Linking the

State of Texas Assessments of Academic Readiness

Reading, English I, and English II with

The Lexile® Framework for Reading

February 2018
Redacted

Prepared by MetaMetrics for:

Texas Education Agency
William B. Travis Building
1701 N. Congress Avenue
Austin, Texas, 78701

MetaMetrics *Linking Assessment with Instruction*

LEXILE® QUANTILE®

Linking the

State of Texas Assessments of Academic Readiness

Reading, English I, and English II with

The Lexile® Framework for Reading

Technical Report
Redacted

Table of Contents

List of Tables

List of Figures

Introduction

Often it is desirable to convey more information about test performance than can be incorporated into a single primary score scale. Two examples arise in large-scale assessment. In one situation, one test can provide a unique type of information,such as national comparisons available from NationalAssessment of Educational Progress (NAEP), but is not administered very often. At the same time another test is administered more often, but is not able to provide the breadth of information (such as a state assessment). An auxiliary score scale for a test can be established to provide this additional information through assessment scale linkages. Once linkages are established between the two assessments, then the results of the more-frequently-administered assessment can be translated in terms of the scale for the other assessment.

In another situation, the linkage between two score scales can be used to provide a context for understanding the results of one of the assessments. For example, sometimes it is hard to explain what a student can read based on the results of a reading comprehension test. Parents typically ask the questions "If my child is in the fourth grade and scores a 1480 on the STAAR Reading assessment, what does this mean?" or "Based on my child's test results, what can he or she read and how well?" or "Is my child well prepared to meet the reading demands of grade level materials?" Once a linkage is established with an assessment that is related to specific book or text titles, then the results of the assessment can be explained and interpreted in the context of the specific titles that a student should be able to read.

Auxiliary score scales can be used to "convey additional normative information, test-content information, and information that is jointly normative and content based. For many test uses, an auxiliary scale conveys information that is more crucial than the information conveyed by the primary score scale. In such instances, the auxiliary score is the one that is focused on, and the primary scale can be viewed more as a vehicle for maintaining interpretability over time" (Petersen, Kolen, and Hoover, 1989, p. 222). One such auxiliary scale is The Lexile® Framework for Reading, which was developed to appropriately match readers with text at a level that provides challenge but not frustration.

Linking assessment results with the Lexile Framework provides a mechanism for matching each student's reading ability with text on a common scale. These scales serve as anchors to which texts and assessments can be connected, allowing parents, teachers, and administrators to speak the same language. In addition, the Lexile Framework provides a common way to monitor if students are "on track" for the reading demands of various postsecondary endeavors. By using the Lexile Framework, the same metric is applied to the books students read, the tests they take, and the results that are reported. Parents often ask questions like the following:

- How can I help my child become a better reader?
- How do I challenge my child to read so that she is ready for various college and career options?

Questions like these can be challenging for parents and educators. By linking STAAR Reading (for Grades 3-8), English I, and Englsih II with The Lexile Framework for Reading, educators

and parents will be able to answer these questions and will be better able to use the results from the test to improve instruction and to develop each student's level of reading comprehension.

This research study was designed to determine a mechanism to provide reading levels that can be matched to text based on the STAAR Reading, English I, and English II results. The study was conducted by MetaMetrics in collaboration with the Texas Education Agency and Educational Testing Service (contracts dated September 30, 2015; September 1, 2016; and September 1, 2017). The Texas Education Agency facilitated the exchange of data between MetaMetrics and Educational Testing Service.

The following are the primary purposes of this study:

- present a solution for matching readers with text;
- provide Texas with Lexile measures on the STAAR Reading, English I, and English II assessments;
- develop tables for converting STAAR scale scores to Lexile measures; and
- produce a report that describes the linking analysis procedures.

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.

Lexile text measures for texts and readers typically range from above 200L to below 1600L but measures can be below 0L for beginning reader materials (e.g., BR150L) to above 2000L for advanced materials. Within any single 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.

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 text is dominated by the familiarity of the semantic units and by the complexity of the syntactic structures used in constructing the text. The Lexile Framework utilizes these two dominant features of language in measuring text complexity by examining the characteristics of word frequency and sentence length. In addition to these features, when measuring early reader texts, the Lexile Framework utilizes characteristics found to be important to the complexity of early reader text such as word decodabilty and patterning and repetition.

Variables that Affect the Text Complexity of Upper Level Text

Semantic Component. Most operationalizations of the semantic component 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 strong 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 first 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). For example, the word "accident" appears 176 times in the 5,088,721-word corpus. The second word frequency measure used was the frequency of the "word family." A word family included: (1) the stimulus word; (2) all plurals (adding "-s" or "-es" 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. For example, the word family for "accident" would include "accidental," "accidentally," "accidentals," and "accidents," and they would all have the same word frequency of 334. The frequency of a word family was based on the sum of the individual word frequencies from each of the types listed.

Correlations were computed between algebraic transformations of these means (mean frequency of the words in the test item and mean frequency of the word families in the test item) 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 log of the mean word frequency provided the strongest correlation with item rank order ($r = -0.779$) for the items on the combined form.

The Lexile Framework currently employs a 1.4 billion-word corpus when examining the semantic component of text. This corpus was assembled from the more than 90,000 texts that were measured by MetaMetrics for publishers from 1998 through 2012.

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

While sentence length has been shown to be a powerful proxy for the syntactic complexity of a passage, an important caveat is that sentence length is not the underlying causal influence (Chall, 1988). Researchers sometimes incorrectly assume that manipulation of sentence length will have

a predictable effect on passage difficulty. Davidson and Kantor (1982), for example, illustrated rather clearly that sentence length can be reduced and difficulty increased and vice versa.

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

Variables that Affect the Text Complexity of Early Reader Texts

Texts designed for early readers are distinct from texts designed for more accomplished readers because they are usually designed specifically to facilitate early readers' progress. For all readers, making meaning of the texts is always the focus, but for early readers, developing an understanding of how to "crack the code" requires specific attention. Early readers must develop the ability to hear sounds in words, develop sight words, and acquire word recognition strategies (Fitzgerald and Shanahan, 2000) as they develop the comprehension and fluency characteristic of more advanced readers. A number of studies support the finding that the presence of specific text features support the development of skills associated with code cracking. For example, word repetition reinforces sight-word learning and development of the sounds associated with spelling patterns (e.g., Vadasy, Sanders, & Peyton, 2005). Repeated phrases also reinforce scaffolding development of a variety of word recognition strategies (e.g., Ehri & McCormick, 1998). The use of words familiar in oral language enhances readers' ability to make meaning from words and permits more attention to word recognition (e.g., Muter, Hulme, Snowling, & Stevenson, 2004). Inclusion of several types of text-characteristic support may further support students' growth as readers. Research suggests that to appropriately describe early reader text complexity it is necessary to consider several text characteristics at multiple linguistic levels (Graesser & McNamara, 2011; Graesser, McNamara, & Kulikowich, 2011; Kintsch, 1998; and Snow, 2002). In general, levels of text characteristics include word level (e.g., word structure, word frequency), within-sentence level (e.g., syntax), and across-sentence/discourse level (e.g., referential cohesion). The research base supporting the importance of multiple levels of text characteristics for early phases of learning to read is extensive (Mesmer, Cunningham, & Hiebert, 2012) and has identified the importance of considering the impact of interaction between the features (Merlini Barbaresi, 2003; and Biber, 1988).

In order to determine which text characteristics had the greatest impact on text complexity for early readers, MetaMetrics identified twenty-two unique text characteristics at four linguistic levels: sounds-in-words, words (structure and meaning), within-sentence syntax, and across-sentence/discourse.

- *Sounds-in-Words*: number of phonemes in words, phonemic Levenshtein Distance, and mean internal phonemic predictability
- *Word Structure*: decoding demand, orthographic Levenshtein Distance, number of syllables in words, and mean internal orthographic predictability
- *Word Meaning*: age of acquisition, abstractness, and word rareness
- *Within-Sentence Syntax*: sentence length and grammar
- *Across-Sentence/Discourse*: linear edit distance, linear word overlap, cohesion triggers, type-token ratio, longest common string, edit distance, Cartesian word overlap, information load, and compression ratio

From these characteristics, 238 operationalizations were developed to capture the varied ways in which the characteristics could be quantified in terms of their presence in the text. Three hundred and fifty early reader texts designed for readers in Kindergarten through Grade 2 were selected to represent the range of text types early readers are likely to encounter. These included decodable books, phonics readers, leveled books, high-frequency readers, and various trade books. Two separate sub-studies were conducted to determine the relative challenge of the texts. One study collected primary-grade educators' ratings of the complexity of the 350 texts and the other gathered Grade 1 and 2 students' responses to a subset of 89 texts from the full set of 350 study texts. From these studies a text-complexity logit scale was created so that each text could be assigned a measure (Fitzgerald, Elmore, Koons, Hiebert, Bowen, Sanford-Moore & Stenner, 2014).

