was passed, nationally, students still have much room for progress, as seen in the 2011 National Assessment of Educational Progress (NAEP) results for reading. At the fourth grade, about two-thirds (67%) of the students performed at or above the *Basic* level, and one-third (34%) performed at or above *Proficient*. Only eight percent performed at the *Advanced* level. At the eighth grade, about 76% of the students performed at or above the *Basic* level, about one third (34%) performed at or above *Proficient*, and just 3% performed at the *Advanced* level (National Center for Education Statistics, 2011).

In June of 2010, the National Governors Association Center for Best Practices and the Council of Chief State School Officers (CCSSO) released the Common Core State Standards (CCSS). These standards, developed for K-12 in English language arts and mathematics, establish clear goals for learning intended to prepare students for success in college and work. The English language arts standards outline challenging goals for student reading and provide guidance regarding the proportions of literary and informational texts students should read. These standards explicitly describe literacy as part of students' educational programs across the content areas, including history/social studies, science, and technical subjects. The CCSS also challenge educators to provide reading materials at a level of complexity necessary to prepare adequately students for college and career success (Standard 10). The Lexile measure is provided as a measure of text difficulty, and Appendix A of the CCSS provides Lexile measures for reading ability targets in Grades 2-12.

Research has shown clearly that there is a positive correlation between reading proficiency and the amount of reading students engage in throughout their schooling years (Cunningham & Stanovich, 1998; O'Connor, Swanson & Geraghty, 2010; O'Connor, Bell, Harty, Larkin, Sackor & Zigmond, 2002; Cain, Oakhill & Lemmon, 2004; Jenkins, Stein & Wysocki, 1984). When students are provided with materials that are appropriate for their reading proficiency level, they exhibit higher levels of understanding of what they read, and when they comprehend what they read, students may learn more. Thus, the more students read, the more likely they are to develop into strong readers. Studies investigating summer reading loss have shown that when students are provided with books at their reading level and interest areas, their gains in reading were comparable to gains one would expect in summer school (Kim, 2006). Since motivation is key to voluntary reading, two critical features of book selection are interest and reading level, and both were addressed in Kim's study. Kim demonstrated in a randomized field study that low-income students are not destined to summer loss; but rather, showed that low-income students' skills could, in fact, grow over the summer if they were able to select books at their interest level and reading level. Kim used The Lexile Framework for Reading – a tool that many states use to make sure that students are appropriately challenge – to match students with books at an appropriate complexity (difficulty) level.

LightSail Inc. has developed this assessment system to address the need for students to read often and read material at the right complexity level. The assessment components of the LightSail software help to personalize the reading experience for students and provide valid and reliable indicators of student reading ability. With up-to-date information about their students' reading ability, instructors can better prepare students to be successful readers.

Features of LightSail

LightSail assessment components are research-based, scientifically valid, and reliable. Several specific features of LightSail assessment components are noteworthy.

- Reading materials are authentic: they are bestselling fiction and nonfiction for students in K-12

- The native-Lexile, two-sentence, and one-sentence item formats used on the LightSail Power Challenge are extensions of the "embedded completion" item format that has 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).

- LightSail assessment components are linked with the Lexile scale and, as such, the item and passage calibrations used to convert a raw score (number correct) into the Lexile metric are provided by the Lexile Theory. The calibration equation used to calibrate LightSail 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 embedded completion items. These rules were precisely followed in developing the Lexile Reading Ability Item Bank items. A multi-stage review process was used to ensure conformance with the item writing specifications and appropriateness for use with students in Grades 2 through 12.

- Assessment items used on the LightSail Power Challenge were selected from the Lexile Reading Ability Item Bank, a proprietary set of items developed by MetaMetrics to meet the guidelines of the Lexile Theory. The performance of these items is informed by the Lexile Theory and, for a portion of the items, data from field administrations is included. Any items that perform poorly are revised or rejected from the item bank.

- The LightSail software assessment components are administered individually online, scored immediately and objectively, and results are used to help guide reading selections for future instruction.

- The online test administration format supports quick administration in an untimed, low-pressure format.

- No extensive or specialized preparation is needed to administer the LightSail assessment components, although proper interpretation and use of the results requires an understanding of The Lexile Framework for Reading.

- The LightSail assessment components use a Bayesian scoring algorithm, which incorporates past performance to predict future performance. Bayesian methodology provides a paradigm for combining prior information with current data, both subject to uncertainty, to calculate an estimate of current status, which is again subject to uncertainty. This methodology connects the administration of each assessment, regardless of type (Power Challenge of in-text embedded assessments), and thus produces more precise measurements when compared with independent assessments.

Using The Lexile Framework for Reading

Teachers, parents, administrators, and students can use the tools provided by the Lexile Framework to plan instruction. When students' Lexile measures are known, teachers, parents and students can work together to choose appropriately challenging texts that also match the students' interests and background knowledge. The Lexile Framework does not prescribe a reading program; it is a tool that 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 they will find challenging but not frustrating.

The Lexile Framework is a system that helps match readers with literature appropriate for their reading skills. When reading a book within his or her Lexile range (50L above his or her Lexile measure to 100L below), the reader should comprehend enough of the text to make sense of it, while still being challenged enough to maintain interest and learning.

There are many factors that affect the relationship between a reader and a text. These factors include 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 complexity (difficulty), is a good starting point in the selection process with other factors then being considered. The Lexile measure should never be the only factor considered when selecting a text.

Purposes and Uses of the LightSail Assessment Components

The LightSail assessment components are designed to measure a reader's ability to comprehend texts of increasing difficulty. The results of the LightSail assessment components can be used to target students' reading materials at an appropriate level of complexity and to serve as a tool for measuring reading growth.

One outcome of the LightSail assessment components is the location of the reader on the Lexile Map (Appendix A). Once a reader is measured, it is possible to forecast how well the reader will likely 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 compare directly a reader and text. When reader and text measures match, the Lexile Framework forecasts 75% comprehension for independent reading. 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%.

In addition to helping to personalize the reading experience for students, the data provided by the LightSail assessment components can help educators make better-informed decisions about materials selection, particularly in cases where differentiated instruction is the goal. Furthermore, LightSail assessment component results provide valuable information for teachers whose students who need extra attention in reading, such as students requiring an Individualized Educational Program (IEP) or students who are classified as English as a Second Language (ESL).

Development Groups

LightSail Inc. provided the vision of the software and collaborated with MetaMetrics on the development of the assessment components.

MetaMetrics managed the overall development of the program's assessments. MetaMetrics designed the Power Challenge, selected the passages and test items from the Lexile Reading Ability Item Bank, coordinated the test development, and designed the scoring and reporting algorithms. MetaMetrics licensed to LightSail Inc. the Lexile Cloze Generation Engine (a computer program that analyzes the difficulty of text and generates cloze items), a Bayesian scoring application to score the tests, and a forecasting application to aid in the identification of students most in need of reading intervention and who may be at risk of performing below proficiency on a state summative assessment.

LightSail Inc. developed the software system and managed the development of in-text embedded assessments using the Lexile Cloze Generation Engine. LightSail approved final passage selection and item sets for the Power Challenge forms and implemented the scoring and reporting algorithms.

Limitations of the LightSail Assessment Components

A well-targeted assessment can provide useful information for matching texts and readers. As with any other assessments, results from the LightSail assessment software are just one source of evidence about a reader's level of comprehension. Obviously, decisions are best made when using multiple sources of evidence about a reader. Other sources include other reading test data, reading group placement, lists of books read, and, most importantly, teacher judgment. *One measure of reader performance, taken on one day, is not sufficient to make high-stakes student-level decisions such as summer school placement or retention.*

The Lexile Framework for Reading provides a common metric for combining different sources of information about a reader into a best overall judgment of the reader's ability expressed in the Lexile metric. LightSail Inc. encourages users to employ multiple measures when deciding where to locate a reader on the Lexile scale.

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: complexity, 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 features: reader ability and text complexity.

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 utilizes these two dominant features of language in measuring text complexity by examining the characteristics of word frequency and sentence length. Lexile text measures typically range from above 200L to below 1600L but measures can be below 0L for emergent reading texts ("BR" for "Beginning Reader") and above 1800L for advanced texts. Within any one classroom, there will be a range of reading materials to reflect the student range of reading ability 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 currently employs a 600-million-word corpus when examining the semantic component of text. This corpus was assembled from the more than 15,000 texts that were measured by MetaMetrics for publishers from 1998 through 2002. When text is analyzed by MetaMetrics, all electronic files are initially edited according to established guidelines used with the Lexile Analyzer software. These guidelines include the removal of all incomplete sentences, chapter titles, and paragraph headings; running of a spell check; and re-punctuating where necessary to correspond to how the book would be read by a child (for example, at the end of a page). The text is then submitted to the Lexile Analyzer that examines the lengths of the sentences and the frequencies of the words and reports a Lexile measure for the book. When enough additional texts have been analyzed to make an adjustment to the corpus necessary and desirable, a linking study will be conducted to adjust the calibration equation such that the Lexile measure of a text based on the current corpus will be equivalent to the Lexile measure based on the new corpus.

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, sentence length was selected as a proxy for the syntactic component of reading complexity in the Lexile Framework.

Calibration of Text Complexity

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 complexity of text, 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, most thermometers have anchor points at 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 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 $1/100^{th}$ of the difference between freezing (0 degrees) and boiling (100 degrees) water. For the Lexile scale the unit size was defined as $1/1000^{th}$ of the difference between the mean difficulty of the primer material and the mean difficulty of the encyclopedia samples. Therefore, a Lexile unit by definition equals $1/1000^{th}$ 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

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

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.

The Lexile Scale ranges from below 200L to above 1600L. There is a not an explicit bottom or top to the scale, but rather two anchor points on the scale (described above) that describe different levels of reading comprehension. The Lexile Map, a graphic representation of the Lexile Scale from 200L to 1600L, provides a context for understanding reading comprehension.

Validity Evidence for The Lexile Framework for Reading

The 2014 Standards for Educational and Psychological Testing (America 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 for proposed uses of tests" (p. 11). 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 complexity 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 and reading comprehension.

Lexile Framework Linked to other Measures of Reading Comprehension. The Lexile Framework for Reading has been linked to numerous 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: the 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 consisting of texts targeted to his specific reading level.

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

Test	Grades in Study	N	Correlation Between Test Score and Lexile measure
TerraNova Assessment Series (CTBS/5)	2, 4, 6, 8	2,713	0.92
Gates-MacGinitie Reading Test	2, 4, 6, 8, 10	4,644	0.90
Texas Assessment of Knowledge and Skills (TAKS)	3, 5, 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	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*
Comprehensive Testing Program (CPT 4 – ERB)	2, 4, 6, and 8	924	0.83 to 0.88
Oklahoma Core Competency Tests (OCCT)	3 – 8	10,691	0.71 to 0.75*
TOEFL iBT	NA	2,906	0.63 to 0.67
TOEIC	NA	2,799	0.73 to 0.74
Kentucky Performance Rating for Educational Progress (K-PREP)	3 – 8	6,480	0.71 to 0.79*
North Carolina ACT	11	3,472	0.84
North Carolina READY End-of-Grades/End-of-Course Tests (NC READY EOG/EOC)	3, 5, 7, 8, and E2	12,356	0.88 to 0.89

Notes: Results are based on final samples used with each linking study.
*Not vertically equated; separate linking equations were derived for each grade.

Lexile Framework and the Difficulty of Basal Readers. Lexile measures are organized in a sequential manner, so a lower Lexile measure for a text means that the text is less complex than text 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 Lexile measures. 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 complexity. So, for example, the latter portion of a third-grade reader is presumably more complex than the first portion of the same book. Likewise, a fourth-grade reader is presumed to be more complex 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 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*).

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.

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.

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 (Poznanski, 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 and 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 Question	Number of Passage	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 correlations 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 $\bar{Z}$ mean disattenuated correlation between theory-based calibration and empirical difficulty 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 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 Lexile measures 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 item response theory (IRT) model equation. 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.

Although the exact relationship between comprehension rate and the pattern of slice calibrations is complicated, Equation 2 is an unbiased approximation.

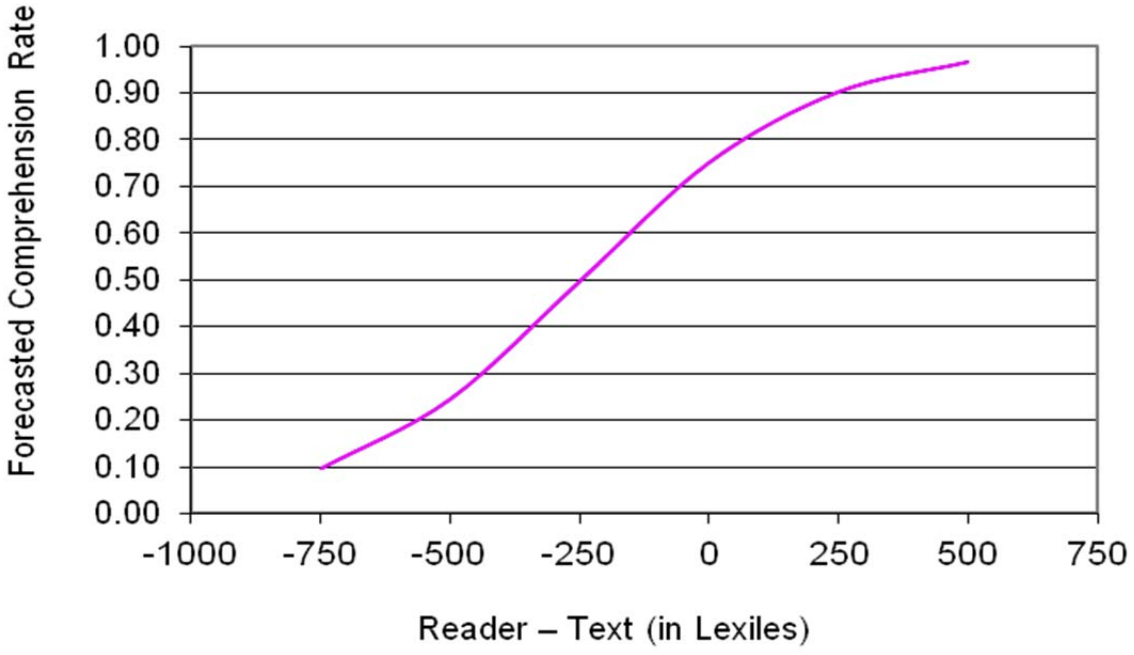
$$Rate = \frac{e^{ELD+1.1}}{1+e^{ELD+1.1}} \qquad \text{(Equation 2)}$$

where ELD is the "effective logit difference" given by

$$\text{ELD} = (\text{Reader Lexile measure} - \text{Text Lexile measure}) \div 225. \qquad \text{(Equation 3)}$$

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	Tornado (Byars)	96%
1000L	750L	The Martian Chronicles (Bradbury)	90%
1000L	1000L	Reader's Digest	75%
1000L	1250L	The Call of the Wild *(London)*	50%
1000L	1500L	On the Equality Among Mankind (Rousseau)	25%

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

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

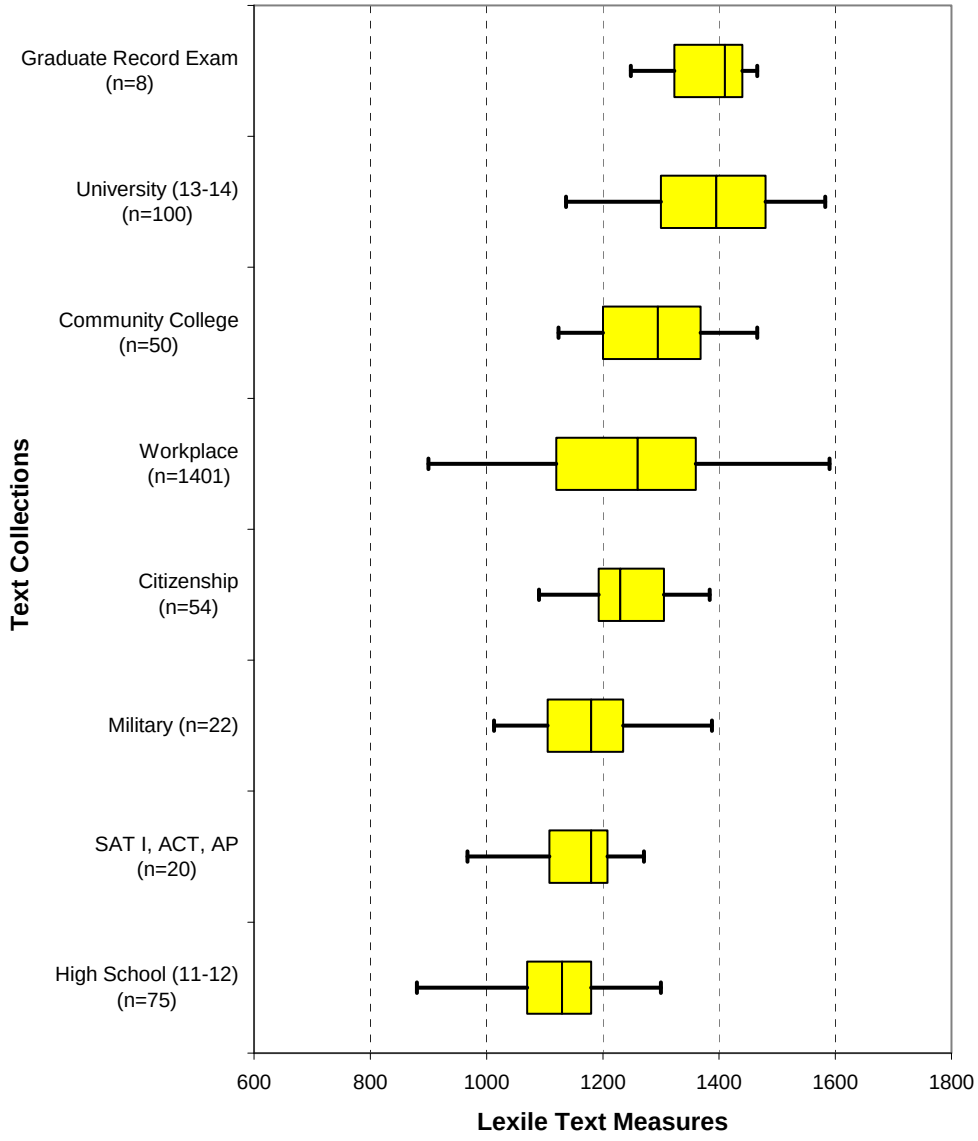
The subjective experience of 50%, 75%, and 90% comprehension as reported by readers varies greatly. A 1000L reader reading 1000L text (75% comprehension) reports confidence and competence. Teachers listening to such a reader report that the reader can sustain the meaning thread of the text and can read 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 Lexile 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.

College and Career Readiness and Text Complexity

There is increasing recognition of the importance of bridging the gap that exists between K-12 and higher education and other postsecondary endeavors. Many state and policy leaders have formed task forces and policy committees such as P-20 councils. In the *Journal of Advanced Academics* (2008), Williamson investigated the gap between high school textbooks and various reading materials across several postsecondary domains. The resources Williamson used were organized into four domains that correspond to the three major postsecondary endeavors that students can choose—further education, the workplace or the military, and, the broad area of citizenship, which cuts across all postsecondary endeavors. Williamson discovered a substantial increase in reading expectations and text complexity from high school to postsecondary domains—"a gap large enough to help account for high remediation rates and disheartening graduation statistics" (Smith, 2011). *Figure 2* illustrates this continuum of text difficulty.

Figure 2. A continuum of text difficulty for the transition from high school to postsecondary experiences (box plot percentiles. 5th, 25th, 50th, 75th, and 95th).[1]



Expanding on Williamson's work, Stenner, Sanford-Moore, and Williamson (2012) aggregated readability information across the various postsecondary options available to a high school graduate to arrive at a standard of reading needed by individuals to be considered "college and career ready." In their study, they included additional citizenship materials beyond those examined by Williamson (e.g., national and international newspapers and other adult reading materials such as Wikipedia articles). Using a weighted mean of the medians for each of the postsecondary options (education, military, work place, and citizenship), a measure of 1300L was defined as the general reading demand for postsecondary options and could be used to judge a student's "college and career readiness."

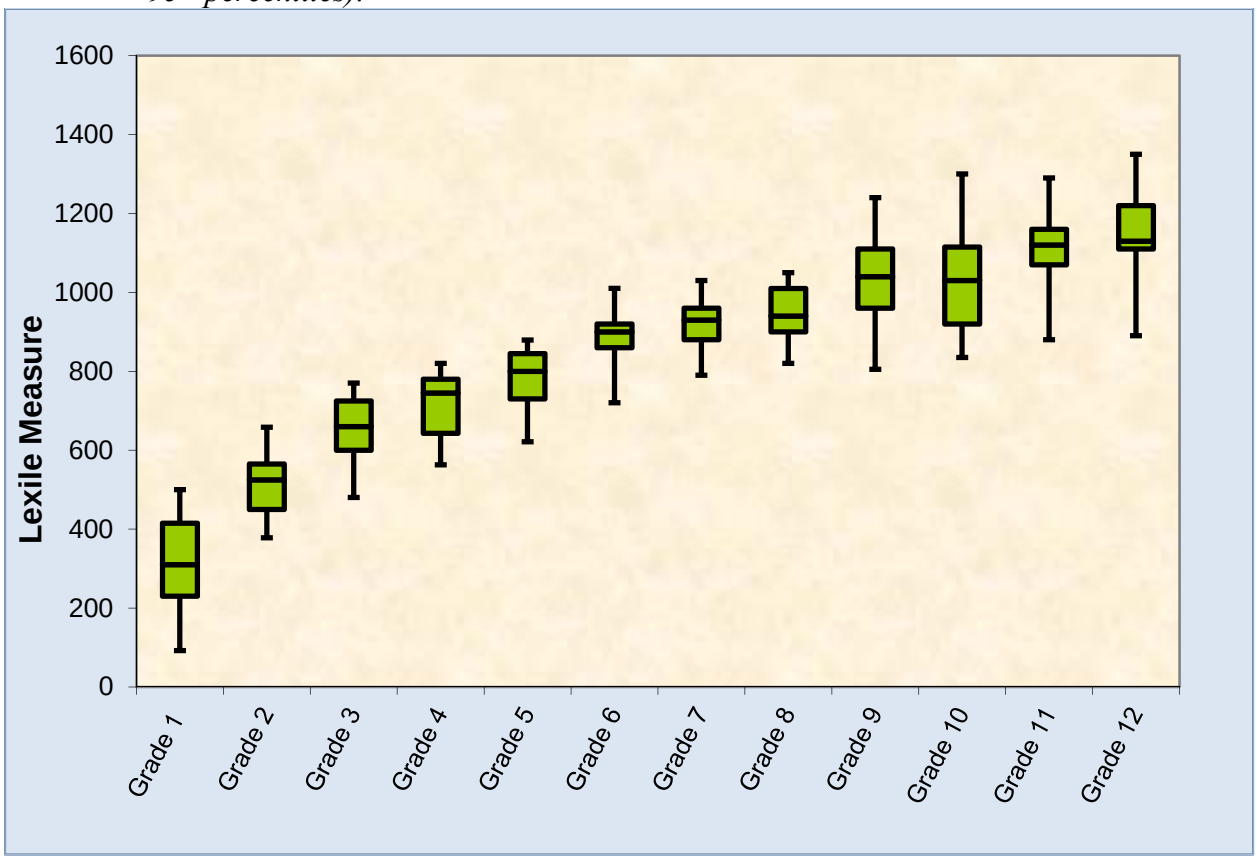
[1] Reprinted from Williamson, G. L. (2008). A text readability continuum for postsecondary readiness. *Journal of Advanced Academics, 19*(4), 602-632.

In Texas, two studies were conducted to examine the reading demands in various postsecondary options – technical college, community college, and 4-year university programs. Under Commissioner Raymond Paredes, the Texas Higher Education Coordinating Board (THECB) conducted a research study in 2007 (and extended in 2008) which addressed the focal question of "how well does a student need to read to be successful in community colleges, technical colleges, and universities in Texas?" THECB staff collected a sample of books that first year students in Texas would be required to read in each setting. These books were measured in terms of their text complexity using The Lexile Framework for Reading. Since the TAKS had already been linked with Lexile measures for several years, the THECB study was able to overlay the TAKS cut scores onto the post high school reading requirements (MetaMetrics, 2008).

After the THECB study was completed, other states have followed the Texas example and used the same approach in examining the gap from high school to the postsecondary world. In 2009, a similar study was conducted for the Georgia Department of Education; and in 2010, a study was conducted for the Tennessee Department of Education. In terms of mean text demand, the results across the three states produced similar estimates of the reading ability needed in higher-education institutions: Texas, 1230L; Georgia, 1220L; and Tennessee, 1260L. When these results are incorporated with the reading demands of other postsecondary endeavors (military, citizenship, workplace, and adult reading materials [national and international newspapers] and Wikipedia articles) used by Stenner, Koons, and Swartz (2010), the college and career readiness standard for reading is 1293L. These results are based on more than 105,000,000 words from approximately 3,100 sources from the adult text space.

Between 2004 and 2008, MetaMetrics (Williamson, Koons, Sandvik, and Sanford-Moore, 2012) collected and measured textbooks across the K-12 educational continuum. The box-and-whisker plot in *Figure 3* shows the Lexile measures (*y*-axis) across grades as defined in the US. For each grade, the box refers to the interquartile range. The line within the box indicates the median. The end of each whisker shows the 5th and 95th percentile text complexity measures in the Lexile metric for each grade. This information can provide a basis for defining at what level students need to be able to read to be ready for various postsecondary endeavors such as further education beyond high school and entering the work force.

Figure 3. Text complexity distributions, in Lexile units, by grade (whiskers represent 5th and 95th percentiles).



This continuum can be "stretched" to describe the reading demands expected of students in Grades 1-12 who are "on track" for college and career (Sanford-Moore and Williamson, 2012). The quantitative aspect of defining text complexity consists of a stair-step progression of increasingly difficult text by grade levels (Common Core State Standards for English Language Arts, Appendix A, NGA Center and CCSSO, 2010, p. 8).

The question for educators becomes how to determine if a student is "on track" for college and career as previously defined in the Common Core State Standards and described above. "As state departments of education, and the districts and schools within those respective states, transition from adopting the new Common Core State Standards to the more difficult task of implementing them, the challenge now becomes how to translate these higher standards into tangible, practical and cost-effective curricula" (Smith, 2012). Implementing the Common Core will require districts and schools to develop new instructional strategies and complementary resources that are not only aligned with these national college- and career-readiness standards, but also utilize and incorporate proven and cost-effective tools that are universally accessible to all stakeholders. The Standards for English Language Arts focus on the importance of text complexity. As stated in Standard 10, students must be able to "read and comprehend complex literary and informational texts independently and proficiently" (Common Core State Standards for English

Language Arts, College and Career Readiness Anchor Standards for Reading, NGA Center and CCSSO, 2010, p.10).

The Common Core State Standards recommend a three-part model for evaluating the complexity of a text that takes into account its qualitative dimensions, quantitative measure, and reader and task considerations. It describes text complexity as "the inherent difficulty of reading and comprehending a text combined with consideration of reader and task variables … a three-part assessment of text [complexity] that pairs qualitative and quantitative measures with reader-task considerations" (NGA Center and CCSSO, 2010, p. 43). In simpler terms, text complexity is a transaction between text, reader, and task. The quantitative aspect of defining text complexity consists of a stair-step progression of increasingly difficult text by grade levels (Common Core State Standards for English Language Arts, Appendix A, NGA Center and CCSSO, 2010, p. 8).

MetaMetrics' research on the typical reading demands of college and careers contributed to the Common Core State Standards as a whole and, more specifically, to the Lexile-based grade bands in *Table 6*.

Table 6. Text complexity standards describing "on track" for college and career reading levels (expansion of CCSS grade).

Grade	Lexile Text Ranges to Guide Reading for College and Career Readiness
2	420L to 650L
3	520L to 820L
4	740L to 940L
5	830L to 1010L
6	925L to 1070L
7	970L to 1120L
8	1010L to 1185L
9	1050L to 1260L
10	1080L to 1335L
11-12	1185L to 1385L

Description of the LightSail Assessment Components

The LightSail software and assessment components are built upon research showing that when students read text at their reading levels, they experience optimal reading comprehension for learning (Crawford, 1978; Guthrie and Davis, 2003; Jalongo, 2007). In addition, students who are better readers are also higher achievers and engage in life-long learning in relation to careers (Crawford, 1978; Kirsch, I., de Jong, J., LaFontaine, D., McQueen, J., Mendelovits, J., and Monseur, C, 2002). In a review of prior studies, Squires and his colleagues (1983) found 75% to be the optimal student success rate for learning. They noted that a reanalysis of the Fischer (Denham and Lieberman, 1980) data by Rim showed that reading achievement by grade 2 students increased up to a 75% success rate and then began to decrease. O'Connor, Swanson, and Geraghty (2010) randomly assigned 123 students in grades 2 and 4 to three different conditions for the difficulty level of reading materials: the grade-appropriate condition, the 'difficult' condition, and a control group. Participants were assessed using a pre-test to measure comprehension and fluency, then given a 20-week intervention course to evaluate comprehension growth over time based on passage difficulty level. Finally, a post-test was administered to determine growth differences between the groups. With respect to both the pre-test and post-test performance, the differences between level and comprehension were found to be significant, where performance was highest for the grade-appropriate condition and lowest for the 'difficult' condition. The results also indicated that there were also significant gains over time for students reading material at their appropriate reading level. The research suggests that students should be given reading level materials that match their comprehension goals.

Similarly, research by O'Connor, Bell, Harty, Larkin, Sackor, and Zigmond (2002) investigated the role of text difficulty on reading ability for students who experienced difficulty with reading. The researchers compared the influence of text difficulty on reading ability growth over an 18-week period for 46 struggling readers who were engaged in one-on-one tutoring. Students were randomly assigned to either receive texts matched to their reading level or matched to their grade level. Three reading tests were used to estimate reading proficiency: the *Peabody Picture Vocabulary Test- 3rd Edition* (PPVT3), the *Woodcock Reading Mastery Tests-Revised (WRMT-R*), and the *Analytic Reading Inventory (ARI)*. These tests were used in a pre-post research design. When groups were compared, students who received texts matched to their reading level made greater learning gains (evidenced by performance on several measures including three subtests of the *Woodcock Reading Mastery Tests-Revised*) as compared to those who received grade-level matched texts.

The LightSail software consists of a Power Challenge for each grade or grade band, 1 through 11-12 (Grades 1, 2, 3, 4, 5, 6, 7-8, 9-10, 11-12). Additionally, the LightSail software includes in-text embedded assessments that can be used to monitor reading ability and update the students' Lexile measures.