Calibration of Text Difficulty of Upper Level Texts

The research study on semantic units (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* (Dunn and Markwardt, 1970). 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 8 other standardized tests. The resulting correlation between the observed logit difficulties and the theoretical calibrations across the 9 tests was 0.93 after correction for range restriction and measurement error.

Once a regression equation is established linking the syntactic and semantic features of text to the difficulty of text, the equation can be used to calibrate test items and text. The result of the research was a regression equation linking the syntactic and semantic features of text to the difficulty of text. This equation can now be used to calibrate test items and text within the Lexile Framework.

The Lexile Scale

In developing the Lexile Scale, the Rasch model (Wright and Stone, 1979) was used to estimate the difficulties of the items and the abilities of the 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 group it can be determined which item is harder and which one is easier. This ordering should hold when the same two items are administered to a second group. If two different items are administered to the second group, there is no way to know which set of items is harder and which set is easier. The problem is that the location of the scale is not known. General objectivity requires that scores obtained from different test administrations be tied to a common zero—absolute location must be sample independent (Stenner, 1990). To achieve general objectivity, the theoretical logit difficulties must be transformed to a scale where the ambiguity regarding the location of zero is resolved.

The first step in developing a scale with a fixed zero was to identify two anchor points for the scale. The following criteria were used to select the two anchor points: they should be intuitive, easily reproduced, and widely recognized. For example, with most thermometers the anchor points are the freezing and boiling points of water. For the Lexile Scale, the anchor points are text from seven basal primers for the low end and text from *The Electronic Encyclopedia* (Grolier, Inc., 1986) for the high end. These points 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 (a Lexile) 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 by definition equals $1/1000^{th}$ of the difference between the difficulty of the primers and the difficulty 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 low-end anchor point and then solving the system of equations.

The Lexile Scale ranges from below 200L to above 1600L. There is 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 1500L+, provides a context for understanding reading comprehension (see the Appendix).

Calibration of Text Difficulty of Early Reader Texts

To bring the observed difficulties (logit scores) of early reader texts from the two studies previously described onto the Lexile scale, a theory-based linking procedure was conducted. First, Lexile text measures were calculated based only on the syntactic and semantic features of the text as done with upper level texts. Next, for approximately 10% of the texts the discrepancy between the observed difficulty and the theoretical Lexile measure was large and the texts were flagged and not used in subsequent analyses. Finally, using the remaining 90% of the texts in the study, a linear linking function (SD line) was calculated. In linear linking, a transformation is chosen such that scores on two sets of data are considered to be linked if they correspond to the same number of standard deviations above (or below) the mean in some group of data elements (Angoff, 1984, cited in Petersen, Kolen, and Hoover, 1989; Kolen and Brennan, 2014). The result of the linear linking function was that the early reader observed difficulties were transformed to Lexile measures while still maintaining the relative ordering of the difficulty of the texts derived from the educator judgments and student performances.

Once observed Lexile measures were calculated, random forest regression technique was employed to evaluate the importance of the 238 operationalizations of characteristics that research suggests affects text complexity of early reader texts. This process was conducted in several stages and is described in detail by Fitzgerald and Elmore and their colleagues (2014). The first step in the analysis was to set baseline performance. Eighty percent of the texts were selected for this training process and twenty percent were held as a validation sample. Three separate random forest regressions were conducted, one each for: (1) the 80% of the 350 texts that the teachers ordered ($n = 279$); (2) the 80% of the texts that the students were presented ($n = 71$), and (3) the two sets of texts combined ($N = 350$). Each random forest regression produced importance values for each of the 238 variables in relation to the text-complexity logit scale. The next step in the analysis involved an iterative variable-selection procedure in which the variables with the smallest importance values were systematically removed and the effect on the model calculated. This process determined whether fewer variables could predict text complexity as well or nearly as well as the 238-variable model. The result was a set of nine variables:

- word level: monosyllable decoding, syllable count, age of acquisition, and word rareness, and abstractness;
- within-sentence and across-sentence/discourse level: intersentential complexity, phrase diversity, non-compressibility, and text density.

Finally, a final set of three random forest regression models was trained using the nine variables with the teacher text set, the student text set, and the two text sets combined. The resulting correlations for the teacher, student, and combined models were 0.89, 0.71, and 0.88, respectively. The validation samples, 20% of the teacher texts ($n = 71$) and 20% of the student texts ($n = 19$), were combined and a final random forest regression was run with the nine selected

variables as predictors. The model was validated with a correlation of 0.85 and RMSE of 9.68. The final model is now used to calibrate texts intended for early-readers.

The nine variables have been grouped into four early-reading indicators based on the linguist level addressed:

- Decoding Demand includes syllable count and monosyllable decoding demand;
- Semantic Demand includes abstractness, word rareness, and age of acquisition;
- Syntactic Demand includes intersentential complexity; and
- Structure Demand includes non-compressibility, phrase diversity, and text density.

The Enhanced Lexile Analyzer®

When text is analyzed by MetaMetrics, all electronic files are initially edited according to established guidelines used with the enhanced Lexile Analyzer software. These guidelines include the removal of all incomplete sentences, chapter titles, and paragraph headings; running of a spell check. The text is then submitted to the enhanced Lexile Analyzer that examines the lengths of the sentences and the frequencies of the words for upper-level texts and the nine early-reader variables for lower-level texts. The enhanced Lexile Analyzer first looks at the text features of a piece of text and attempts to determine if the text is written for early readers (early reader texts) or for more advanced readers (upper level texts). Based on the results of the examination, the enhanced Lexile Analyzer applies the most appropriate word and sentence/discourse variables to the measurement process. The enhanced Lexile Analyzer then reports a Lexile measure for the text. If the Lexile measure of the text is 650L or below, then the four early-reading indicators are also reported.

Validity of 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) states 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 ability and reading comprehension.

Lexile Framework and other Measures of Reading Comprehension. The Lexile Framework 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 a number of linking studies conducted with the Lexile Framework. In these studies, students were administered a Lexile assessment and another assessment of reading comprehension. There is a strong relationship between reading comprehension ability as measured by the Lexile Framework and reading comprehension ability as measured by other assessments. 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 information, the student can receive a reading list consisting of texts (books and articles) targeted to his specific reading level.

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

Standardized Test	Grades in Study	N	Correlation Between Test Score and Lexile Measure
Gates-MacGinitie Reading Test	2, 4, 6, 8, 10	4,644	0.90
Metropolitan Achievement Test (8th ed.)	2, 4, 6, 8, 10	2,382	0.93
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 linked; 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 indicates that the text is less complex than text with a higher Lexile measure. 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 difficulty. So, for example, the latter portion of a third-grade reader is presumably more difficult than the first portion of the same book. Likewise, a fourth-grade reader is presumed to be more difficult than a third-grade reader. 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*).

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

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

r_{OT} = raw correlation between observed difficulties (O) and theory-based calibrations (T).
R_{OT} = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction.
$R´_{OT}$ = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction and measurement error.
*Mean correlations are the weighted averages of the respective correlations.

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

Lexile Framework and Fountas & Pinnell Reading Levels. Koons, Elmore, Sanford-Moore, and Stenner (2017) explored the relationship between Fountas & Pinnell reading levels for a set of texts A through M (i.e. Kindergarten through Grade 2) and the their corresponding Lexile measures to obtain construct validity evidence for the measurement of early reader texts. The Spearman correlation coefficient between the two text sets was $r_s = 0.84$, indicating a strong positive relationship. Because Fountas & Pinnell reading levels are "larger grained" than the Lexile measures, some variation of Lexile measures within each Fountas & Pinnell reading level was expected. *Figure 1* shows a series of box-and-whisker plots of the results. The box in each box-and-whisker plot depicts the IQR with the bottom of the box at the 25th percentile of the distribution of Lexile measures, the line between the shaded portions at the median (50th percentile), and the top of the box at the 75th percentile. The bottom whisker depicts the text measure at the 5th percentile of the distribution and the top whisker depicts the text measure at the 95th percentile. *Figure 1* shows steadily increasing Lexile text measures across Fountas & Pinnell reading levels for each represented percentile except the 95th percentile of Level C (351L), which has a greater value than the 95th percentile of the two following levels (D: 288L; and E: 350L).