Upon entry into the program, a new user will be administered the Power Challenge at the appropriate grade level and will receive a Lexile measure based on the test results; in-text embedded assessments for the student can then be targeted based on the student's Lexile measure (included in "power" texts).

LightSail Power Challenge

The LightSail Power Challenge includes a total of 11 test levels with one test per grade or grade-range level (2 through 11-12). The tests are untimed, but each is designed to take about 35 minutes for a student to complete. The items on the Power Challenge are composed of informational (nonfiction) and narrative (fiction) passages. Each Power Challenge form consists of 32 multiple-choice items as shown in *Table 7*. (A description of item types is provided later in this technical manual in the section entitled *Development of LightSail Power Challenge*.)

Table 7. LightSail Power Challenge item types by grade.

	Grade 1	Grade 2	Grades 3 through 11-12
Picture Items	5	-	-
One-Sentence Items	9	5	-
Two-Sentence Items	9	9	-
Regular Native Items	9	18	32
Total Number of Items	32	32	32

Student results are reported as a Lexile measure. There are many reasons to use scale scores, in this case Lexile measures, 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 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 875 on one form of a test represents a similar level of reading ability as receiving a scale score (Lexile measure) of 875 on another form of the test.

The typical range of the Lexile Scale is from below 200L to above 1600L. There is a not an explicit bottom or top to the scale, but rather two anchor points on the scale that describe different levels of reading comprehension. The Lexile Map, a graphic representation of the Lexile Scale from 200L to 1500L+, provides a context for understanding reading comprehension (see Appendix A). Lexile reader measures are reported in 5-unit intervals. Scores below 0L are reported as BRxxxL (Beginning Reader).

LightSail In-text Embedded Assessments

For students using LightSail software, reading ability is continuously monitored through online "power" texts. Power texts are suggested books from the extensive digital library within the LightSail software that are within 100L of the student's reading ability. In these books, students complete in-text embedded assessments as they read. The in-text embedded assessments consist of up to two cloze items presented on a page of text. Typically, every two to three pages in the book include cloze items.

LightSail Assessment Components Sequence

When students sign in to LightSail software for the first time, they are administered a grade-specific Power Challenge that determines their initial Lexile measure. As part of the administration, "step-down" logic is employed when a student misses a significant number of items at the beginning of the test (i.e., in first five items and in first ten items) and "step-up" logic is employed when student completes almost all of the items correct on the test. This logic directs the student to a lower-level Power Challenge or a higher-level Power Challenge that is more targeted to his or her ability. The resulting Lexile measure determines the Lexile level of the "power" texts selected for the student with in-text embedded assessments.

When the student is administered an in-text embedded assessment, the student's prior information (i.e., previous test results) is incorporated into the Lexile Scoring (Bayesian scoring) algorithm and a new Lexile measure and a new estimate of uncertainty for the student is produced. This data is entered into the LightSail software to allow the program to continue to offer targeted text selections to the student.

Interpreting and Using LightSail Assessment Component Results

The Lexile Framework for Reading provides teachers and educators with tools to help them link assessment results with subsequent instruction. Assessments such as the ones in the LightSail software that are linked to the Lexile scale provide tools for monitoring the progress of students at any time during the course of instruction.

When a reader takes the LightSail Power Challenge or completes an in-text embedded assessment, his or her results are reported as a Lexile measure. 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. When the reader and text are matched (same Lexile measures), the reader is "targeted." 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 experiences 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 experiences 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 should comprehend enough of the text to make sense of it, while still being challenged enough to maintain interest and learning.

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

The Lexile Framework reporting scale is not bounded by grade level, although typical Lexile measure ranges have been identified for students in specific grades. Because the Lexile Framework reporting scale is not bounded by grade level, it makes provisions for students who read below or beyond their grade level. See the Lexile Framework Map for literary and informational titles, leveled reading samples, and approximate grade ranges (Appendix A).

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 200,000 books, 60 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.

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 assessment systems such as the LightSail software components that are specifically developed to match the Lexile Framework levels 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.

Suggestions for Using The Lexile Framework for Reading

Use the Lexile Framework to Select Books. Teachers, parents, and students can use the tools provided by the Lexile Framework to select materials 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 knowledge of the variables involved when they design reading instruction.* The Lexile Framework facilitates 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.

Many factors 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 complexity, 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 students 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 tests throughout the year can help to monitor students’ progress toward their goals.

Monitor Reading Program Goals. As a student's Lexile measure increases, the set of reading materials he can likely 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 or program 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: S*tudents' 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 should be ready to read with at least 75% comprehension these kinds of materials which are at the next grade level." Or, "Your child will need to increase his/her Lexile measure by 400L-500L 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. Fluency is highly correlated to comprehension (Fuchs, Fuchs, Hops, & Jenkins, 2001; Rasinski, 2009). 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 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 adjust text targeting 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 likely be challenging for a group of readers, they can systematically plan instruction that will allow students to encounter difficult text in a controlled fashion and make use of instructional scaffolding to build student success and confidence with more challenging text. The teacher can model appropriate learning strategies for students, such as rereading or rephrasing text in one's own words, so that students can then learn what to do when comprehension breaks down. Students can then practice these metacognitive strategies on selected text while the teacher monitors their progress.

Teachers can use Lexile measures to guide a struggling student toward texts at the lower end of the student's Lexile range (100L above 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.

Differentiating instruction for the reading experience also involves 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.

Targeting Instruction to Students' Abilities. To encourage optimal progress with the use of any reading materials, teachers need to be aware of the complexity 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 and texts containing new content, 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 or the content of such texts can be challenged with more difficult reading levels, which will reduce 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.

Apply Lexile measures Across the Curriculum. Over 450 publishers provide Lexile measures for their trade books and textbooks, enabling educators to make connections among all of the different components of the curriculum to plan instruction more effectively. With a student's Lexile measure, teachers can connect him or her to hundreds of thousands of books. Using periodical databases, teachers and students can also find appropriately challenging newspaper and magazine articles that have Lexile measures.

Using the Lexile Framework in the Classroom

- Develop individualized reading lists that are tailored to provide appropriately challenging reading while still reflecting student interest and motivations.
- Build text sets that include texts at varying levels to enhance thematic teaching. These texts might not only support the theme, but also provide a way for all students to successfully learn about and participate in discussions about the theme, building knowledge of common content for the class while building the reading skills of individual students. Such discussions can provide important collaborative brainstorming opportunities to fuel student writing and synthesize the curriculum.
- Sequence materials in a reading program to encourage growth in reading ability. For example, an educator might choose one article a week for use as a read-aloud. In addition

to considering the topic, the educator could increase the complexity 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 in which the readings in anthologies are presented to the students 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 texts within the student's Lexile range, reports of recent assessments, and a form to record reading that occurs at home. This is an important opportunity to encourage individualized goal setting and engage families in monitoring the progress of students in reaching those goals.
- Choose texts lower in the student's Lexile range when factors make the reading situation more challenging 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 to provide all students with exposure to differentiated, challenging text at least once every two to three weeks as suggested by the lead authors of the Common Core State Standards.
- Use the free Find a Book website (at www.lexile.com/fab) to support book selection and create booklists within a student's Lexile range to help the student make more informed choices when selecting texts.
- Use database resources to infuse research into the curricula while tailoring reading selections to specific Lexile levels. In this way, students can explore new content at an appropriate reading level and then demonstrate their assimilation of that content through writing and/or presentations. A list of the database service providers that have their collections measured can be found at www.lexile.com/using-lexile/lexile-at-library.

Using the Lexile Framework 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/using-lexile/lexile-at-library/)
- Use the free Find a Book website (at www.lexile.com/fab) to support book selection and help students make informed choices when selecting texts.

Using the Lexile Framework 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 Find a Book website (at www.lexile.com/fab).

- 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 LightSail Assessment Components and Grade Levels. Lexile measures do not translate specifically to grade levels. Within any grade, there will be a range of readers and a range of materials to be read. In a fifth-grade classroom there will be some readers who are far ahead of the others and there will be some readers who are behind the others in terms of reading ability. To say that some books are "just right" for fifth graders assumes that all fifth 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 would 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%, 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.

The Lexile Reading Ability Item Bank

The Lexile Reading Ability Item Bank (LRAIB) is a proprietary collection of original MetaMetrics' assessment items available for licensing by partners such as LightSail Inc. The items in this bank have been written to assess a large range of reading ability. Experienced MetaMetrics Editorial, Content, and Research staff have developed the items, including passages and stems, adhering to clear guidelines described below. The items have been reviewed throughout the process to ensure the highest quality possible and, where possible, field tested to empirically examine the difficulty of the items.

When licensing items from the LRAIB, test publishers work with MetaMetrics to determine the appropriate specifications of the each test based on student population, programming focus, and other factors. Then the tests are created by selecting the best items to fit the specifications.

LRAIB Item Development

The LRAIB includes four types of items: native items, one- and two-sentence items, and picture items. The components of the items and their descriptions are included below.

- *Passage.* The passage is the ancillary text for which an item is written. For most items, the Lexile measure of the passage is considered the Lexile measure of the item. For picture items, the image is considered the text or passage and the Lexile measure of the item is calculated empirically. Each passage is used for only one item.

- *Stem.* The stem is the question or embedded completion statement. For embedded completion statements, they should appear as if they were written as part of the passage. The statement portion of the embedded 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.

- *Correct Answer.* The correct answer is the correct response. The correct answer (key) typically has a Lexile measure similar to the measure of the passage.

- *Distractor(s).* The distractors are the three wrong responses that are semantically and syntactically correct. These should be attractive responses if the reader has not read the passage. The distractors have similar Lexile measures as the correct answer.

All of the items, including their associated passages, in the LRAIB are either written by staff at MetaMetrics or commissioned and then reviewed by staff at MetaMetrics. MetaMetrics staff who are experienced in item development and who have experience with the everyday reading

ability of students at various levels intensively review and revise each component of the items at each stage of development.

Each reading passage was analyzed using the Lexile Analyzer to determine its Lexile measure. Item writers are given general guidelines for passage length to help ensure that the overall length of each test was uniform and that the reading demand of each test form allowed administration within a single class period.

The answer choices come from a word list created by MetaMetrics that includes Lexile ranges for words. The level of the answer choices corresponds to the Lexile measure of the passage. Item writers are provided with vocabulary lists to use during item development. The vocabulary lists have been compiled by MetaMetrics based on research to determine Lexile word measures (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 if the words had likely been encountered in easier texts (those with lower Lexile measures), then those words should be a part of a reader's "working" vocabulary.

Native Items

The native-Lexile item is the primary item type used in the development of the Lexile Framework. Its format consists of a passage of text (maximum of 125 words) followed by the stem written by the item author. The stem consists of an embedded completion statement, a correct answer, and three distractors. The embedded completion statement is similar to the fill-in-the-blank format and should assess the student's ability to form a generalization based on the passage or draw an inference from the passage. From the four answer choices, the reader is asked to select the "best" word that completes the statement. The statement is written to ensure that by reading and comprehending the passage the reader is able to select the correct option. When the native-Lexile completion statement is read by itself, each of the four options is plausible. Items are written so that the correct response is not stated directly in the passage and is not suggested by the item itself. Rather, the examinee must determine the correct answer through comprehension of the passage.

When properly written, this native-Lexile item format 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.

There are two main advantages to using the native-Lexile item format. 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. *Figure 4* provides sample items across a range of difficulty.

Figure 4. Sample native-Lexile items.

150L

It's the middle of the day. Workers stop working to eat. They take out sandwiches or soup.

They have _____.

A. lunch*
B. pets
C. cars
D. honey

710L

Caleb sat on the grass near the fence, trying to read his class assignment. But the warm sun and gentle breeze made him drowsy. He closed his eyes and listened to the bees buzzing nearby. He opened his eyes and watched the clouds travel slowly across the sky. He heard the screen door open suddenly as his mother came outside to look for him. "Have you finished your homework yet?" she asked. Caleb sighed quietly and picked up his book again.

Caleb was _____.

A. distracted*
B. ignored
C. protected
D. confused

1060L

Galen was born in 129 CE. His avid interest in experimentation completely transformed medicine for the next thousand years. A keen supporter of observation and dissection, Galen performed basic experiments to understand how systems of the body worked. Though Roman law forbade the dissection of humans, Galen was able to work with other animals to understand the functions of the nervous system. Galen theorized, as a result of his experiments, that the brain controlled the muscles. This was a radical idea for the time. When human dissection became legal long after his death, Galen's theories regarding the human brain were proven correct while his other theories were not.

Galen's discovery about the brain was _____.

A. fundamental*
B. predictable
C. symbolic
D. unnecessary

One- and Two-sentence Items

The one-sentence and two-sentence items are developed in a manner similar to the native items. The sentence in the item sets up a context for the reader and provides enough information for reader to use to select the correct answer.

In the one-sentence items, the single sentence contains the missing word. In this item type, only the correct answer is a plausible option given the context of the one sentence.

In the two-sentence items, the second sentence has a missing word followed by four options. From the four options, the reader is asked to select the "best" option that completes the sentence. The two sentences are written to ensure that by reading and comprehending the first sentence, the reader is able to complete correctly the second sentence. With this format, all options are syntactically appropriate completions of the sentence, but one option is unambiguously the "best" option when considered in the context of the first sentence.

The one-sentence and two-sentence item types are designed to measure reading comprehension at a targeted Lexile zone. The target Lexile zone guides the item development. For two-sentence items, the Lexile zone of the item is determined by using the Lexile Analyzer to measure the level of the complete two-sentence item. The second sentence includes the correct answer. Key context words should be from the target Lexile level. *Figure 5* provides examples of one- and two-sentence items.

Figure 5. Sample one- and two-sentence items.

-170L

Can Sal _____ the cat?

A. pat*
B. ban
C. pit
D. dot

130L

The dog could not catch the cat.

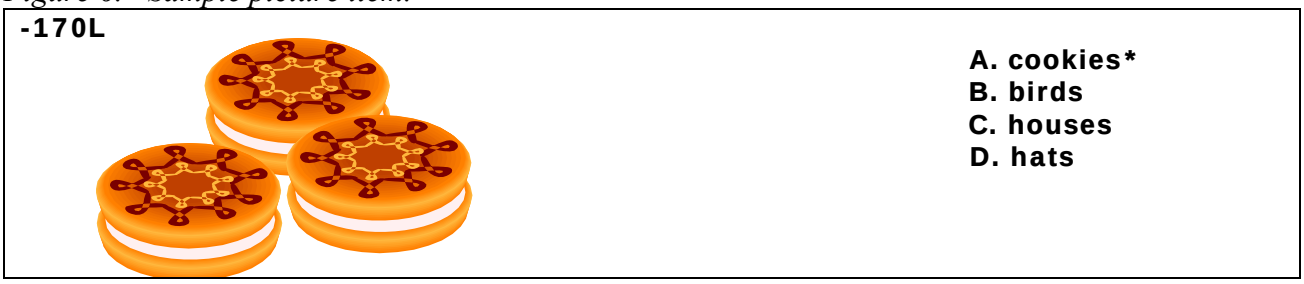
The cat was very _____.

A. fast*
B. big
C. long
D. funny

Picture Items

Picture items are developed to align with key words chosen from LVA and grade-level appropriate word lists. The distractors are also chosen from LVA and grade level lists. The distractors are the same part of speech as the key and are unambiguously incorrect. In some items, the distractors are purposely selected to correspond to additional characteristics (such as syllable count or initial sound) of the key. MetaMetrics field-tests all of the picture items. Before use in a field study, these items are reviewed by MetaMetrics staff for content, sensitivity, and grade appropriateness. After the field study, the items are calibrated to the Lexile scale using the empirical data. Additional data from each item's performance is also reviewed by content specialists and psychometricians; any item that performs poorly is removed from the LRAIB. The difficulty of a picture item is based on the Lexile measure derived from the field test. *Figure 6* provides an example of a picture item.

Figure 6. Sample picture item.



LRAIB Item Review

All items developed for the LRAIB undergo a thorough two-stage review process prior to submission for client approval. First, items are reviewed and edited according to the item development criteria and for sensitivity issues described below. Items are then reviewed and edited by a group of specialists that represent various perspectives, including test developers, English Language Learner (ELL) consultants, literacy teachers, and editors. These individuals examine each item for sensitivity issues and for the quality of the item and response options. During this second stage of the item review process, additional edits may be incorporated.

The following criteria are used during the development and review of Lexile items.

Item Conventions--

- Stems should require the student to draw an unambiguous conclusion or inference from the passage.
- Stems should be clear as to what or whom the sentence question is about.
- Stems should attempt to avoid the use of negatives.
- The correct answer should not be a word that is the same as or closely similar to words that appear elsewhere in the passage.
- Answer choices should be one word.
- Answer choices should be reasonably Lexile-targeted. For native items, 100L below to 100L above the passage is a general guideline.
- Answer choices should logically complete the statement to force passage dependence for answering correctly. All answer choices should make sense in the context of the stem, but only the correct choice should make sense in the context of the paragraph.
- Answer choices should not be homonyms, as this may merely confuse the reader. Answer choices should not be antonyms; if two choices are opposite there is a high probability that one is correct.
- Answer choices should be balanced. If the correct answer 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. With higher-level texts, it is best to try and make all of the words positive or negative. Additionally, correct answer choices should not stand out in length, beginning letters or any other property.
- Answer choices should be selected in accordance with sensitivity guidelines.

Bias and Sensitivity Guidelines--

- Reading passages should be age-appropriate for the intended student population.
- Standard English conventions appropriate for students at the targeted grade and reading level should be used in all passages. Some fictional passages can incorporate non-biased colloquial expressions that are appropriate for the targeted grade and reading level.
- Bias based on race, gender, age, ethnicity, religion, disability, sexual orientation, or socioeconomic status should not be present in passages or items. No group should have an advantage over another because of values, vocabulary, phrasing, or assumptions in a passage. Passages and items should avoid stereotypes of ethnic or gender groups.
- To the degree possible, unique prior knowledge should not be required for the examinee to understand or appreciate the passage; that is, whatever prior knowledge is required should be judged to be already possessed by all likely examinees. References to events, people, and places should be explained within the passage unless considered common knowledge. Figurative language should be explained within the passage or be defined through context.
- Topics that may be offensive to, or induce an emotional reaction from, a student, parent, or citizen group (e.g. violence, abuse, terminal illness, poverty) should be avoided in passages and items.
- Registered trademarks and brand names should not appear in passages or items. Common business names should also be avoided in passages and items.

Editorial Guidelines—

- Use of Real-World Contact Information. 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.
- Style Manual. The Editorial Department of MetaMetrics uses *The Chicago Manual of Style* (16th edition, 2010) and *Merriam-Webster's Unabridged Dictionary* (Online edition).

LRAIB Item Field Testing and Calibration

In addition to content and sensitivity reviews during the development process, LRAIB items are field-tested as part of MetaMetrics on-going research. LRAIB items may be field-tested as part of stand-alone research field tests or they may be embedded within research tests for concurrent projects. Several recent research studies including LRAIB items are described below.

Research Studies

Study 1. In the spring of 2012, 160 LRAIB items were field tested with high school students in a small school district as the Lexile Research Test (LRT). The LRT was administered as a stand-alone research field test, with one form for each grade (Grades 9-12). Each form contained 40

items, including literary and informational texts. Means and ranges of the item difficulties for the LRT were designed to reflect the text complexity ranges recommended in the Common Core State Standards. A total of 411 students participated in the study, with between 85 and 125 students per grade.

Study 2. As part of a 2012 research study designed to link a nationally normed, standards-based test with the Lexile Framework, 12 LRAIB items were embedded into the assessment. A total of 3,217 Grade 8 students participated in the study. Each LRAIB item was administered to an average of 1,686 students.

Study 3. In the spring of 2012, a Southern state administered a study designed to link its state reading assessment with the Lexile Framework. Across Grades 3-8, 6,480 students participated in the study. The study included 12 LRAIB items, with each item being administered to an average of 1,080 students.

S*tudy 4.* In the spring of 2013, a state in the central region of the East Coast administered a study designed to link its state reading assessment with the Lexile Framework. Across Grades 3- 8, and 11. 3,128 students participated in the study. The study included 55 LRAIB items, with each item being administered to an average of 432 students.

Study 5. In the spring of 2013, a state in the southern region of the East Coast administered a study designed to link its state reading assessment with the Lexile Framework. The study included 12,490 students in Grades 3, 5, 7, 8 and English II. For each grade, two forms of the linking test were developed. The study included 102 LRAIB items, with each item being administered to an average of 1,280 students.

Study 6. As part of a 2014 research study designed to link a nationally-normed computer-adaptive reading assessment with the Lexile Framework, 205 LRAIB items were embedded in the assessment. A total of 294,967 students in Grades 1 through 12 participated in the study. Each LRAIB item was administered to an average of 8,169 students.

Study 7. In the summer of 2014, the Lexile Research Test (LRT) was administered to high school students learning English in Japan. The LRT was administered as a stand-alone research field test, with one form for Grades 9 and 10, one form for Grade 11, and one form for Grade 12. Each form contained 30 items, with a total of 90 LRAIB items field tested. A total of 414 students participated in the study, with each item being administered to an average of 138 students.

Tables 8 and *9* present descriptive data of the LRAIB items in research studies.

Table 8. Research studies including LRAIB items administered in the United States.

Study	Year administered	Number of LRAIB Items Tested	Number of participants in study	Grades Tested
1	2012	160	411	9-12
2	2012	12	3,217	8
3	2012	12	6,480	3-8
4	2013	55	3,128	3-8, 11
5	2013	102	12,490	3, 5, 7, 8, English II
6	2014	205	294,967	1-12

Table 9. International research studies including LRAIB items administered to students who are not native English speakers.

Study	Year administered	Number of LRAIB Items Tested	Number of participants in study	Grades Tested
7	2014	90	414	9-12

Field-Test Analyses

During these field studies, LRAIB items were analyzed using both the classical measurement model and the Rasch (one-parameter logistic item response theory) model. Item statistics and descriptive information (item number, field test form and item position, and answer key) were compiled for each item

Classical Measurement. For each item, the p-value (percent correct) and the point-biserial correlation between the item score (correct response) and the total test score were computed. Point-biserial correlations were also computed between each of the incorrect responses and the total score. In addition, frequency distributions of the response choices (including omits) were tabulated (both actual counts and percents). *Table 10* and *Table* 11 display the classical item statistics.

Rasch Item Response Theory. Classical test theory has two basic shortcomings: (1) the use of item indices whose values depend on the particular group of examinees from which they were obtained, and (2) the use of examinee ability estimates that depend on the particular choice of items selected for a test. The basic premises of item response theory (IRT) overcome these shortcomings by predicting the performance of an examinee on a test item based on a set of underlying abilities (Hambleton and Swaminathan, 1985). The relationship between an examinee's item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an item characteristic curve (ICC). This function specifies that as the level of the trait increases, the probability of a correct response to an item increases.

The conversion of observations into measures can be accomplished using the Rasch (1980) model, which states a requirement for the way that item calibrations and observations (count of

correct items) interact in a probability model to produce measures. The Rasch IRT model expresses the probability that a person (*n*) answers a certain item (*i*) correctly by the following relationship:

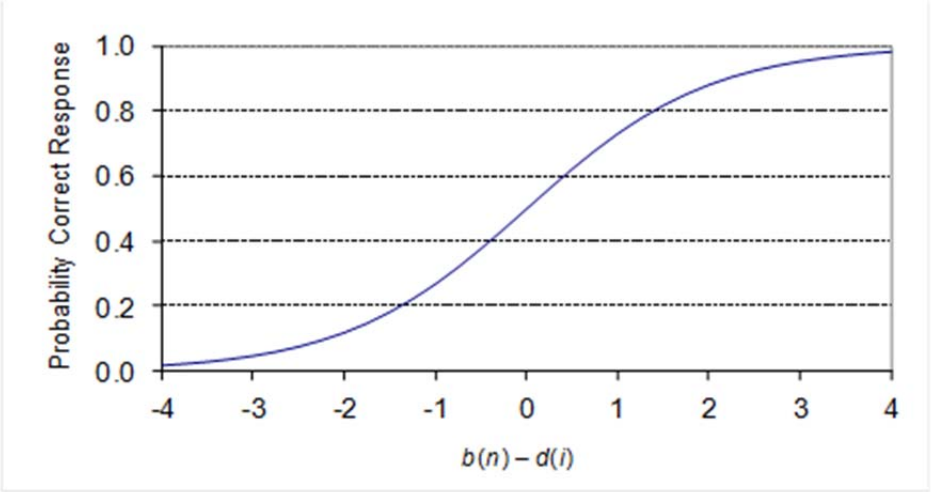
$$P_{ni} = \frac{e^{b_n - d_i}}{1 + e^{b_n - d_i}} \qquad \text{(Equation 4)}$$

where d_i is the difficulty of item i (i = 1, 2, …, number of items);
b_n is the ability of person n (n = 1, 2, …, number of persons);
$b_n - d_i$ is the difference between the ability of person n and the difficulty of item i; and
P_{ni} is the probability that examinee n responds correctly to item i
(Hambleton and Swaminathan, 1985; Wright and Linacre, 1994).

This measurement model assumes that item difficulty is the only item characteristic that influences the examinee's performance such that all items are equally discriminating in their ability to identify low-achieving persons and high achieving persons (Bond and Fox, 2001; Hambleton, Swaminathan, and Rogers, 1991). In addition, the lower asymptote is zero, which specifies that examinees of very low ability have zero probability of correctly answering the item. The Rasch model has the following assumptions: (1) unidimensionality—only one ability is assessed by the set of items; and (2) local independence—when abilities influencing test performance are held constant, an examinee's responses to any pair of items are statistically independent (conditional independence, i.e., the only reason an examinee scores similarly on several items is because of his or her ability, not because the items are correlated). The Rasch model is based on fairly restrictive assumptions, but it is appropriate for criterion-referenced assessments.

Figure *Figure 7* graphically shows the probability that a person will respond correctly to an item as a function of the difference between a person's ability and an item's difficulty.

Figure 7. The Rasch Model--the probability person n responds correctly to item i.



An assumption of the Rasch model is that the probability of a response to an item is governed by the difference between the item calibration (d_i) and the person's measure (b_n). From an examination of the graph in *Figure 7*, when the ability of the person matches the difficulty of the item ($b_n - d_i = 0$), then the person has a 50% probability of responding to the item correctly.

The number of correct responses for a person is the probability of a correct response summed over the number of items. When the measure of a person greatly exceeds the calibration (difficulties) of the items ($b_n - d_i > 0$), then the expected probabilities will be high and the sum of these probabilities will yield an expectation of a high "number correct." Conversely, when the item calibrations generally exceed the person measure ($b_n - d_i < 0$), the modeled probabilities of a correct response will be low and the expectation will be a low "number correct."

Thus, Equation 3 can be rewritten in terms of the number of correct responses of a person on a test

$$O_p = \sum_{i=1}^{L} \frac{e^{b_n - d_i}}{1 + e^{b_n - d_i}} \qquad \text{(Equation 5)}$$

where O_p is the number of correct responses of person p and L is the number of items on the test. When the sum of the correct responses and the item calibrations (d_i) is known, an iterative procedure can be used to find the person measure (b_n) that will make the sum of the modeled probabilities most similar to the number of correct responses. One of the key features of the Rasch IRT model is its ability to place both persons and items on the same scale. It is possible to predict the odds of two individuals being successful on an item based on knowledge of the relationship between the abilities of the two individuals. If one person has an ability measure that is twice as high as that of another person (as measured by *b*—the ability scale), then he or she has twice the odds of successfully answering the item.

Equation 4 possesses several distinguishing characteristics:

- The key terms from the definition of measurement are placed in a precise relationship to one another.
- The individual responses of a person to each item on an instrument are absent from the equation. The only information that appears is the "count correct" (O_p), thus confirming that the raw score (i.e., number of correct responses) is "sufficient" for estimating the measure.

For any set of items the possible raw scores are known. When it is possible to know the item calibrations (either theoretically or empirically from field studies), the only parameter that must be estimated in Equation 4 is the person measure that corresponds to each observable count correct. Thus, when the calibrations (d_i) are known, a correspondence table linking observation and measure can be constructed without reference to data on other individuals.

The item responses were submitted to a Winsteps IRT analysis. The resulting item difficulties (in logits) were assigned Lexile measures by multiplying by 180 and anchoring each set of items to the mean theoretical difficulty of the items on the form.

Tables 10 and *11* present the item response theory results (Lexile measures).

Table 10. Item-level descriptive statistics of LRAIB items included in studies administered in the United States.

Study	Number of LRAIB Items Tested	Number of Student Responses Per Item Mean (Range)	*p*-value Mean (SD)	Point-biserial Mean (Range)	Lexile Measure Mean (SD)
1	160	95 (70 – 125)	0.93 (0.19)	0.37 (-0.35 – 0.66)	1193.72 (209.09)
2	12	1686 (1,492 – 1,820)	0.59 (0.15)	0.37 (0.16 – 0.51)	1118.33 (117.77)
3	12	1080 (593 – 1,415)	0.48 (0.28)	0.30 (-0.01 – 0.52)	1082.5 (222.51)
4	55	432 (315 – 584)	0.56 (0.20)	0.38 (0.11 – 0.64)	1109.82 (187.95)
5	102	1280 (1,110 – 2,654)	0.61 (0.19)	0.38 (0.08 – 0.59)	1053.53 (268.91)
6	205	8169 (715 – 16,488)	0.55 (0.17)	0.26 (-0.18 – 0.42)	836 (361.13)

Table 11. Item-level descriptive statistics of LRAIB items included in international studies administered to students who are not native English speakers.

Study	Number of LRAIB Items Tested	Number of Student Responses Per Item Mean (Range)	*p*-value Mean (SD)	Point-biserial Mean (Range)	Lexile Measure Mean (SD)
7	90	138 (99 – 213)	0.36 (0.16)	0.15 (-0.31 – 0.50)	820.67 (225.07)

Where necessary, items are deleted from the item bank or revised and recalibrated. The item data from the field studies is also used to inform item selection from the LRAIB for future projects.

Development of LightSail Power Challenge Assessment

The LightSail Diagnostic reading assessment was designed to measure initial reading ability. LightSail Inc. identified criteria for the development of the assessment:

- Simplified test administration that could be accomplished through a web-based environment.
- Minimum number of items per test form and minimum administration time while still ensuring minimal measurement error when determining each student's reading ability.
- Adapted test level administration during the test administration to best measure a student's estimated reading ability.

Test specification for the LightSail Diagnostic reading assessment began during June 2015 with test development, final test evaluation, and operational materials being completed during summer 2015.

LightSail Power Challenge Specifications

The LightSail Diagnostic reading assessment specifications consisted of an assessment that covered Grades 1 through 12 with 32 items per test form and with one test form developed for each grade level. It was determined that test forms would be developed specifically for the following levels: Grade 1, Grade 2, Grade 3, Grade 4, Grade 5, Grade 6, Grades 7-8, Grades 9-10, and Grades 11-12.