Figure 1. *Progression of Lexile text measures and Fountas & Pinnell reading levels, Levels A through M.*
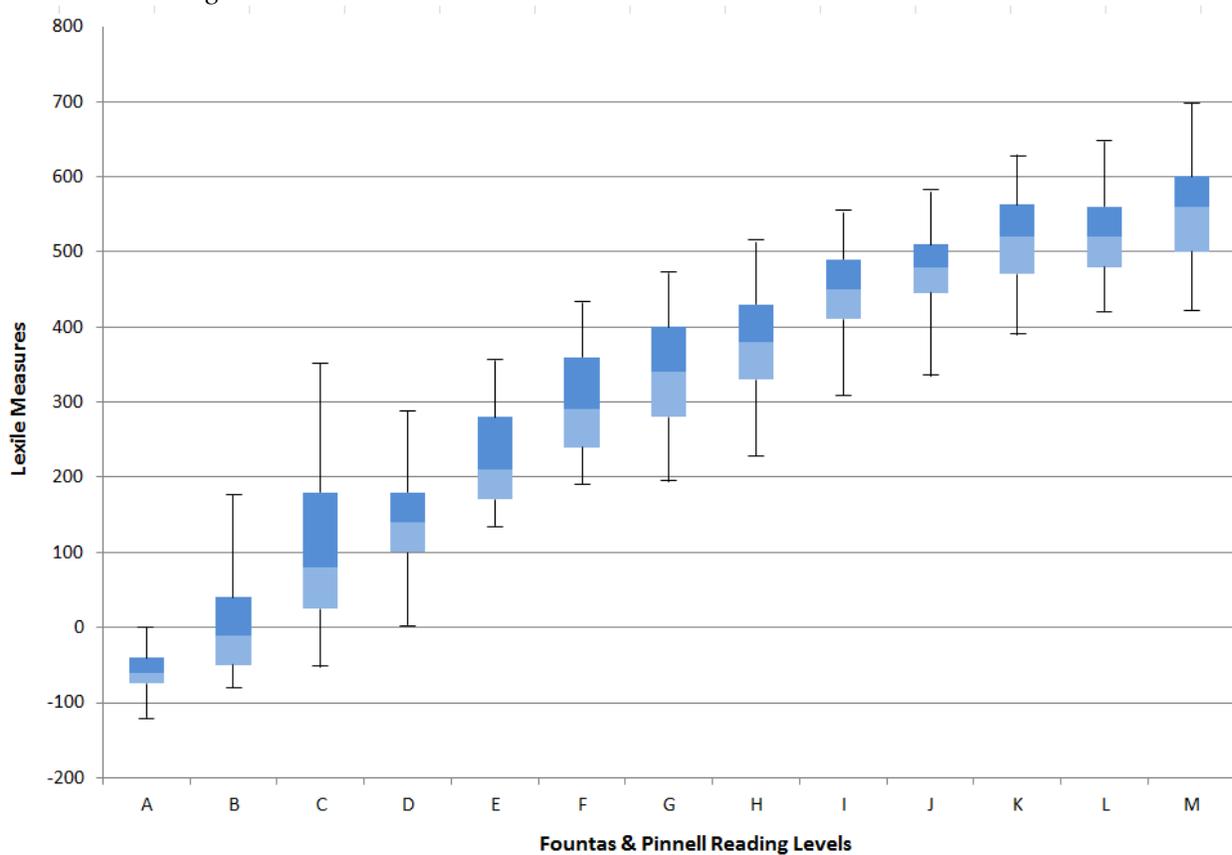
Lexile Framework and the Difficulty of Reading Test Items. Additional construct validity evidence was obtained by exploring the relationship between Lexile calibrations of item difficulties and actual item difficulties of reading comprehension tests. In a study conducted by Stenner, Smith, Horabin, and 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 publishers 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; 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 Lexile equation did not fit poetry items or noncontinuous 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 noncontinuous prose items were removed and correlations were recalculated. *Table 3* contains the results of this analysis.

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

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

r_{OT} = raw correlation between observed difficulties (O) and theory-based calibrations (T).
R_{OT} = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction.
$R´_{OT}$ = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction and measurement error.
**Means are computed on Fisher Z transformed correlations.

The last three columns in *Table 3* show the raw correlation between observed (O) item difficulties and theoretical (T) item calibrations, with the correlations corrected for restriction in range and measurement error. The Fisher Ƶ mean of the raw correlations (r_{OT}) is 0.84. When corrections are made for range restriction and measurement error, the Fisher Ƶ 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 show that most attempts to measure reading comprehension, no matter what the item form, type of skill or objectives assessed, or item type used, measure a common comprehension factor specified by the Lexile theory.

Text Measure Error Associated with the Lexile Framework

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). The Lexile Analyzer automates this process, but what "certainty" can be attached to each text measure?

Using a bootstrap procedure to examine error due to the text samples, the above analysis could be repeated (Efron, 1981; Sitter, 1992). 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. To examine this source of error, 200 items that had been previously calibrated and shown to fit the model were administered to 3,026 students in Grades 2 through 12 in a large urban school district. For each item the observed item difficulty calibrated from the Rasch model was compared with the theoretical item difficulty calibrated from the regression equation used to calibrate texts. A scatter plot of the data is presented in *Figure 2*.

Figure 2.   Scatter plot between observed item difficulty and theoretical item difficulty.
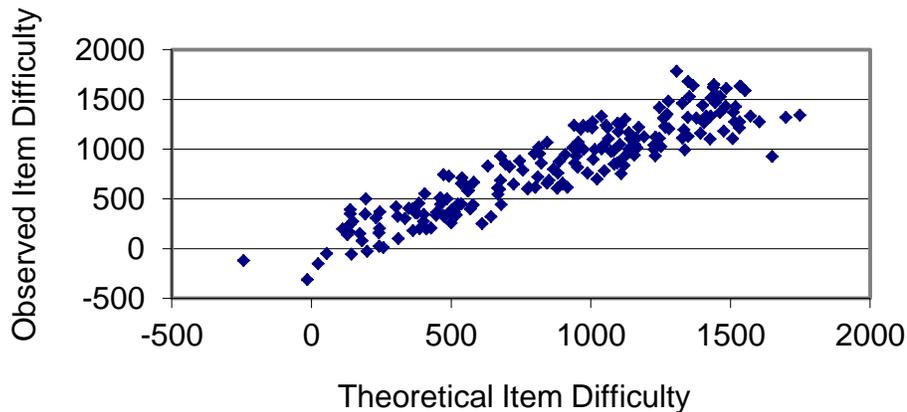
The correlation between the observed and the theoretical calibrations for the 200 items was 0.92 and the root mean square error was 178L. Therefore, for an individual slice of text the measurement error is 178L.

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

Table 4.   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

A typical Grade 3 reading test has approximately 2,000 words in the passages. To calibrate this text, it would be sliced into 16 125-word passages. The error associated with this text measure would be 45L. A typical Grade 7 reading test has approximately 3,000 words in the passages and

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

The Find A Book (fab.lexile.com) contains information about each book analyzed: author, Lexile measure and Lexile Code, awards, ISBN, and developmental level as determined by the publisher. Information concerning the length of a book and the extent of illustrations—factors that affect a reader's perception of the difficultly of a book—can be obtained from MetaMetrics.

Lexile Item Bank

The Lexile Item Bank contains over 10,000 items that have been developed since 1986 for research purposes with the Lexile Framework.

Passage Selection. Passages selected for use are selected from "real world" reading materials that students may encounter both in and out of the classroom. Sources include textbooks, literature, and periodicals from a variety of interest areas and material written by authors of different backgrounds. The following criteria are used to select passages:

- the passage must develop one main idea or contain one complete piece of information;
- understanding of the passage is independent of the information that comes before or after the passage in the source text; and
- understanding of the passage is independent of prior knowledge not contained in the passage.

With the aid of a computer program, item writers develop native-Lexile items by examining blocks of text (minimum of three sentences) that are calibrated to be within 100L of the source text. From these blocks of text item writers are asked to select four to five that could be developed as items. If it is necessary to shorten or lengthen the passage in order to meet the criteria for passage selection, the item writer can immediately recalibrate the text to ensure that it is still targeted within 100L of the complete text (source targeting).