The LightSail Diagnostic reading assessment target mean was set at approximately the 60th percentile of each grade, which places it within the stretch targets proposed in Appendix A of the Common Core State Standards for English Language Arts (need citation). The low end of the targeted range corresponds to a Lexile measure at the 10th percentile student measure and the high end corresponds to approximately the 95th percentile. The mean and range values allow assessment use with a wide range of general education and intervention programs. *Table 12 s*hows more detailed specifications for each of the LightSail Power Challenge forms.

Table 12. Specifications for LightSail Power Challenge forms.

Test (Grade)	Target Mean	Target Minimum	Target Maximum
1	80L	-330L	530L
2	230L	-150L	650L
3	550L	100L	880L
4	660L	210L	980L
5	770L	330L	1080L
6	880L	480L	1170L
7-8	990L	580L	1260L
9-10	1095L	690L	1365L
11-12	1145L	800L	1420L

For the Grades 1 through 5 test forms, the proportion of literary (narrative) and informational (nonfiction expository) content will progress through the grades to reflect the 2009 NAEP Reading Framework. The Common Core Standards for English Language Arts advocate for instruction and content to align with the NAEP framework, with the inclusion of informational texts in English language arts and other content area classrooms. Tests forms for Grades 1 through 5 will contain approximately 50% literary (narrative) and 50% informational (nonfiction expository) content. Tests forms for Grades 6 through 11-12 will contain 40% literary (narrative) and 60% informational (non-fiction expository) content.

All items developed for the LightSail Power Challenge forms are native-Lexile items, with the exception of a small proportion of items developed for Grades 1 and 2. Because some readers at this level are not ready for the challenge of a test consisting only of native-Lexile items, these forms include the more accessible picture, one-sentence, and two-sentence items. By including these items in addition to native-Lexile items, early and developing readers can be measured appropriately and placed on the Lexile scale with a Lexile measure. *Table 13* includes information about item types in Grades 1 and 2.

Table 13. Item types for the Grades 1 and 2 LightSail Power Challenges.

Test (Grade)	Picture Items per Form	One-sentence Items per Form	Two-sentence Items per Form	Native-Lexile Items per Form
1	5	9	9	9
2	-	5	9	18

LightSail Power Challenge Development

Using the specifications for the LightSail Diagnostic reading assessments described in *Tables 12* and *13*, one operational test form was created for Levels 1 through 11-12.

The form review process for LightSail Diagnostic was conducted in a three-stage process. First, the test and passage specifications were reviewed: Lexile measures of items and means and standard deviations of test forms, word counts across the forms, and distributions of correct responses. Next, the tests were 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 test form:

Curricular Perspective

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

Psychometric Perspective

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

The final item parameter information for each of the LightSail Power Challenge forms is presented in *Table 14*.

Table 14. Operational test form statistics for LightSail Diagnostic.

Test (Grade)	Mean (SD)	Minimum	Maximum
1	78.13 (240.49)	-360L	530L
2	232.50 (222.62)	-190L	690L
3	549.38 (200.71)	120L	860L
4	657.81 (177.21)	210L	960L
5	770.31 (187.59)	340L	1060L
6	875.94 (155.87)	480L	1150L
7-8	989.69 (172.54)	560L	1280L
9-10	1087.19 (175.60)	670L	1370L
11-12	1144.06 (179.90)	800L	1420L

Development of the LightSail In-Text Embedded Assessments

The Lexile Cloze Generation Engine was developed by MetaMetrics to meet the criteria for an effective assessment system identified by teachers. The genesis for the engine was from educators who desired to use automation for two formative assessment tasks: (1) auditing students' completion of well-targeted reading assignments and (2) assessing their level of performance. The first version of the Lexile Cloze Generation Engine was employed in a prototype product developed by MetaMetrics named Inline Reader (IR). IR was designed to audit students' completion of reading assignments by recording the total amount of time spent reading the assignments. Performance was monitored by requiring students to complete auto-generated embedded cloze tasks and writing a summary of the assigned reading.

Specifications of the Lexile Cloze Generation Engine

The Lexile Cloze Generation Engine auto-generates embedded items within a passage. The traditional cloze procedure for measuring reading comprehension is based on the deletion of every 5th to 7th word (or some variation) regardless of part of speech (Bormuth, 1967, 1968, 1970). It can also consist of selectively deleting certain categories of words. Selective deletions have shown greater instructional effects than random deletions (Greene, 2001). There is evidence to support that the cloze procedure reveals both text comprehension and language mastery levels. Some of the research done with metacognition shows that better readers use more strategies (and the appropriate strategy) when they read. In addition, the cloze procedure has been shown to require more re-reading of the passage and an increase in the use of context clues, both characteristics of better readers. The cloze-item format has been shown in multiple studies to measure the same reading construct as norm- and criterion-referenced tests (Stenner, Smith, Horiban, and Smith, 1987a).

The Lexile Cloze Generation Engine, version 4 (MetaMetrics, 2009) is designed to cloze a specific number of words in a text based on its length. It attempts to produce the optimal set of clozes based on two criteria: Lexile targeting and even distribution of clozed words. Clozes and distractors are selected based on a set of configurable criteria. First, the text is preprocessed and analyzed using the Lexile Analyzer. All words are tagged with Lexile word measures. Then parts of speech are tagged using a Maximum Entropy method. Next, the engine calculates an ideal number of clozes to be selected based on the article's length. Once the text is tagged and the engine knows how many clozes to select, an initial pass to find clozes is made with strict limits on inter-cloze distance and Lexile targeting. After that initial selection of clozes, the engine looks for the sparsest region (i.e. the area of the text furthest from an already clozed word) and attempts to find a cloze nearest the center of this region while maintaining the same strict limits. If it is unable to find a word to cloze within those limits, it will gradually and systematically reduce the criteria, starting first with reducing limits on inter-cloze distance and eventually reducing limits on Lexile targeting. Once it finds a word to cloze, it resets the criteria and finds the next sparsest region. The engine continues this process until it has (a) found the target number of clozes or (b) exhausted all possible regions of the text for potential clozes.

Specific rules are used within the Lexile Cloze Generation Engine related to target cloze count (the number of words clozed within a passage), cloze density (the proportion and location of clozed words within a passage), cloze word selection, and distractor word selection (MetaMetrics, 2009).

Research with the Lexile Cloze Generation Engine

In the spring of 2005, MetaMetrics conducted a linking study to determine if IR could produce Lexile measures and, if so, check for what adjustments to measures the new item type required.

Reading Riches, a reading and writing motivation program employing the Lexile Cloze Generation Engine, was implemented in two large school districts during the 2004-2005 school year. Reading was assessed with Inline Reader technology. The goal of the study was to link the IR item type (i.e., auto-generated embedded sentence cloze task) with the Lexile scale. The following section describes the version of IR used by participants and the results of the linking study.

A sample of 1,498 students in Grades 5 through 12 was administered both a Lexile linking test comprised of native-Lexile items and an online administration of IR passages targeted at 50%, 75%, and 90% comprehension for the typical student at each grade. There were two forms of IR passages, one with conditioned items (Form B) and one with unconditioned items (Form A). Conditioning involved human interaction with the foils to edit out any perceived foibles caused by the computer algorithm. The unconditioned items were completely auto-generated without any post editing. The specifications for each form are presented in *Table 15*.

Table 15. IR linking study test form specifications.

Grade	Form	Article 1		Article 2		Article 3	
		Lexile	# of items	Lexile	# of items	Lexile	# of items
5	A	500L	5	810L	15	900L	12
	B	520L	8	800L	16	890L	8
6	A	650L	12	900L	7	1100L	8
	B	680L	6	900L	10	1100L	8
7*	A	780L	8	1000L	8	1200	11
	B	770L	9	1000L	11	1200L	10
8	A	800L	10	1100L	13	1300L	8
	B	869L	14	1100L	9	1300L	10
9/10	A	950L	9	1200L	13	1400L	8
	B	950L	11	1200L	9	1300L	10
11/12	A	1000L	8	1250L	10	1450L	9
	B	1000L	12	1250L	8	1440L	9

A computer error was discovered in the implementation of the Grade 7 unconditioned form, therefore, Grade 7 data was removed from further analyses. The first analyses examined the point measure correlations to see if the IR items were performing as expected. The correlations presented in *Table 16* were lower than had been observed previously for other reading item types, even when controlling for potential artifacts like range restriction (MetaMetrics, 1999a, 1999b, 2000, 2006b).

Table 16. Mean point-biserial and point measure correlations from various reading research studies.

	Date	Number of Items	Mean Point Measure Correlation	Mean Point-Biserial Correlation
PASeries Reading	2004	342	0.42	
Native-Lexile items (Duval, FL)	1999	427		0.42
Native-Lexile items (Miami, FL)	1999	300		0.42
IR	2005	280	0.33	
Lingos Vocabulary Assessment	2000	65		0.39

The correlations for the conditioned and unconditioned IR items were similar, although the unconditioned items had slightly higher values at .33 (.32 for the conditioned form). In addition, the person data from the IR linking study was examined to determine whether the degree to which the item performance affected person measures. Most of the 1,498 students in the study took two native-Lexile tests -- one in the winter and one in the spring -- as well as the IR field study form. This design enabled a within-grade "roundabout" to be employed to determine if the correlations between tests comprised of IR items produced measures as highly correlated with native-Lexile item tests as two native-Lexile item tests are correlated with each other. *Table 14* provides descriptive statistics for forms used in the study.

Only in Grade 8 does the IR unconditioned item correlation not compare favorably with the native-Lexile item correlation. In Grade 5, only one native-Lexile form was administered, so there is no native-to-native correlation with which to compare. However, a correlation of .80 for the native to IR unconditioned form is high for a within-grade raw score correlation. These results support the premise that IR items and native-Lexile items measure the same construct—reading comprehension. The data suggests that measures produced by IR items will link suitably to the Lexile scale. Because the conditioned items produced slightly lower correlations than the unconditioned items, and because the conditioned items require human intervention, conditioning IR items was not considered necessary in further research.

Table 17. Descriptive statistics for test forms in linking study (IR forms standardized to 45 items on a form).

Grade	Native to Native		Native to IR (Unconditioned Form)		Native to IR (Conditioned Form)	
	r	*N*	*r*	*N*	*r*	*N*
5			.80	46	.76	38
6	.59	197	.59	91	.35	106
8	.73	169	.51	86	.57	83
9	.81	331	.82	295	.77	290
10	.76	254				
11	.66	140	.66	130	.63	140
12	.73	130				
Mean	*.72*		*.68*		*.62*	

Checking for theory fit of the items with the Lexile Theory was the last stage of the IR item-validation process. The goal for the IR engine was to auto-generate items that can produce Lexile measures based on theory alone. Given the multiple-items-per-passage nature of IR items, *PASeries* Reading's passage-native Lexile items provided an appropriate interpretative framework for accessing how well the Lexile theory predicts the difficulty of the items for each passage. The root mean square error (RMSE) indicates how closely the difficulty based on theory and on observed results match. Results are presented in *Table 18*.

Table 18. Comparison of RMSEs on passage means for two item formats.

Item Type	Number of Passages	Root Mean Square Error (RMSE)	Within passage Standard Deviation (SD)
Passage Native from *PASeries* Reading	42	151L	159L
IR	15	152L	207L

The theoretical prediction for IR items (RMSE 152L) is nearly identical to that of the passage natives (151L). The within-passage variability on the IR items is much higher, but no adjustment to the Bayesian scoring algorithm when computing measures within IR is made at this time.

Table 19 compares Lexile passage measures produced using theoretical Lexile measures with measures produced using observed Lexile measures from auto-generated items. The observed measures were anchored on the native-Lexile items' theoretical measures.

Table 19. Difference in mean passage difficulty when computed by theory and when observed (weighted).

	Number of Passages	Mean	SD
Theory	15	1023L	268L
Observed	15	1041L	161L

The results from the administration support the conclusion that a format adjustment was not necessary to link reader Lexile measures from IR with reader Lexile measures from native-Lexile items.

Cloze Engine Tuning -- LightSail In-Text Embedded Assessment Development

To create the LightSail in-text embedded assessments, LightSail identified every third page in a book to use as a passage and include up to two cloze items. Passages with Lexile measures greater than 100L different from the complete text were not retained for assessment development.

Each passage was passed through the Lexile Cloze Generation Engine using default rules related to target cloze count (the number of words clozed within a passage), cloze density (the proportion and location of clozed words within a passage), cloze word selection, and distractor word selection. MetaMetrics staff reviewed preliminary passage and item sets to determine if the default rules maximized the assessment value of the passages. Closer examination focused on one potential change to the default rules: (1) the selection of distractors for each item based on passage Lexile measure.

To examine the appropriate Lexile ranges of potential distractors, MetaMetrics staff examined passages spanning the text complexity range of books employed by the LightSail software. Each passage was passed through the Cloze Generation Engine twice, once with the default range of 600L and once with the alternative range of 300L. Few differences were seen in the selection of distractors between these ranges. It was concluded that the reduced range had no negative impact and the decision was made to use the lower limit for LightSail in-text embedded assessments.

Three other modifications were made to the Lexile Cloze Generation Engine for use within the LightSail software:

1. adverbs were removed from cloze selection given the limited number of words that could be used as distractors;
2. distractors were selected to be consistent with the cloze word in terms of whether “a” or “an” preceded the cloze word; and
3. distractors were selected to have the same number of characters (length) as the cloze word.

Scoring and Reporting

The two main purposes of the LightSail assessment components are to measure student-reading comprehension so reading materials can be appropriately targeted and to measure iteratively growth in reading comprehension throughout the school year. In order to meet these goals, a developmental scale must be used to report the results. The LightSail assessment components are reported on the Lexile scale. This section describes the procedures and the analyses used to score and report the results of the LightSail assessment components.

Test Use Guidelines. 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, 2014).

LightSail Power Challenge Scoring

LightSail Diagnostic reading assessment scores are reported on the Lexile scale. For the Power Challenge, 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. But, receiving a scale score (Lexile measure) of 675 on one form of a test represents a similar level of reading ability as receiving a scale score (Lexile measure) of 675 on another form of the test.

Correspondence tables were provided for each test form based upon the difficulties of the items on the form.

LightSail In-Text Embedded Assessment Scores

The LightSail in-text embedded assessments are scored using a distributed difficulty algorithm. The distributed difficulty method is designed to handle the case when individual item difficulties are not known, but can be regarded as a sample from a population whose parameters are known (MetaMetrics, 2011). For items produced with the Lexile Cloze Generation Engine, the items are assumed to be from a distribution with a mean equal to the Lexile measure of the passage and the standard deviation is assumed constant across passages.

For the in-text embedded assessments, 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 passed to the Bayesian Scoring algorithm module along with the Lexile measure of the passage and the number of cloze items in the passage.

Scoring LightSail Assessments: The Bayesian Paradigm

Bayesian methodology provides a paradigm for combining prior information with current data, both of which are subject to uncertainty, and for arriving at an estimate of current status, which is again subject to uncertainty. Uncertainty is modeled mathematically using probability.

In the LightSail context, when a student is administered the Power Challenge, the results from the test become the prior information for the following test administration—in-text embedded assessments. Each subsequent assessment uses prior information from all previous assessments.

The current data in this context is the performance on the current test (i.e., in-text embedded assessment), which can be summarized as the number of items answered correctly out of the total number of items attempted.

Both prior information and current data are represented via probability models reflecting uncertainty. The need for incorporating uncertainty when modeling prior information is intuitively clear. The need for incorporating uncertainty when modeling test performance is, perhaps, less intuitive. Once the test has been taken and scored, and assuming that no scoring errors were made, the performance, i.e. raw score, is known with certainty. Uncertainty arises because test performance is associated with, but not determined by, the ability of the student, and it is that ability, rather than the test performance per se, that we are endeavoring to measure. Any single performance may over- or underestimate a student's ability, based on factors such as luck, prior knowledge, etc. Thus, although we are certain about the test performance once the results have been calculated, we remain uncertain about the ability that produced the performance.

The uncertainty associated with prior knowledge is modeled by a probability distribution for the ability parameter. This distribution is called the prior distribution and it is usually represented by a probability density function (e.g., the normal bell-shaped curve). The uncertainty arising from current data is modeled by a probability function for the data when the ability parameter is held fixed. When roles are reversed so that the data are held fixed and the ability parameter is allowed to vary, this function is called the likelihood function. In the Bayesian paradigm, the posterior probability density for the ability parameter is proportional to the product of the prior density and the likelihood, and this posterior density is used to obtain the new ability estimate along with its uncertainty.

Modeling Growth and Its Impact on the Prior. Once a posterior has been obtained from current data, that posterior can serve as the prior for an immediate repeat assessment. If a substantial amount of time has passed since the last assessment, however, then allowance should be made for an uncertain amount of growth since the last assessment. This allowance is accomplished by

means of a growth model, which estimates as a function of elapsed time both student growth and the augmentation in uncertainty.

Bayesian Scoring Process: Overview of Flow

1. *Administer Power Challenge*. The information from the Power Challenge (Lexile measure, uncertainty) becomes the prior information used by the Bayesian Scoring algorithm to calculate subsequent updated Lexile reader measures.

   During the administration of the Power Challenge, a student's performance is considered periodically to determine whether he or she is performing poorly enough to warrant ending the testing session or administering a lower test level to better estimated his or her reading ability. After 5 or 10 questions, the student's results are examined and the test administration can be stopped if warranted.

   - If the student responds to all of the first five items incorrectly, then the administration is stopped and the student is presented with a lower level of the test to complete.
   - If the student responds to five or more of the first 10 items incorrectly, then the administration is stopped and the student is presented with a lower level of the test to complete.

   From the correspondence tables, students will receive a raw score and Lexile measure based on performance on the final test level completed.

   At the end of the administration of the Power Challenge, a student's performance is also considered to determine whether he or she is performing well enough to warrant administering a higher test level to better estimate his or her reading ability.

   - If the student responds to 28 or more of the 32 items correctly, then the administration continues and the student is presented with the last 10 items from a higher level of the test to complete.

   From the correspondence tables, students will receive a raw score and Lexile measure based on performance on the last 22 items on the complete test administered and the last 10 items on the higher-level test administered.

2. *Administer an In-text Embedded Assessment and Compute New Values.* This step uses the information from student performance on in-text embedded assessments to produce a posterior density. This value is used to create the new Lexile measure and associated uncertainty for the student. The new Lexile measure and uncertainty for the student will be incorporated into the prior information for the scoring of subsequent tests if the student has responded to between 30% and 90% of the cloze items correctly. For each subsequent administration of an in-text embedded assessment, all of the information on the student's reading ability from the previous test administrations is incorporated into the student's prior.

3. *Update reported Lexile measure and "power" text range.* If the student's updated uncertainty measure is less than 60L, then the student's Lexile measure, uncertainty, and "power" text range are reported to the student.

Conditions

1. Negative growth (negative differences in days since last test) is not permitted. If a student takes a test that is not scored and then takes another test, either (1) the first test should not be scored or (2) the first is scored and the second test is re-scored. If the first test is scored, the information will need to be used as the priors for the second test when re-scoring. Zero time (i.e., tests taken on the same day) will follow the standard process. Zero time means that sigma old will be automatically used as sigma update.

2. Changes in answer key and item difficulty should result in a re-score of any test affected. All tests taken after that rescore will need to have the Bayesian Score recalculated.

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 Lexile scale is a developmental scale for reporting reader ability and text complexity, ranging from below 200L for beginning readers and beginning-reader materials to above 1600L for advanced readers and materials. Reader Lexile measures are reported in 5-unit intervals.

Prior to May 1, 2014, all Lexile reader measures at or below 0L were reported as BR (Beginning Reader). Starting in spring 2014, Lexile reader measures below 0L may be reported with a more specific measure. These BR measures are shown as "BRxxxL." For example, a Lexile reader measure of -150 is reported as BR150L where "BR" stands for "Beginning Reader" and replaces the negative sign in the number. The Lexile scale is like a thermometer, with numbers below zero indicating decreasing reading ability as the number moves away from zero. The smaller the number following the BR code, the more advanced the reader is. For example, a BR150L reader is more advanced than a BR200L reader. Above 0L, measures indicate increasing reading ability as the numbers increase. For example, a 200L reader is more advanced than a 150L reader.

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 (e.g., 90^{th} percentile point based on prior research by MetaMetrics with the Lexile Framework). In instructional environments where the purpose of the Lexile measure is to appropriately match readers with texts, all scores below 0L should be reported as "BRxxxL." No student should receive a negative Lexile measure on a score report. It is suggested that the lowest reported value below 0L is BR400L.

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 items that were used to determine the score, and (3) the condition of the reader when the items 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 LightSail assessment components—text error and reader error.

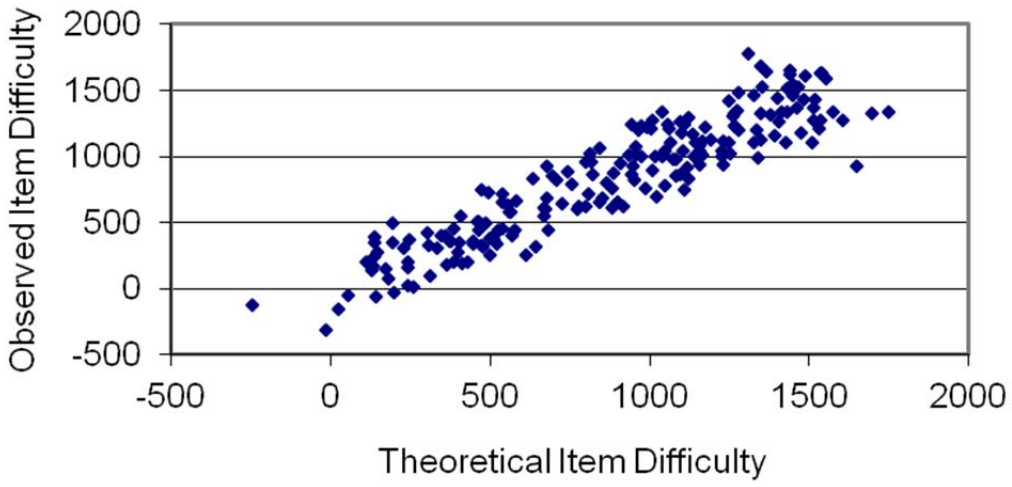
Text Measure Error Associated with The Lexile Framework for Reading

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 *The Stories that Julian Tells* (520L) would be lower than the measure corresponding to a comparable raw score on *USA Today* (1200L). 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. 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.

Study 1. To examine text measurement error, 200 items that had been previously calibrated to the Lexile scale and shown to fit the Rasch 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 8*.

Figure 8. 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 89L whereas a longer book such as *War and Peace* (4,082 slices of text) would only have an uncertainty of 3L (33).

A typical grade 3 reading test has appropriately 2,000 words in the passages. To calibrate this text, it would be sliced into sixteen 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.

Table 20. Standard errors for selected values of the length of the text.

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

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 all of the items that could be developed from a selected piece of text. The theoretical Lexile measure of a piece of text is the mean theoretical difficulty of all items associated with the text. Stenner and his colleagues state that the "Lexile Theory replaces statements about individual items with statements about ensembles. The ensemble interpretation enables the elimination of irrelevant details. The extra-theoretical details are taken into account jointly, not individually, and, via averaging, are removed from the data text explained by the theory" (p. 314). The result is that when making text-dependent generalizations, text readability can be measured with high accuracy and the uncertainty in expected comprehension is largely due to the unreliability in reader measures.

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 is because one of the four forms was 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 21 presents the ensemble study data in which three independent teams wrote one item for each of thirty passages to make a total of 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 21. 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
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
10	820L	967	740	675	794	153	23,445	26
11	510L	267	602	468	446	169	28,413	64
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
15	790L	866	557	553	659	180	32,327	131
16	770L	1077	1095	893	1022	112	12,446	-252
17	850L	747	864	674	762	96	9,257	88
18	870L	974	1197	870	1014	167	28,007	-144
19	880L	1093	733	692	839	221	48,739	41
20	1020L	888	1372	863	1041	287	82,429	-21
21	812L	902	1133	715	917	209	43,753	-105
22	866L	819	809	780	803	20	419	63
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
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 Lexile measures. Mean (O) is the observed ensemble mean. SD is the standard deviation within ensemble.

The difference between passage text calibration and observed ensemble mean is provided in the last column. The root mean square error (RMSE) from regressing observed ensemble means on

text calibrations is 110L. *Figures 9* and *10* show plots of observed ensemble means against theoretical text calibrations.

Figure 9. Plot of observed ensemble means and theoretical calibrations (RMSE = 110L).

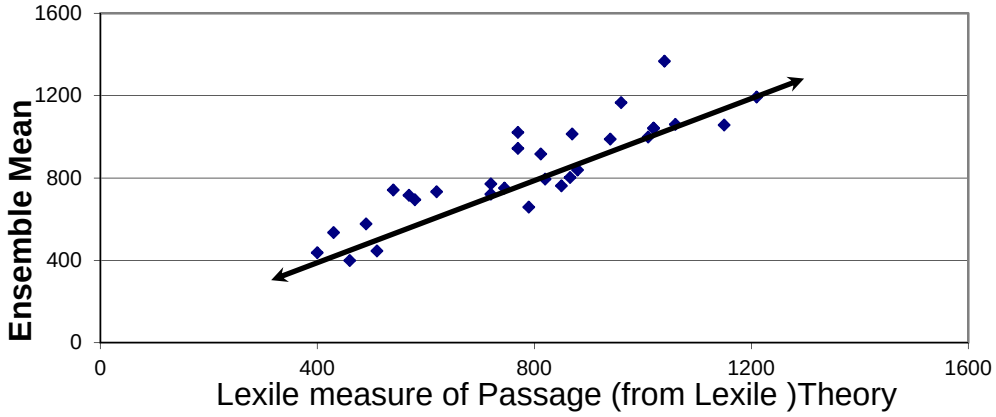
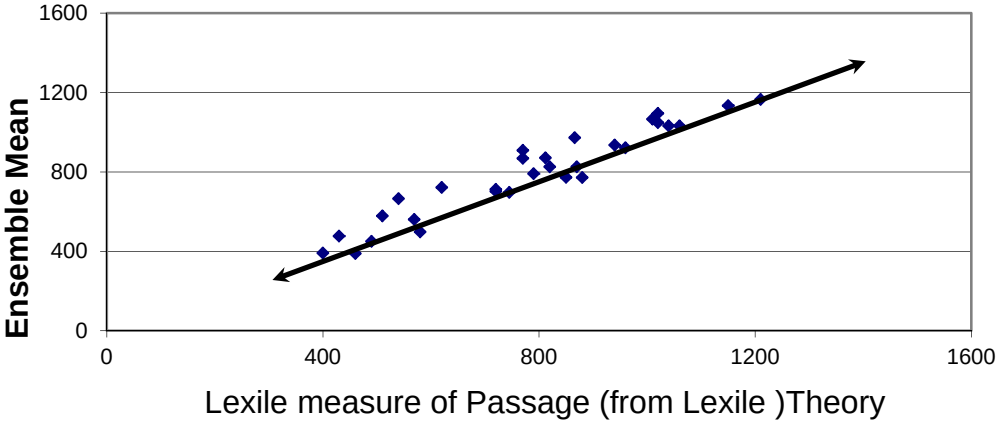


Figure 10. Plot of simulated "true" ensemble means and theoretical calibrations.



Note that some of the deviations around the identity line are because ensemble means are poorly estimated given that each mean is based on only three items. *Figure 4* depicts simulated data when an error term [distributed ~ $N(0, \sigma = 64L)$] is added to each theoretical value. Contrasting the two plots in *Figures 4* and *5* provides 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 22* 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 22. 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 sample 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

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 composed 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. LightSail's Power Challenges 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. *Tables 23* and *24* describe the uncertainties due to model misspecification for the LightSail Power Challenge. The Lexile ranges shown in the table indicate reader measures associated with scores of approximately 25% to approximately 75% correct.

Table 23. Uncertainties for Power Challenge forms by Lexile range (approximately 25% - 75% correct), Grades 1 through 5.

Reader Measure	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
BR400L to BR305L	79				
BR300L to BR205L	75	81			
BR200L to BR105L	74	77			
BR100L to BR5L	76	73			
0 to 95L	80	73			
100L to 195L	83	75	79		
200L to 295L		80	73	77	
300L to 395L			72	72	79
400L to 495L			72	70	74
500L to 595L			79	71	71
600L to 695L				76	71
700L to 795L					75
800L to 895L					79
Median	74	73	72	70	71

Table 24. Uncertainties for Power Challenge forms by Lexile range (approximately 25% - 75% correct), Grades 6 through 11-12.

Reader Measure	Grade 6	Grade 7-8	Grade 9-10	Grade 11-12
400L to 495L	77			
500L to 595L	72	78		
600L to 695L	69	73	78	
700L to 795L	69	70	73	77
800L to 895L	73	70	70	71
900L to 995L	78	73	70	70
1000L to 1095L		79	73	71
1100L to 1195L			78	76
Median	68	69	70	70

Validity

The *2014 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 for proposed uses of tests" (p. 11). 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 the LightSail assessment components will be used as a measure of the reading ability of a student and will be used to target reading materials and instruction, validity evidence should primarily focus on the degree to which the LightSail assessment components measures reading comprehension of appropriate reading material. For convenience, the various sources of validity evidence—content and construct validity evidence—will be described as if they are unique, independent components rather than interrelated parts. A primary source of validity evidence comes from examination of the content of the LightSail assessment components 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. LightSail's Power Challenge and the in-text embedded assessments have been designed to measure comprehension of informational and literary texts. To this end, informational and literary texts have been included in the test forms. In-text embedded assessments are found in "power" texts, which are part of the robust library of digital books across a variety of topics, genres, and levels. 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 for whom the tests were designed. The difficulty of the item vocabulary was also matched to the difficulty of the passage. The sections in this technical report entitled *Development of LightSail Power Challenge* and *Development of LightSail In-Text Embedded Assessments* describe the difficulty of the test passages and the item development process. The passages and items were thoroughly reviewed prior to placement on a test.

In addition to reading complex text, students must use the information to answer questions about the text. The CCSS (NGA and CCSSO, 2010a) identifies three standards related to the key ideas and details in the text that define what students should understand and be able to do--

1. Read closely to determine what the text says explicitly and to make logical inferences from it; cite specific textual evidence when writing or speaking to support conclusions drawn from the text.
2. Determine central ideas or themes of a text and analyze their development; summarize the key supporting details and ideas.