Item Format. The native-Lexile item format is an embedded completion task. The embedded completion format is similar to the fill-in-the-blank format. When properly written, this format directly assesses the reader's ability to draw inferences and establish logical connections between the ideas in the passage (Haladyna, 1994). The reader is presented with a passage of approximately 30 to 150 words in length. The passages are shorter for beginning readers and longer for more advanced readers. The passage is then response illustrated (a statement is added at the end of the passage with a missing word or phrase followed by four options). From the four presented options, the reader is asked to select the "best" option that completes the statement. With this format, all options are semantically and syntactically appropriate completions of the sentence, but one option is unambiguously the "best" option when considered in the context 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. When the embedded completion statement is read by itself, each of the four options is plausible.

Item Writer Training. Item writers are classroom teachers and other educators who have had experience with the everyday reading ability of students at various levels. The use of individuals with these types of experiences helps to ensure that the items are valid measures of reading comprehension. Item writers are provided with training materials concerning the embedded completion item format and guidelines for selecting passages, developing statements, and selecting options. The item writing materials also contain incorrect items that illustrate the criteria used to evaluate items and corrections based on those criteria. The final phase of item writer training is a short practice session with three items. Item writers are provided vocabulary lists to use during statement and option development. The vocabulary lists were compiled from spelling books one grade level below the level where the item would typically be used. The rationale was that these words should be part of a reader's "working" vocabulary since they had been learned the previous year.

Item writers are also given extensive training related to "sensitivity" issues. Part of the item writing materials address these issues and identify areas to avoid when selecting passages and developing items. The following areas are covered: violence and crime, depressing situations/death, offensive language, drugs/alcohol/tobacco, sex/attraction, race/ethnicity, class, gender, religion, supernatural/magic, parent/family, politics, animals/environment, and brand names/junk food. These materials were developed based on material published by McGraw-Hill (*Guidelines for Bias-Free Publishing,* 1983). This publication discusses the equal treatment of the sexes, fair representation of minority groups, and the fair representation of disabled individuals.

Item Review. All items are subjected to a two-stage review process. First, items are reviewed and edited by an editor according to the 19 criteria identified in the item writing materials and for sensitivity issues. Approximately 25% of the items developed are deleted for various reasons. Where possible items are edited and maintained in the item bank.

Items are then reviewed and edited by a group of specialists that represent various perspectives—test developers, editors, and curriculum specialists. These individuals examine each item for sensitivity issues and for the quality of the response options. During the second stage of the item review process, items are either "approved as presented," "approved with edits," or "deleted." Approximately 10% of the items written are "approved with edits" or "deleted" at this stage. When necessary, item writers receive additional on-going feedback and training.

Item Analyses. As part of the linking studies and research studies conducted by MetaMetrics, items in the Lexile Item Bank are evaluated in terms of difficulty (relationship between logit [observed Lexile measure] and theoretical Lexile measure), internal consistency (point-biserial

correlation), and bias (ethnicity and gender where possible). Where necessary, items are deleted from the item bank or revised and recalibrated.

During the spring of 1999, 8 levels of a Lexile assessment were administered in a large urban school district to students in Grades 1 through 12. The 8 test levels were administered in Grades 1, 2, 3, 4, 5, 6, 7-8, and 9-12 and ranged from 40 to 70 items depending on the grade level. A total of 427 items were administered across the 8 test levels. Each item was answered by at least 9,000 students (the number of students per level ranged from 9,286 in grade 2 to 19,056 in grades 9-12). The item responses were submitted to a Winsteps Rasch 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.

The STAAR Reading, English I, and English II—
Lexile Framework Linking Process

Description of the Assessments

State of Texas Assessments of Academic Readiness (STAAR™). The Texas Education Agency (TEA), in collaboration with the Texas Higher Education Coordinating Board (THECB) and Texas educators, developed the STAAR program in response to requirements of Senate Bill (SB) 1031 and House Bill (HB) 3 enacted by the 80th and 81st Texas legislative sessions, respectively. Implemented in the school year 2011-2012, the STAAR assessments measure the extent to which students have learned and are able to apply the skills and knowledge described by the state-mandated curriculum standards known as the Texas Essential Knowledge and Skills (TEKS; TEA, 2015). Students are tested in mathematics and reading in Grades 3 through 8, writing in Grades 4 and 7, science in Grades 5 and 8, and social studies in Grade 8. STAAR end-of-course (EOC) assessments are required for students enrolled in Algebra I, Algebra II, English I, English II, English III, Biology, and U.S. History.

Readiness standards are a subset of the TEKS and are identified as being essential for success in the current grade level or course, important for preparedness for the next grade level or course, and supportive of postsecondary readiness. Measurement of the readiness standards is emphasized on the STAAR assessments. Other TEKS are considered to be supporting standards and are assessed on the STAAR assessments, though not emphasized.

The STAAR Reading, English I, and English II assessments are administered annually and are used for local and federal accountability purposes. Three reading categories represent the content and skills assessed by STAAR Reading for Grades 3-8, English I, and English II:

1. Understanding/analysis across genres (Reading),
2. Understanding/analysis of literary texts (Reading), and
3. Understanding/analysis of informational texts (Reading).

In addition to receiving scores for each reporting category, students in Grades 3-8 receive an overall reading score. This overall reading score was used to link with the Lexile Framework. In addition to the reading component reporting categories, the STAAR English I and English II assessments incorporate three writing components: composition, revision, and editing. Students receive scores in all six reporting categories, in addition to a combined overall scale score. For English I and II, the overall score was used to produce the link with the Lexile Framework.

STAAR Reading, English I, and English II assessments and blueprints were examined for two time periods for the linking study. Grades 3 through 5 test blueprints were reviewed for the 2015-2016 school year, and Grades 6 through 8, English I, and English II test blueprints were reviewed for the 2016-2017 school year. The Grades 3 through 8 reading assessments contained from 40 to 46 operational multiple-choice items, depending on the grade. Both the English I and English II assessments contained 52 operational multiple-choice items and one writing

composition (TEA, 2016d). The STAAR Reading Grades 3 through 8 assessments are administered in a paper/pencil format. The STAAR English I and English II assessments are administered in both paper/pencil and online formats.

The STAAR program uses the Rasch model to place items on the same scale across administrations (TEA, 2015). Under Texas Education Code (TEC) §39.036, TEA was required to develop a vertical scale for monitoring student performance in Grades 3-8 reading. The vertical scale allows for the comparison of student scores across grade levels. Scale scores range from approximately 600 to 2300. STAAR English I and English II scores are reported on horizontal scales which range from approximately 1000 to 6000.

The Lexile Framework for Reading. The Lexile Framework is a tool built to help teachers, parents, and students locate appropriate reading materials. Text complexity and reader ability are measured in the same unit—the Lexile. Text complexity is determined by examining word frequency and sentence length characteristics. Items and text are calibrated using the Rasch model. Lexile measures typically range from 200L to 1600L. Actual Lexile measures can range from below zero (BR) to above 2000L.

The Lexile Framework measures reading ability by focusing on skills readers use when studying written materials sampled from various content areas. Each test item consists of a passage that is response-illustrated (a statement is added at the end of the passage with a missing word or phrase followed by four options, or distractors). The skills measured by these items include referring to details in the passage, drawing conclusions, and making comparisons and generalizations. Lexile items do not require prior knowledge of ideas outside of the passage, vocabulary taken out of context, or formal logic.

Lexile Linking Test forms were developed for administration to students in Grades 3 through 8, English I, and English II. For Grades 3 through 8, characteristics of the linking test forms were as similar as possible to the STAAR Reading assessment, including the number of operational items per test and difficulty of the items. The Grades 3, 4, and 5 linking test forms contained 40, 44, and 46 items, respectively. The Grades 6, 7, and 8 linking test forms contained 40, 42, and 44 items, respectively. The English I and English II Lexile Linking Test forms each contained 48 items.

The items for the linking test forms were chosen to optimize the match to the target test (e.g., STAAR Reading). The IRT difficulty values associated with STAAR assessments were provided to MetaMetrics and converted to Lexile measures using a computer program developed by MetaMetrics (no date). Each linking test form had a mean Lexile measure established through analysis of the difficulties of the passages on the STARR assessments, normative grade-level means, and the item difficulties of the STAAR assessments (Spring 2016 item difficulties were used for Grades 3-4, 6-8, English I, and English II; Spring 2015 item difficulties were used for Grade 5). The Lexile Linking Test form difficulty means were as follows: Grade 3, 671L; Grade 4, 802L; Grade 5, 972L; Grade 6, 1034L; Grade 7, 1124L; Grade 8, 1136L; English I, 1216L; and English II, 1235L.