3. Analyze how and why individuals, events, and ideas develop and interact over the course of a text.

PARCC describes close reading as follows:

> Close, analytic reading stresses engaging with a text of sufficient complexity directly and examining meaning thoroughly and methodically, encouraging students to read and reread deliberately. Directing student attention on the text itself empowers students to understand the central ideas and key supporting details. It also enables students to reflect on the meanings of individual words and sentences; the order in which sentences unfold; and the development of ideas over the course of the text, which ultimately leads students to arrive at an understanding of the text as a whole. (PARCC, 2011, p. 7)

With the embedded completion statement item format used with the LightSail Power Challenge, the student is asked to read a passage taken from an actual text and then choose the option that best fills the blank in the last statement. In order to complete the statement, the student must respond on an inferential level (determine the main idea of the passage, draw an inference from the material presented, or make a connection between sentences in the passage). This inferential level is consistent with Depth of Knowledge (DOK) Level 2 (skills and concepts) and Level 3 (strategic thinking) (Webb, 2007).

- Level 2 (skills and concepts) includes the engagement of some mental processing beyond recalling or reproducing a response. The content knowledge or process involved is more complex than in Level 1. Keywords that generally distinguish a Level 2 item include 'classify,' 'organize,' 'estimate,' 'make observations,' 'collect and display data,' and 'compare data.'

- Level 3 (strategic thinking) requires reasoning, planning, using evidence, and higher level of thinking than the previous two levels. The complexity results because the multistep task requires more demanding reasoning.

Construct Validity Evidence

Evidence for the construct validity of the LightSail assessment components is provided by the extensive body of research supporting The Lexile Framework for Reading. The development of the LightSail assessment components 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 Power Challenge are 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. (2014). *Standards for educational and psychological testing*. Washington, DC. American Educational Research Association.

Bond, T.G. & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

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.

Cain, K., Oakhill, J. & Lemmon, K (2004). Individual differences in the inference of word meanings from context. The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology,* 96, 671-681.

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.

Crawford, J. (1978). Interactions of learner characteristics with the difficulty level of instruction. *Journal of Educational Psychology*, *70*(4), 523-531.

Cunningham, A. & Stanovich, K. (1998). What reading does for the mind. *American Educator.* Spring/Summer.

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.

Denham, C., & Lieberman, A., Eds. (1980). *Time to Learn: A review of the beginning teacher evaluation study.* Sacramento: California State Commission of Teacher Preparation and Licensing.

Dunn, L.M. & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test*-Revised, Forms L and M. Circle Pines, MN. American Guidance Service.

Fuchs, L. S., Fuchs, D., Hops, M.K., & Jenkins, J.R. (2001). Oral Reading as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-245.

Greene, B.B. (2001). Testing reading comprehension of theoretical discourse with cloze. *Journal of Research in Reading*. *24*(1), 82-98.

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading and Writing Quarterly*, *19*, 59-85.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and appplications*. Boston: Kluwer · Nijhoff Publishing.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Hare, M.G. (2002, Septemebr 17). `Reading czar' has talk with educators. *The Baltimore Sun*. Retrieved from http://articles.baltimoresun.com/2002-09-17/news/0209170382_1_reading-problems-reading-education-lyon

Jalongo, M. R. (2007). Beyond benchmarks and scores. Reasserting the role of motivation and interest in children's academic achievement. *Childhood Education: International Focus Issue,* 395-407.

Jenkins, J., Stein, M., & Wysocki, K. (1984). Learning vocabulary through reading. *American Education Research Journal, 21*(4), 767-787.

Kim, J.S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis*, *28(*4). 335-355.

Kirsch, I., de Jong, J., LaFontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). Reading for change. Performance and engagement across countries. Paris. Organisation for Economic Co-operation and Development.

Klare, G.R. (1963). *The measurement of readability*. Ames, IA. lowa State University Press.

LightSail Inc. (2015) "LightSail". Retrieved from http://LightSailed.com/

Liberman, 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.

MetaMetrics, Inc. (2008). *Text Measurement and Analysis: MetaMetrics technical report update for the Texas Higher Education Coordinating Board.* Durham, NC: Author.

MetaMetrics, Inc. (2009, December 3). *Inline Reader Engine.* Metawiki.

MetaMetrics, Inc. (2011). Distributed Difficulty Algorithm (unpublished manuscript).

Miller, G.A. & Gildea, P.M. (1987). How children learn words. Scientific American, 257, 94-99.

National Center for Education Statistics. (November, 2011). The Nation's Report Card. Reading 2011.

National Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2010b). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science and Technical Subjects: Appendix A*. Retrieved from http://www.corestandards.org/assets/Appendix_A.pdf

National Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2010a). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. New York: Student Achievement Partners.

O'Connor, R.E., Bell, K.M., Harty, K.R., Larkin, L.K., Sackor, S., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, *94* (3), 474-485.

O'Connor, R.E., Swanson, H.L., & Geraghty (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology*, *102*(1), 1-19.

Partnership for Assessment of Readiness for College and Careers. (2011). *PARCC model content frameworks: English language arts/literacy grades 3–11*. Retrieved from www.parcconline.org/sites/parcc/files/PARCCMCFELALiteracyAugust2012_FINAL.pdf

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.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attachment Tests*. Chicago: The University of Chicago Press (first published in 1960).

Rasinski, T.V. (Ed.). (2009). *Essential readings on fluency*. Newark, DE: International Reading Association.

Sanford-Moore, E., & Williamson, G. L. (2012). *Bending the text complexity curve to close the gap* (MetaMetrics Research Brief). Durham, NC: MetaMetrics, Inc.

Shankweiler, D. & Crain, S. (1986). Language mechanisms and reading disorder. A modular approach. *Cognition*, *14*, 139-168.

Smith, M. (2011, March 30). Bending the Reading Growth Trajectory: Instructional Strategies to Promote Reading Skills and Close the Readiness Gap. MetaMetrics Policy Brief. Durham, NC: MetaMetrics, Inc.

Smith, M. (2012), Not so common: Comparing Lexile® measures with the standards' other text complexity tools. Durham, NC: MetaMetrics.

Squires, D. A., Huitt, W. G., & Segars, J. K. (1983). Effective Schools and Classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

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., Koons, H., & Swartz, C. W. (2010, unpublished manuscript). *Text complexity and developing expertise in reading.* Durham, NC: MetaMetrics, Inc.

Stenner, A. J., Sanford-Moore, E., & Williamson, G. L. (2012). *The Lexile® Framework for Reading quantifies the reading ability needed for "College & Career Readiness."* MetaMetrics Research Brief. Durham, NC: MetaMetrics, Inc.

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.

Webb, N. (2007, September). "Aligning Assessments and Standards". Retrieved from: http://www.wcer.wisc.edu/news/coverStories/aligning_assessments_and_standards.php

Williamson, G. L. (2008). A text readability continuum for postsecondary readiness. *Journal of Advanced Academics, 19*(4), 602-632.

Williamson, G. L., Koons, H., Sandvik, T., & Sanford-Moore, E. (2012). *The text complexity continuum in grades 1-12* (MetaMetrics Research Brief). Durham, NC: MetaMetrics, Inc.

Wright, B.D., & Linacre, J.M. (1984/2003). *A user's guide to WINSTEPS Rasch-Model computer program, 3.38.* Chicago, Illinois: Winsteps.com.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago. MESA Press.

Appendix

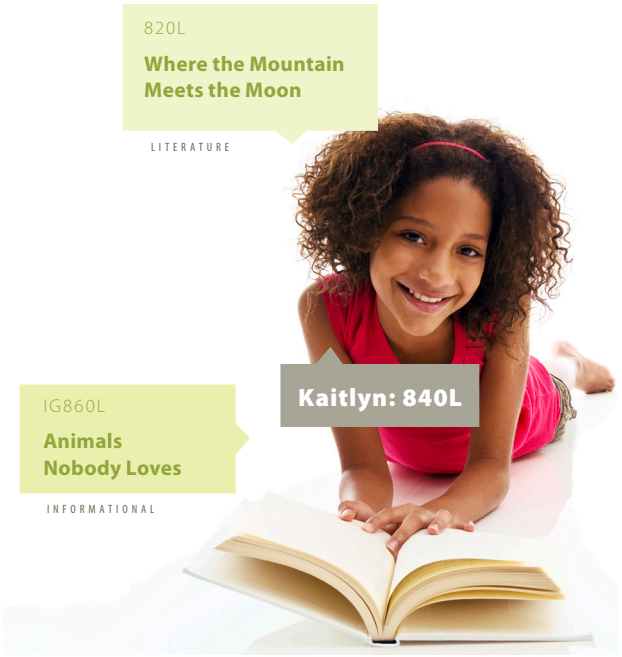
Appendix A. The Lexile Framework for Reading Map

THE LEXILE® FRAMEWORK FOR READING

Matching Readers with Text

Imagine getting students excited about reading while also improving their reading abilities. With the Lexile® Map, students have a chance to match books with their reading levels, and celebrate as they are able to read increasingly complex texts!

Let your students find books that fit them! Build custom book lists for your students by accessing our "Find a Book" tool at Lexile.com/fab.



HOW IT WORKS

The Lexile® Map provides examples of popular books and sample texts that are matched to various points on the Lexile® scale, from 200L for early reading books to 1600L for more advanced texts. The examples on the map help to define text complexity and help readers identify books of various levels of text complexity. Both literary and informational texts are presented on the Lexile Map.

HOW TO USE IT

Lexile reader and text measures can be used together to forecast how well a reader will likely comprehend a text at a specific Lexile level. A Lexile reader measure is usually obtained by having the reader take a reading comprehension test. Numerous tests report Lexile reader measures including many state end-of-year assessments, national norm-referenced assessments, and reading program assessments. A Lexile reader measure places students on the same Lexile scale as the texts. This scale ranges from below 200L to above 1600L. The Lexile web site also provides a way to estimate a reader measure by using information about the reader's grade level and self-reported reading ability.

Individuals reading within their Lexile ranges (100L below to 50L above their Lexile reader measures) are likely to comprehend approximately 75 percent of the text when reading independently. This "targeted reading" rate is the point at which a reader will comprehend enough to understand the text but will also face some reading challenge. The result is growth in reading ability and a rewarding reading experience.

For more guidance concerning targeting readers with books, visit www.Lexile.com/fab to access the "Find a Book" tool. "Find a Book" enables users to search from over 130,000 books to build custom reading lists based on Lexile range and personal interests and to check the availability of books at the local library.

The Lexile Framework for Reading

1300 L ‣ 1500L+
LEXILE RANGE

1500L+

1500L ***Don Quixote**** CERVANTES SAAVEDRA

The Words were to me so many Pearls of Eloquence, and his Voice sweeter to my Ears than Sugar to the Taste. The Reflection on the Misfortune which these Verses brought on me, has often made me applaud Plato's Design of banishing all Poets from a good and well governed Commonwealth, especially those who write wantonly or lasciviously. For, instead of composing lamentable Verses, like those of the Marquiss of Mantua, that make Women and Children cry by the Fireside, they try their utmost Skill on such soft Strokes as enter the Soul, and wound it, like that Thunder which hurts and consumes all within, yet leaves the Garment sound. Another Time he entertained me with the following Song.

SAMPLE TITLES

LITERATURE	1640L	**The Plot Against America** (ROTH)
	1560L	**Rob Roy** (SCOTT)
	1530L	**The Good Earth** (BUCK)
	1520L	**A Fable** (FAULKNER)
	1500L	**The Decameron** (BOCCACCIO)
INFORMATIONAL	1600L	**Sustaining Life: How Human Health Depends on Biodiversity** (CHIVIAN & BERNSTEIN)
	1550L	**The Art of War** (TZU)
	1560L	**The United States' Constitution**
	1520L	**Fair Play: The Ethics of Sport** (SIMON)
	1500L	**Critique of Pure Reason** (KANT)

1400L ‣ 1495L

1400L ***Nathaniel's Nutmeg*** MILTON

Setting sail once again they kept a sharp look-out for Busse Island, discovered thirty years previously by Martin Frobisher, but the rolling sea mists had grown too thick. Storms and gale—force winds plagued them for days on end and at one point grew so ferocious that the foremast cracked, splintered and was hurled into the sea. It was with considerable relief that the crew sighted through the mist the coast of Newfoundland—a vague geographical term in Hudson's day—at the beginning of July. They dropped anchor in Penobscot Bay, some one hundred miles west of Nova Scotia.

SAMPLE TITLES

LITERATURE	1460L	**The Legend of Sleepy Hollow** (IRVING)
	1450L	**Billy Budd**** (MELVILLE)
	1430L	**The Story of King Arthur and His Knights** (PYLE)
	1420L	**Life All Around Me by Ellen Foster** (GIBBONS)
	1420L	**The Scarlet Letter**** (HAWTHORNE)
INFORMATIONAL	1480L	**America's Constitution: A Biography**** (AMAR)
	1430L	**The Declaration of Independence**
	1420L	**Gettysburg Address** (LINCOLN)
	1410L	**Profiles in Courage** (KENNEDY)
	1400L	**The Life and Times of Frederick Douglass** (DOUGLASS)

1300L ‣ 1395L

1300L ***1776: America and Britain at War**** MCCULLOUGH

But from this point on, the citizen-soldiers of Washington's army were no longer to be fighting only for the defense of their country, or for their rightful liberties as freeborn Englishmen, as they had at Lexington and Concord, Bunker Hill and through the long siege at Boston. It was now a proudly proclaimed, all-out war for an independent America, a new America, and thus a new day of freedom and equality. At his home in Newport, Nathanael Greene's mentor, the Reverend Ezra Stiles, wrote in his diary almost in disbelief: Thus the Congress has tied a Gordian knot, which the Parl [iament] will find they can neither cut, nor untie. The thirteen united colonies now rise into an Independent Republic among the kingdoms, states, and empires on earth...And have I lived to see such an important and astonishing revolution?

SAMPLE TITLES

LITERATURE	1360L	**Robinson Crusoe** (DEFOE)
	1350L	**The Secret Sharer** (CONRAD)
	1340L	**The Hunchback of Notre Dame** (HUGO)
	1340L	**The Metamorphosis**** (KAFKA)
	1340L	**Fever Pitch** (HORNBY)
INFORMATIONAL	1390L	**In Defense of Food: An Eater's Manifesto** (POLLAN)
	1380L	**Politics and the English Language**** (ORWELL)
	1370L	**Jane Austen's Pride and Prejudice** (BLOOM)
	1340L	**Walden**** (THOREAU)
	1300L	**Arctic Dreams: Imagination and Desire in a Northern Landscape** (LOPEZ)

**Common Core State Standards Text Exemplar

The Lexile Framework for Reading

1000 L ‣ 1295L
LEXILE RANGE

1200L ‣ 1295L

1200L ***Why We Can't Wait*** KING

We sing the freedom songs today for the same reason the slaves sang them, because we too are in bondage and the songs add hope to our determination that "We shall overcome, Black and white together, We shall overcome someday." I have stood in a meeting with hundreds of youngsters and joined in while they sang "Ain't Gonna Let Nobody Turn Me 'Round." It is not just a song; it is a resolve. A few minutes later, I have seen those same youngsters refuse to turn around from the onrush of a police We sing the freedom songs today for the same reason the slaves sang them, because we too are in bondage and the songs add hope to our determination that "We shall overcome, Black and white together, We shall overcome someday."