Evaluation of Lexile Linking Test Items. After administration, the Lexile linking items were reviewed. *Table 5* presents the raw score descriptive statistics for the Lexile Linking Test forms. The minimum possible score on each assessment was 0 and the maximum possible raw score ranged from 39 to 48 depending on the test level. A total of 11,301student answer sheets were returned to MetaMetrics; 723 were removed in Grade 5 because answer sheet analysis revealed that a portion of the students failed to finish the Lexile Linking Test, and 59 students across all assessments were removed due to invalid testing results. All remaining item responses were submitted to a Winsteps analysis (Linacre, 2011). Subsequently, 341 students were removed because of misfit to the Rasch model, leaving a total of 10,178 students across all test levels.

Table 5.   Descriptive statistics for the Lexile Linking Test forms.

Test Level	N* (Persons)	Raw Score Mean (SD)	Minimum Score		Maximum Score	
			Observed	Possible	Observed	Possible
3	806	25.34 (11.3)	0	0	39	39
4	1,027	28.19 (11.9)	0	0	44	44
5	1,818**	31.92 (10.9)	0	0	46	46
6	1,318	25.28 (9.5)	0	0	40	40
7	1,500	25.48 (10.6)	0	0	42	42
8	1,510	24.83 (11.3)	0	0	44	44
English I	866	27.97 (11.0)	1	0	48	48
English II	1,333	33.08 (11.2)	0	0	48	48

 *N (Persons) reflects the removal of 341 students due to misfit to the Rasch model
 **Reflects the removal of 723 students who did not finish the linking test.

All items were reviewed and evaluated for use in the linking study based on item difficulty or potential alternate answer choices being more attractive than the correct answer choice (i.e. low point-biserial). One item in Grade 3 was flagged for removal based on this criterion. As shown in *Table 6*, the item statistics for the remaining linking test items were in an acceptable range and were retained in the subsequent analyses. Additionally, the coefficient alpha correlations were evaluated for signs of internal test consistency. The estimated reliability for each of the linking test forms ranged from 0.92 to 0.94. This indicates strong internal consistency reliability for each test and high consistency across all eight linking tests.

Table 6.   Item statistics from the administration of the Lexile Linking Tests.

Test Level	N* (Persons)	N** (Items)	Percent Correct Mean (Range)	Point-Biserial Range	Coefficient Alpha
3	806	39	73 (30 to 99)	0.15 to 0.58	0.92
4	1,027	44	70 (36 to 96)	0.27 to 0.48	0.93
5	1,818	46	73 (43 to 97)	0.26 to 0.57	0.92
6	1,318	40	64 (39 to 87)	0.30 to 0.57	0.92
7	1,500	42	41 (33 to 48)	0.23 to 0.64	0.94
8	1,510	44	58 (27 to 87)	0.30 to 0.61	0.94
English I	856	48	.58 (.33 - .85)	0.23 to 0.57	0.93
English II	1,317	48	.69 (.33 - .92)	0.25 to 0.62	0.94

*N (Persons) reflects the removal of 341 students for misfit to the Rasch model and 723 students in Grade 5 who did not finish the linking test.
**N (Items) reflects removal of 1 item for poor performance.

Study Design

A single-group/common-person design was chosen for this study (Kolen and Brennen, 2014). This design is most useful "when (1) administering two sets of items to examinees is operationally possible, and (2) differential order effects are not expected to occur" (pp. 16–17).

The Grades 3 and 4 Lexile Linking Tests were administered between April 11, 2016 and April 18, 2016 to students shortly after their Grades 3 and 4 Spring 2016 STAAR administration. The Grade 5 Lexile Linking Test was administered between November 30, 2015 and December 4, 2015 to Grade 6 students and matched with their Spring 2015 Grade 5 STAAR administration. Grades 6 through 8 Lexile Linking Tests were administered between January 9, 2017 and January 24, 2017 to students in Grades 7, 8, and 9 and matched with their Spring 2016 STAAR Reading assessment administration for Grades 6 through 8. English I and English II Lexile Linking Tests were administered between September 6, 2017 and October 27, 2017 to students who had completed the Englsih I or English II courses and matched with their Spring 2017 STAAR English I and English II assessment administrations.

Description of the Sample

Educational Testing Service (ETS) selected the sample of students for the study. The Grades 3 through 8 students were located in 14 districts and 70 schools across Texas; the English I and English II students were located in 2 districts and 16 schools across Texas.

Table 7 presents the number of students tested in the linking study and the percentage of students with complete data (i.e. both a STAAR scale score and a Lexile Linking Test record), referred to as the matched sample. For the Grades 3 thorugh 8 STAAR Reading linking study, a total of 9,333 students had valid tests scores on the STAAR Reading assessment and 9,043 students had

valid scores on the Lexile Linking Tests. A total of 7,771 students had valid test scores on both the STAAR Reading and the Lexile Linking Test, or 85.93% of the Llinking test student sample.

For the STAAR English I and English II study, a total of 3,052 students had valid STAAR English I and English II scores. After 194 students were removed for text-to-speech accommodations from the initial sample, a total of 2,858 studenst remained. The matched sample for the linking study consisted of 827 English I students (95.50% of the linking test sample) and 1,310 English II students (98.27% of the linking test sample).

Table 7.   Number of students sampled and number of students in the matched sample.

Test Level	STAAR Initial Samples	Lexile Linking Test Samples	Matched STAAR to Lexile Linking Test Samples*	Percent of Linking Test Samples Retained in the Matched Samples
3	1,019	904	865	95.69
4	1,278	1,119	1,071	95.71
5	2,482	2,632	1,781	67.67
6	1,482	1,338	1,223	91.41
7	1,507	1,504	1,368	90.96
8	1,565	1,546	1,463	94.63
English I	1,150	866	827	95.50
English II	1,708	1,333	1,310	98.27

 * Reflects the removal of 723 students in Grade 5 who did not finish the linking test.

All students in the matched sample and remaining items were submitted to a Winsteps (Linacre, 2011) analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.003. This process helped identify students misfitting the Rasch model in the Winsteps analysis.

To account for individual differences in motivation when responding to the two assessments, the sample was trimmed. Test scores from the matched sample for each assessment were rank ordered and converted to percentiles. For each student, the difference in percentiles between the STAAR and the Lexile Linking Test assessments was examined. A screen of a 30-percentile-point difference was selected for Grades 3 through 8 tests, and a 25-percentile-point difference was selected for the English I and English II tests. This process helped minimize the number of students removed from the sample and maintain distributional characteristics, while removing students that were obvious outliers on one or both of the assessments.

To identify the final sample of students, students in the matched sample with the following characteristics were removed:

- Texas STAAR Reading Lowest Observable Scale Score (LOSS),

- Texas STAAR Reading Highest Observable Scale Score (HOSS),
- Misfit to the Rasch model,
- Oral administration accommodation: Reading the test passages aloud,
- 0% and 100% correct on the Lexile Linking Test, or
- Percentile rank difference of scores between the two tests.

Table 8 shows, for each grade, the number of students (*N*) in the matched sample, the final sample, and the percent of each matched sample represented in the final sample. The table also summarizes the number of student test scores (by test level) removed from analysis and the reason for their removal. For the Grades 3 through 8 STAAR Reading study, of the 7,771 students in the matched sample, 5,856 (75.36%) remained in the final sample. For the STAAR English I and English II study, of the 2,137 students in the matched sample, 1,683 (78.76%) remained in the final sample.

Table 8.   Percentage of students in the linking study matched and final samples and reason for removal.