SAMPLE TITLES

LITERATURE	1280L	**The House of the Spirits** (ALLENDE)
	1270L	**Tarzan of the Apes** (BURROUGHS)
	1270L	**Chronicle of a Death Foretold** (GARCIA MARQUEZ)
	1220L	**Annie John** (KINCAID)
	1210L	**The Namesake**** (LAHIRI)
INFORMATIONAL	1290L	**A Brief History of Time** (HAWKING)
	1280L	**Black, Blue, and Gray: African Americans in the Civil War**** (HASKINS)
	1240L	**Blood Done Sign My Name** (TYSON)
	1230L	**Stiff: The Curious Lives of Human Cadavers** (ROACH)
	1200L	**The Dark Game: True Spy Stories** (JANECZKO)

1100L ‣ 1195L

1100L ***Pride and Prejudice**** AUSTEN

Lydia was a stout, well-grown girl of fifteen, with a fine complexion and good-humoured countenance; a favourite with her mother, whose affection had brought her into public at an early age. She had high animal spirits, and a sort of natural self-consequence, which the attentions of the officers, to whom her uncle's good dinners and her own easy manners recommended her, had increased into assurance. She was very equal therefore to address Mr. Bingley on the subject of the ball, and abruptly reminded him of his promise; adding, that it would be the most shameful thing in the world if he did not keep it. His answer to this sudden attack was delightful to their mother's ear.

SAMPLE TITLES

LITERATURE	1180L	**The Curious Incident of the Dog in the Night-time** (HADDON)
	1170L	**The Amazing Adventures of Kavalier & Clay** (CHABON)
	1150L	**A Wizard of Earthsea** (LE GUIN)
	1130L	**All the King's Men** (WARREN)
	1110L	**A Separate Peace** (KNOWLES)
INFORMATIONAL	1160L	**The Longitude Prize**** (DASH)
	1160L	**In Search of Our Mothers' Gardens** (WALKER)
	1140L	**Winterdance: The Fine Madness of Running the Iditarod** (PAULSEN)
	1130L	**The Great Fire**** (MURPHY)
	1100L	**Vincent Van Gogh: Portrait of an Artist**** (GREENBERG & JORDAN)

1000L ‣ 1095L

1000L ***Mythbusters Science Fair Book*** MARGLES

There may be less bacteria on the food that's picked up quickly, but playing it safe is the best idea. If it hits the floor, the next thing it should hit is the trash. If putting together petri dishes and dealing with incubation seems like a bigger project than you're ready to take on, there's a simpler way to observe bacterial growth. Practically all you need is some bread and your own two hands. Cut the edges off each slice of bread so that they'll fit into the plastic containers. Put one slice of bread into each container. Measure one tablespoon of water and splash it into the first piece of bread. Put the lid on the container and use your pen and tape to label this your control.

SAMPLE TITLES

LITERATURE	1080L	**I Heard the Owl Call My Name** (CRAVEN)
	1070L	**Savvy** (LAW)
	1070L	**Around the World in 80 Days** (VERNE)
	1010L	**The Pearl** (STEINBECK)
	1000L	**Hobbit or There and Back Again** (TOLKIEN)
INFORMATIONAL	1070L	**Geeks: How Two Lost Boys Rode the Internet Out of Idaho**** (KATZ)
	1030L	**Phineas Gage** (FLEISCHMAN)
	1020L	**This Land Was Made for You and Me: The Life and Songs of Woody Guthrie** (PARTRIDGE)
	1010L	**Travels With Charley: In Search of America**** (STEINBECK)
	1000L	**Claudette Colvin: Twice Toward Justice** (HOOSE)

**Common Core State Standards Text Exemplar

The Lexile Framework for Reading

700L ▸ 995L
LEXILE RANGE

900L ▸ 995L

900L ***We are the Ship: The Story of Negro League Baseball*** NELSON

Rube ran his ball club like it was a major league team. Most Negro teams back then weren't very well organized. Didn't always have enough equipment or even matching uniforms. Most times they went from game to game scattered among different cars, or sometimes they'd even have to "hobo"—which means hitch a ride on the back of someone's truck to get to the next town for a game. But not Rube's team. They were always well equipped, with clean, new uniforms, bats, and balls. They rode to the games in fancy Pullman cars Rube rented and hitched to the back of the train. It was something to see that group of Negroes stepping out of the train, dressed in suits and hats. They were big-leaguers.

SAMPLE TITLES

LITERATURE	980L	**Dovey Coe** (DOWELL)
	950L	**Bud, Not Buddy** (CURTIS)
	940L	**Harry Potter and the Chamber of Secrets** (ROWLING)
	940L	**Heat** (LUPICA)
	900L	**City of Fire** (YEP)
INFORMATIONAL	990L	**Seabiscuit** (HILLENBRAND)
	970L	**The Kid's Guide to Money: Earning It, Saving It, Spending It, Growing It, Sharing It**** (OTFINOSKI)
	950L	**Jim Thorpe, Original All-American** (BRUCHAC)
	930L	**Colin Powell A & E Biography** (FINLAYSON)
	920L	**Talking with Artists** (CUMMINGS)

800L ▸ 895L

800L ***Moon Over Manifest*** VANDERPOOL

There wasn't much left in the tree fort from previous dwellers. Just an old hammer and a few rusted tin cans holding some even rustier nails. A couple of wood crates with the salt girl holding her umbrella painted on top. And a shabby plaque dangling sideways on one nail, FORT TREECONDEROGA. Probably named after the famous fort from Revolutionary War days. Anything else that might have been left behind had probably been weathered to bits and fallen through the cracks. No matter. I'd have this place whipped into shape lickety-split. First off, I picked out the straightest nail I could find and fixed that sign up right. Fort Treeconderoga was open for business.

SAMPLE TITLES

LITERATURE	GN840L*	**The Odyssey** (HINDS)
	830L	**Baseball in April and Other Stories** (SOTO)
	820L	**Maniac Magee** (SPINELLI)
	820L	**Where the Mountain Meets the Moon**** (LIN)
	800L	**Homeless Bird** (WHELEN)
INFORMATIONAL	880L	**The Circuit** (JIMENEZ)
	870L	**The 7 Habits of Highly Effective Teens** (COVEY)
	IG860L*	**Animals Nobody Loves** (SEYMOUR)
	860L	**Through My Eyes: Ruby Bridges** (BRIDGES)
	830L	**Quest for the Tree Kangaroo: An Expedition to the Cloud Forest of New Guinea**** (MONTGOMERY)

700L ▸ 795L

700L ***The Miraculous Journey of Edward Tulane*** DICAMILLO

Edward, for lack of anything better to do, began to think. He thought about the stars. He remembered what they looked like from his bedroom window. What made them shine so brightly, he wondered, and were they still shining somewhere even though he could not see them? Never in my life, he thought, have I been farther away from the stars than I am now. He considered, too, the fate of the beautiful princess who had become a warthog. Why had she become a warthog? Because the ugly witch turned her into one-that was why. And then the rabbit thought about Pellegrina. He felt, in some way that he could not explain to himself, that she was responsible for what had happened to him. It was almost as if it was she, and not the boys, who had thrown Edward overboard.

SAMPLE TITLES

LITERATURE	770L	**Walk Two Moons** (CREECH)
	760L	**Hoot** (HIAASEN)
	750L	**Esperanza Rising** (RYAN)
	720L	**Nancy's Mysterious Letter** (KEENE)
	GN720L*	**Sherlock Holmes and the Adventure at the Copper Beeches** (DOYLE)
INFORMATIONAL	790L	**Be Water, My Friend: The Early Years of Bruce Lee** (MOCHIZUKI)
	760L	**Stay: The True Story of Ten Dogs** (MUNTEAN)
	IG760L*	**Mapping Shipwrecks with Coordinate Planes** (WALL)
	720L	**Pretty in Print: Questioning Magazines** (BOTZAKIS)
	720L	**Spiders in the Hairdo: Modern Urban Legends** (HOLT & MOONEY)

*GN denotes Graphic Novel, IG denotes Illustrated Guide
**Common Core State Standards Text Exemplar

The Lexile Framework for Reading

400L ▸ 695L
LEXILE RANGE

600L ▸ 695L

600L ***You're on Your Way, Teddy Roosevelt*** ST. GEORGE & FAULKNER

But from his first workout in Wood's Gymnasium he had been determined to control his asthma and illnesses rather than letting his asthma and illnesses control him. And he had. On that hot summer day in August he had proved to himself—and everyone else—that he had taken charge of his own life. In 1876 Teedie—now known as Teddy—entered Harvard College. He was on his own ...without Papa. That was all right. "I am to do everything for myself," he wrote in his diary. Why not? He was stronger and in better health than he had ever been. And ready and eager for the adventures and opportunities that lay ahead.

SAMPLE TITLES

LITERATURE	680L	**Charlotte's Web** (WHITE)
	660L	**Holes** (SACHAR)
	620L	**M.C. Higgins, the Great**** (HAMILTON)
	610L	**Mountain Bike Mania** (CHRISTOPHER)
	610L	**A Year Down Yonder** (PECK)
INFORMATIONAL	690L	**Where Do Polar Bears Live?**** (THOMSON)
	680L	**An Eye for Color: The Story of Josef Albers** (WING)
	660L	**Remember: The Journey to School Integration** (MORRISON)
	660L	**From Seed to Plant**** (GIBBONS)
	630L	**Sadako and the Thousand Paper Cranes** (COERR)

500L ▸ 595L

500L ***A Germ's Journey*** ROOKE

Excuse me! Let's blow out of this place! In real life, germs are very small. They can't be seen without a microscope. Rudy forgot to use a tissue. His cold germs fly across the room at more than 100 miles an hour. Whee! I can fly! Best ride ever! A few germs land on Ernie. But skin acts like a suit of armor. It protects against harm. The germs won't find a new home there. Healthy skin keeps germs out. But germs can sneak into the body through cuts, scrapes, or cracks in the skin. Most germs enter through a person's mouth or nose. Rudy's germs continue to fall on nearly everything in the room—including Brenda's candy.

SAMPLE TITLES

LITERATURE	560L	**Sarah, Plain and Tall** (MACLACHLAN)
	530L	**It's All Greek to Me** (SCIESZKA)
	520L	**John Henry: An American Legend** (KEATS)
	500L	**Judy Moody Saves the World** (MCDONALD)
	500L	**The Curse of the Cheese Pyramid** (STILTON)
INFORMATIONAL	IG590L*	**Claude Monet** (CONNOLLY)
	560L	**Lemons and Lemonade: A Book about Supply and Demand** (LOEWEN)
	560L	**Molly the Pony** (KASTER)
	530L	**Langston Hughes: Great American Poet** (MCKISSACK)
	510L	**A Picture for Marc** (KIMMEL)

400L ▸ 495L

400L ***How Not to Babysit Your Brother*** HAPKA

I continued to search. I checked under Steve's bed. Then I checked under my bed. I searched the basement, the garage, and my closet. There was no sign of Steve. This was going to be harder than I thought. Where was Steve hiding? CRASH! Uh-oh, I thought. I heard Buster barking in the kitchen. I ran to see what was going on. When I got there, the dog food bin was tipped over. Steve's head and shoulders were sticking out of the top. Dog food was stuck in his hair, on his clothes, and up his nose. He looked like an alien from the planet Yuck. He giggled as Buster licked some crumbs off his ear.

SAMPLE TITLES

LITERATURE	460L	**Chrysanthemum** (HENKES)
	410L	**The Enormous Crocodile** (DAHL)
	GN400L*	**Pilot And Huxley** (MCGUINESS)
	400L	**The Fire Cat**** (AVERILL)
	400L	**Cowgirl Kate and Cocoa**** (SILVERMAN)
INFORMATIONAL	480L	**Martin Luther King, Jr. and the March on Washington**** (RUFFIN)
	460L	**True Life Treasure Hunts** (DONNELLY)
	460L	**Half You Heard of Fractions?** (ADAMSON)
	420L	**Rally for Recycling** (BULLARD)
	400L	**Animals in Winter** (RUSTAD)

*GN denotes Graphic Novel, IG denotes Illustrated Guide
**Common Core State Standards Text Exemplar

The Lexile Framework for Reading

200L ▸ 400L
LEXILE RANGE

300L ▸ 395L

300L ***Princess Posey and the Next-Door Dog*** GREENE

"We have to stop now," said Miss Lee. "It's time for reading." "Ohhh..." A disappointed sound went up around the circle. "Here's what we'll do." Miss Lee stood up. "You are all very interested in dogs. So this week, you can write a story about your own dog or pet. Then you can read it to the class." Everyone got excited again. Except Posey. She didn't have a pet. Not a dog. Not a cat. Not a hamster. "Those of you who don't have a pet," Miss Lee said, "can write about the pet you hope to own someday." Miss Lee had saved the day! Now Posey had something to write about, too. Posey told her mom about Luca's puppy on the way home.

SAMPLE TITLES

LITERATURE	380L	**Martha Bakes a Cake** (BARSS)
	380L	**Junie B. Jones is (Almost) a Flower Girl** (PARK)
	360L	**Poppleton in Winter**** (RYLANT)
	340L	**Never Swipe a Bully's Bear** (APPLEGATE)
	330L	**Frog and Toad Together**** (LOBEL)
INFORMATIONAL	GN380L*	**BMX Blitz** (CIENCIN)
	380L	**Lemonade for Sale** (MURPHY)
	350L	**A Snowy Day** (SCHAEFER)
	330L	**Freedom River** (RAPPAPORT)
	300L	**From Tree to Paper** (MARSHALL)

200L ▸ 295L

200L ***Ronald Morgan Goes to Bat*** GIFF

He smacked the ball with the bat. The ball flew across the field. "Good;' said Mr. Spano. "Great, Slugger!" I yelled. ''We'll win every game. It was my turn next. I put on the helmet, and stood at home plate. "Ronald Morgan," said Rosemary. "You're holding the wrong end of the bat." Quickly I turned it around. I clutched it close to the end. Whoosh went the first ball. Whoosh went the second one. Wham went the third. It hit me in the knee. "Are you all right?" asked Michael. But I heard Tom say, "I knew it. Ronald Morgan's the worst." At snack time, we told Miss Tyler about the team.

SAMPLE TITLES

LITERATURE	280L	**Hi! Fly Guy**** (ARNOLD)
	260L	**The Cat in the Hat** (SEUSS)
	GN240L*	**Lunch Lady and the Cyborg Substitute** (KROSOCZKA)
	200L	**Dixie** (GILMAN)
	200L	**The Best Bug Parade** (MURPHY)
INFORMATIONAL	290L	**The Story of Pocahantas** (JENNER)
	250L	**Math in the Kitchen** (AMATO)
	230L	**What makes Day and Night** (BRANLEY)
	220L	**I Love Trains!** (STURGES)
	210L	**Sharks!** (CLARKE)

*GN denotes Graphic Novel

**Common Core State Standards Text Exemplar

Please note:

The Lexile measure of a book (the book's text complexity level) is an excellent starting point for a student's book selection. It's important to understand that the book's Lexile measure should not be the only factor in a student's book selection process. Lexile measures do not consider factors such as age-appropriateness, interest, and prior knowledge. These are also key factors when matching children and adolescents with books they might like and are able to read.

Lexile codes provide more information about developmental appropriateness, reading difficulty, and common or intended usage of books. For more information on Lexile codes, please visit **Lexile.com.**

MetaMetrics®
LINKING ASSESSMENT WITH INSTRUCTION

TEXT LEXILE RANGES TO GUIDE READING FOR COLLEGE AND CAREER READINESS

GRADES	CCSS LEXILE TEXT RANGE
11–12	1185L–1385L
9–10	1050L–1335L
6–8	925L–1185L
4–5	740L–1010L
2–3	420L–820L
1	190L–530L

COMMON CORE STATE STANDARDS FOR ENGLISH, LANGUAGE ARTS, REVISED APPENDIX A, NGA AND CCSSO, 2012