Matched Sample		Number of Students Removed by Reason					Final Sample	
Test Level	N	TX STAAR LOSS HOSS	Rasch Misfit	STAAR Accommodations	0 and 100% LT	Percentile Rank Difference	N	Percent of Matched
3	865	0	93	34	25	58	655	75.72
4	1,071	0	81	76	29	121	764	71.34
5	1,781	16	79	165	29	181	1,311	73.61
6	1,223	3	7	54	12	129	1,018	83.24
7	1,368	1	5	124	17	170	1,051	76.83
8	1,463	1	15	129	12	249	1,057	72.25
English I	827	0	7	0	0	200	620	74.97
English II	1,310	0	27	0	0	220	1,063	81.15

Table 9 presents the demographic characteristics of the STAAR Reading Grades 3 through 8 initial, matched, and final samples, and *Table 10* presents the demographic characteristics of the STAAR English I and English II initial, matched, and final samples. Through the trimming process, it is important to preserve the demographic characteristics of the original sample to ensure that bias is not introduced. After removals, some differences can be seen between the initial and final sample percentages for special education students and students receiving a read-aloud accommodation. In several categories, the proportion of students for whom data were "Not Available" decreased in the final sample. All other proportions are very similar to the initial sample.

Table 9.   Percentage of students in the STAAR Reading initial, matched, and final samples for selected demographic characteristics.

Student Characteristic	Category	Initial Sample N = 9,333	Matched Sample N = 7,771	Final Sample N = 5,856
Test Level	3	9.41	11.13	11.19
	4	20.46	13.77	13.03
	5	27.70	22.92	22.39
	6	12.76	15.74	17.39
	7	14.32	17.61	17.95
	8	15.35	18.83	18.05
Gender	Female	43.39	48.98	50.16
	Male	45.27	51.02	49.84
	Not Available	11.33	0.00	0.00
Ethnicity	American Indian	0.41	0.44	0.50
	Asian	1.63	1.53	1.67
	Black/African American	10.81	11.97	11.68
	Hispanic	55.90	53.06	54.64
	Native Hawaiian/ Pacific Islander	0.19	0.22	0.22
	Two or More	1.57	1.71	1.45
	White	24.70	29.33	28.25
	Non-Hispanic	1.43	1.74	1.59
	Not Available	3.36	0.00	0.00
Migrant	Yes	0.33	0.26	0.29
	No	87.77	99.05	99.33
	Not Available	11.90	0.69	0.38
Bilingual Program	Dual Language X1	1.00	1.22	1.02
	Dual Language X2	0.76	0.88	0.91
	Early Exit	0.18	0.15	0.15
	Late Exit	2.94	2.24	2.68
	Not in Bilingual Program	95.12	95.51	95.23
Special Education Students	Yes	6.68	7.21	4.20
	No	81.95	92.75	95.78
	Not Available	11.37	0.04	0.02
Read-Aloud Accommodation	Yes	7.20	7.89	0.00
	No	81.46	92.11	100.00
	Not Available	11.33	0.00	0.00

Table 10. Percentage of students in the STAAR English I and English II initial, matched, and final samples for selected demographic characteristics.

Student Characteristic	Category	Initial Sample N = 3,598	Matched Sample N = 2,137	Final Sample N = 1,683
Test Level	English I/Grade 9	44.08	38.7	36.84
	English II/Grade 10	55.92	61.3	63.16
Gender	Female	50.39	49.7	51.22
	Male	49.61	50.3	48.78
Ethnicity	American Indian	0.25	0.23	0.18
	Asian	3.03	1.73	1.72
	Black/African American	14.95	13.1	12.95
	Hispanic	70.98	75.2	76.29
	Native Hawaiian/ Pacific Islander	0.14	0.19	0.24
	Two or More	1.03	1.08	0.95
	White	9.62	8.47	7.66
English Language Learner	LEP	30.18	28.46	28.46
	Bilingual	7.56	8.38	8.38
	Other	62.23	64.67	63.1
	Not Available	0.03	0.05	0.06
Special Education Students	Yes	7.09	4.96	5.29
	No	92.88	94.99	94.65
	Not Available	0.03	0.05	0.06
Free and Reduced Lunch Program	Free Lunch	74.51	74.45	75.1
	Reduced Lunch	3.67	4.07	4.46
	Other	0.53	0.51	0.53
	Not FRL	21.26	20.92	19.85
	Not Available	0.03	0.05	0.06

Relationship between the STAAR Reading, English I, and English II and the Lexile Framework.
Table 11 presents the descriptive statistics for the STAAR Reading, English I, and English II scale scores and the Lexile Linking Test Lexile measures for the matched sample. STAAR Reading scale scores for Grades 5 through 8 were projected based on a quadratic function of STAAR state level scale score means regressed on grade and grade squared (TEA, 2017). This procedure enabled a student's STAAR scale score to be projected to the date on which the student was administered the Lexile Linking Test.

The correlations between the STAAR Reading scale scores and the Lexile measures range from 0.63 to 0.80 with a total correlation of 0.76 across Grades 3 through 8. The correlations for the STAAR English I and English II scale scores and the Lexile Linking Test Lexile measures were 0.65 and 0.75, respectively. Moderate correlations were observed between the STAAR scale scores and the Lexile measures, supporting the notion of the two assessments measuring a

similar construct. For the student sample, scale scores increase on the STAAR and the Lexile Linking Test as test level increases, with the exception of Grade 5, which has a higher mean score than Grade 6. Overall, given that grade-level samples were not specifically selected to be representative of a statewide student population, the generally increasing monotonic nature of student scores across all grades suggests that the vertical scales are maintained through the linking process for Grades 3 through 8. STAAR English I and English II scale scores were developed on separate horizontal scales.

Table 11. Descriptive statistics for the STAAR Reading estimated scale score and STAAR English I and English II scale scores and Lexile measures for the matched sample (N = 9,908).

Test Level	N	Matched Sample STAAR Reading SS Mean (SD)	Matched Sample Linking Test Lexile Measure Mean (SD)	r
3	865			0.80
4	1,071			0.76
5	1,781			0.76
6	1,223			0.75
7	1,368			0.72
8	1,463			0.63
3 - 8	7,771			0.76
English I	827			0.65
English II	1,310			0.75

* Estimated STAAR Reading scale scores based on quadratic function regressing state scale score means on grade and grade squared.

Table 12 presents the descriptive statistics and correlations for the STAAR Reading, English I, and English II scale scores and the Lexile Linking Test Lexile measures for the final sample. The within-grade correlations between the STAAR Reading scale scores and the Lexile measures range from 0.82 to 0.87 with an overall correlation of 0.86 across Grades 3 through 8. These correlations between the two scores indicate a strong relationship between STAAR scale scores and the Lexile Linking Test Lexile measures used for the linking study.

To account for the time difference between the STAAR English I and English II administration and the Lexile Linking Test administration, a mean adjustment was performed on both the STAAR scale scores and the Lexile measures for the final sample. The state STAAR mean scale score was used to center the STAAR scale for the final sample for both English I and English II (TEA, 2017). The Lexile Linking Test Lexile measures were adjusted to the expected Lexile measure based on a quadratic function estimated from Grades 3 through 8 STAAR Reading state level scale score means regressed on grade and grade squared using the Grades 3 through 8 Lexile linking formula (see *Table 13*).

The correlations between the STAAR English I and English II scales scores and the Lexile Linking Test Lexile measures are 0.87 for both STAAR assessments. These correlations indicate a strong relationship between STAAR English I and English II scale scores and the Lexile Linking Test Lexile measures used for the linking study.

Table 12. *Descriptive statistics for the STAAR Reading estimated scale score and STAAR English I and English II scale scores and Lexile measures for the final sample (N = 7,539).*

Test Level	*N*	Final Sample STAAR Reading SS Mean (SD)	Final Sample Linking Test Lexile Measure Mean (SD)	*r*
3	655			0.87
4	764			0.86
5	1,311			0.87
6	1,018			0.85
7	1,051			0.84
8	1,057			0.82
3 - 8	5,856			0.86
English I	620			0.87
English II	1,063			0.87

* Estimated STAAR Reading scale scores based on quadratic function regressing state scale score means on grade and grade squared.
** Centered on STAAR state level scale score mean.
*** Centered based on quadratic function regressing state Lexile measure means on grade and grade squared from Grade 3 through 8 linking function.

Figures 3 and *4* show the relationship between the STAAR Reading scale scores and the Lexile Linking Test Lexile measures for the matched and final samples. The final sample in *Figure 4* illustrate the linear relationship between the STAAR Reading scale scores and the Lexile Linking Test Lexile measures, reinforcing the use of linear linking.

Figures 5 and 6 illustrate the relationship between the STAAR EOC scale scores and the Lexile Linking Test Lexile measures for the matched sample where linear relationship between the two scales is observed. *Figures 7* and 8 illustrates the relationship between the STAAR EOC scale scores and the Lexile Linking Test Lexile measures for the final sample after the trimming process. The final sample scatterplots, *Figures 6* and 8, illustrate the strong linear relationship between the two scales, supporting the use of a linear link for the study.

Figure 3. Scatter plot of the STAAR Reading scale scores and the Lexile Linking Test Lexile measures, matched sample (N = 7,771).
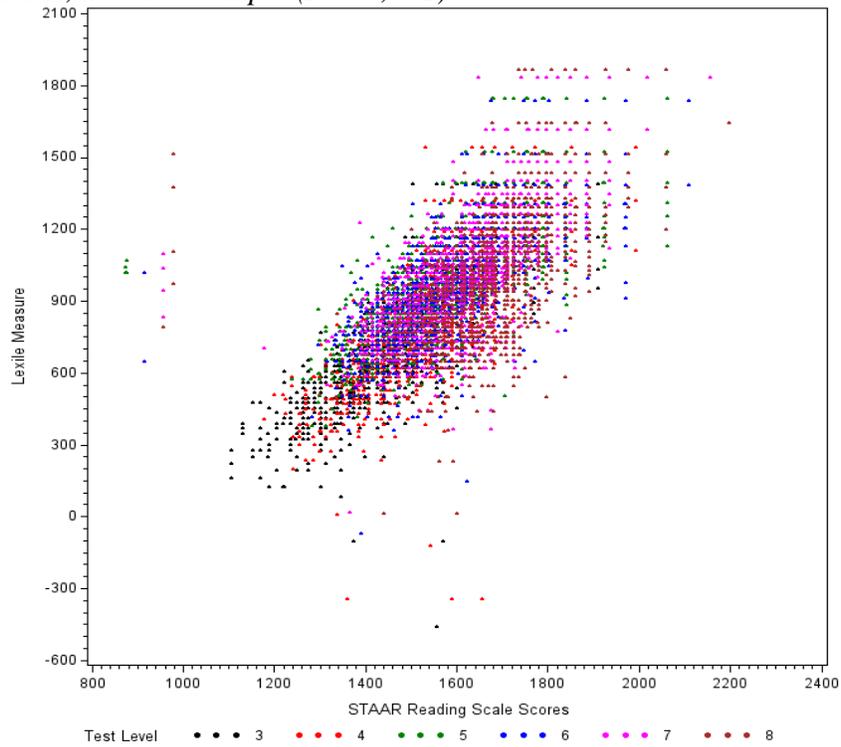
Figure 4. Scatter plot of the STAAR Reading scale scores and the Lexile Linking Test Lexile measures, final sample (N = 5,856).
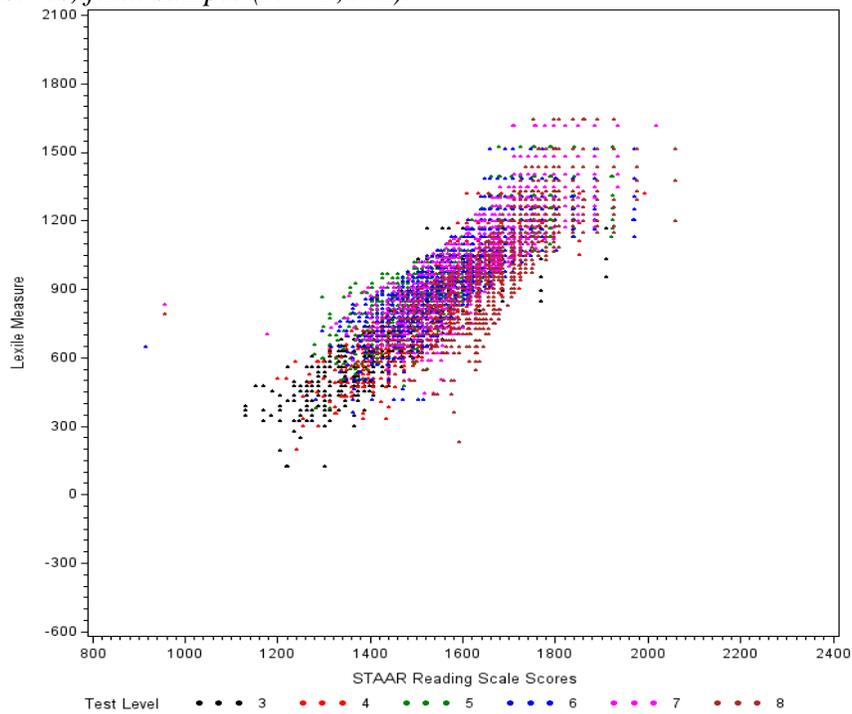
Figure 5. *Scatter plot of the STAAR EOC English I scale scores and the Lexile Linking Test Lexile measures, matched sample (N = 827).*
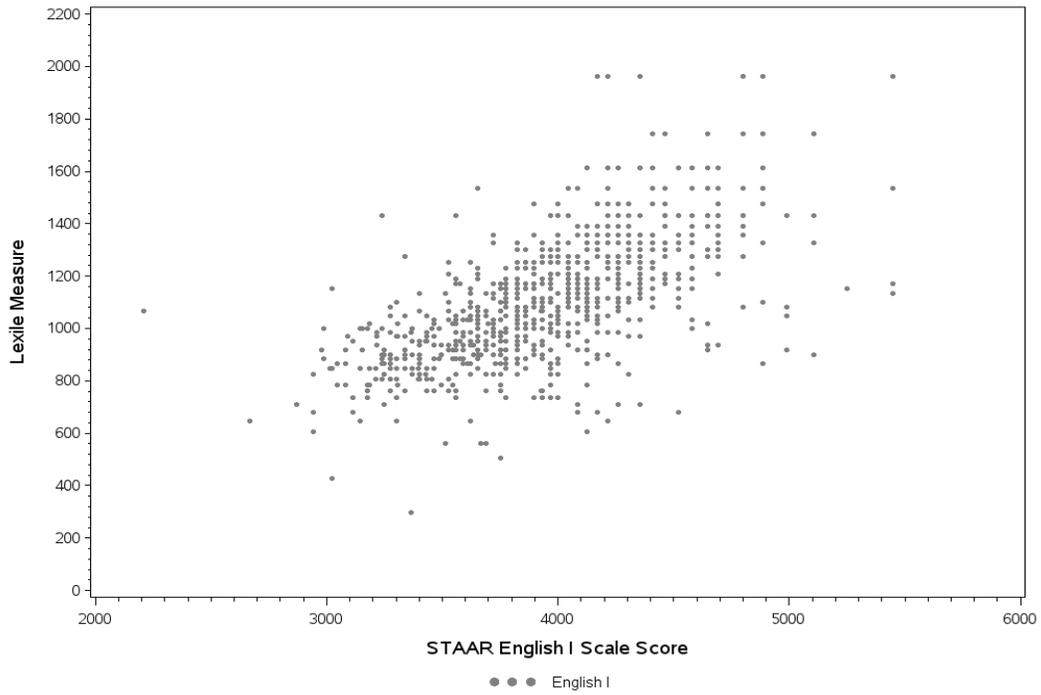
Figure 6. *Scatter plot of the STAAR EOC English II scale scores and the Lexile Linking Test Lexile measures, matched sample (N = 1,310).*
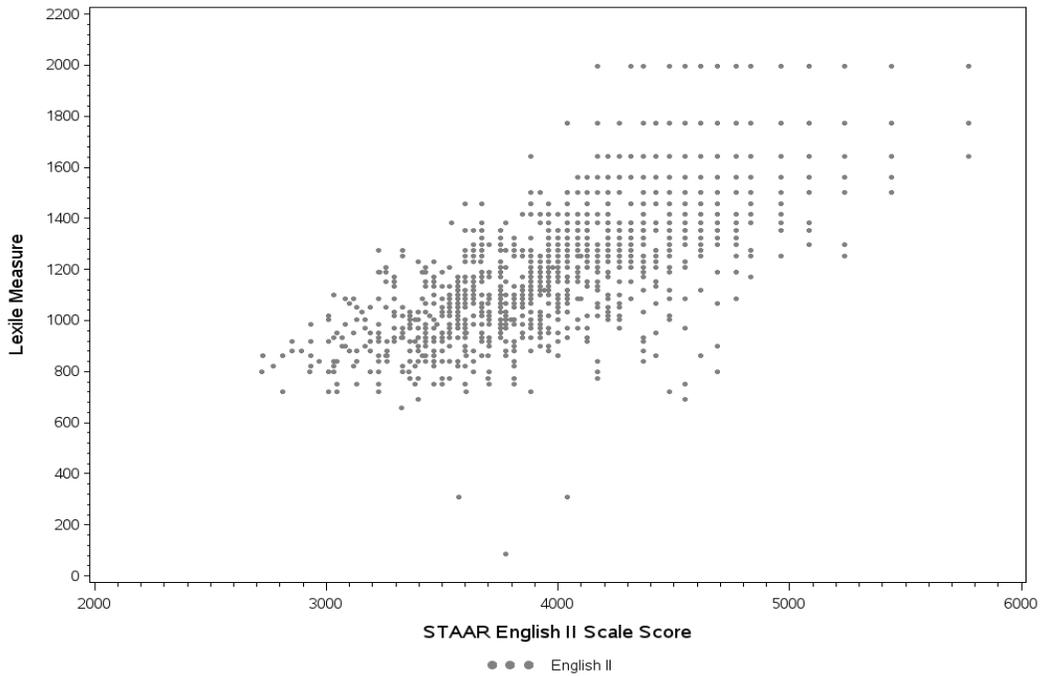
Figure 7. *Scatter plot of the STAAR EOC English I scale scores and the Lexile Linking Test Lexile measures, final sample (N = 620).*
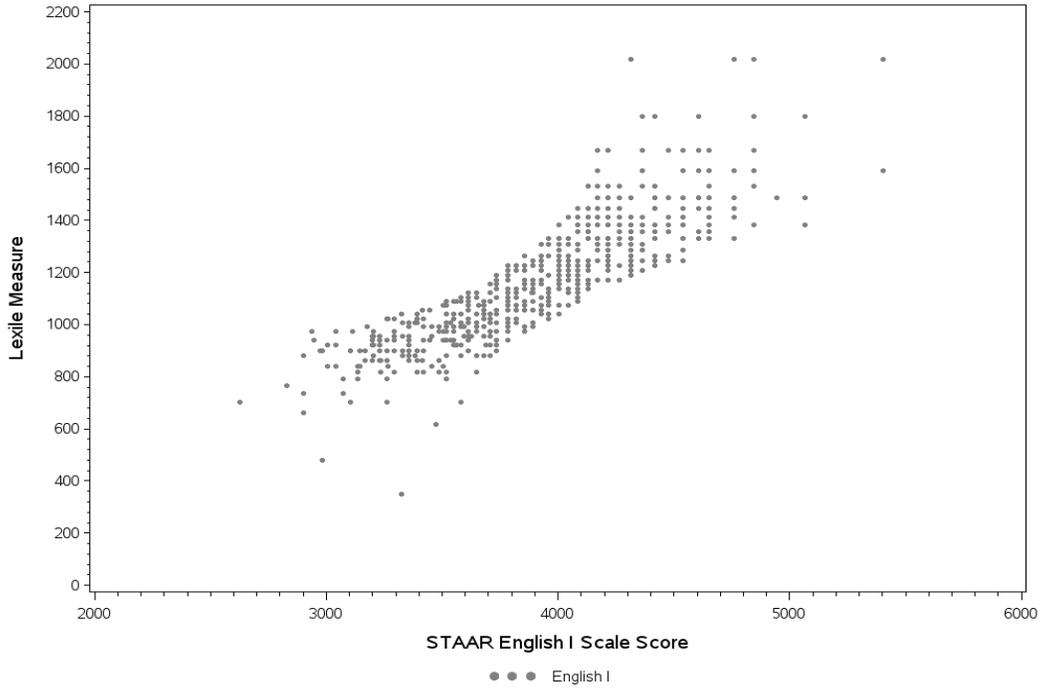
Figure 8. *Scatter plot of the STAAR EOC English II scale scores and the Lexile Linking Test Lexile measures, final sample (N = 1,063).*
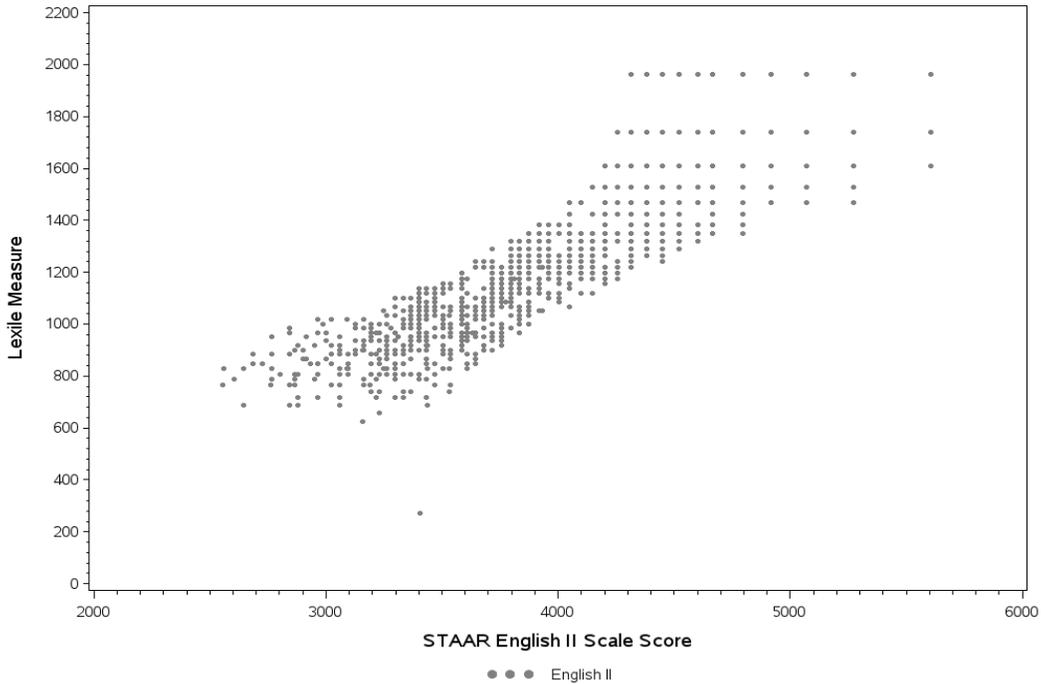
Linking the STAAR Reading, English I, and English II scales with the Lexile Scale

Linking in general means "putting the scores from two or more tests on the same scale" (National Research Council, 1999, p.15). MetaMetrics and the Texas Education Agency conducted this linking study for the purpose of matching students with books and texts—to predict the books and texts a student should be matched with for successful reading experiences, given their performance on the STAAR Reading, English I, and English II assessments.

Evaluation of linkage assumptions. Factors that affect the linkage between two assessments include the domain to be assessed, the definition of the framework for assessment, the test specifications, and the items sampled.

The correlation between the STAAR Reading, English I, and English II scale scores and the Lexile Linking Test Lexile measures support the conclusion that a similar construct is being measured, as illustrated in *Table 12*. The correlations between the two assessments are above or within the typical range of alternate-form reliability coefficients; therefore, the Lexile Linking Tests may be considered T-parallel forms of the STAAR Reading, English I, and English II assessments (see Note 1). By using alternate-form reliability coefficients as a comparison, similar sources of variation are accounted for (differences in testing occasions and items). In addition, the Lexile Linking Tests were constructed to have a similar number of score points and difficulty level as the STAAR Reading, English I, and English II assessments.

Linking Analyses. Two score scales (e.g., the STAAR Reading, English I, and English II assessment scale and the Lexile scale) can be linked using a linear method when (1) test forms have similar difficulties; and (2) simplicity in conversion tables or equations, in conducting analyses, and in describing procedures are desired (Kolen and Brennan, 2014).

In scale alignment which uses the same methods as linear equating (Dorans, Moses, and Eignor, 2010), a transformation is chosen such that two sets of scores are considered to be linked if they correspond to the same number of standard deviations above (or below) the mean in some group of examinees (Angoff, 1984, cited in Petersen, Kolen, and Hoover, 1989; Kolen and Brennan, 2014). Given scores *x* and *y* on tests *X* and *Y*, the linear relationship is

$$\frac{(x - \mu_x)}{\sigma_x} = \frac{(y - \mu_y)}{\sigma_y}$$
Equation (2)

and the linear transformation l_x (called the SD line in this report) used to transform scores on test *Y* to scores on text *X* is

$$x = l_x(y) = \left(\frac{\sigma_x}{\sigma_y}\right)y + \left(\mu_x - \frac{\mu_y \sigma_x}{\sigma_y}\right)$$
Equation (3)

Linear linking by definition has the same mean and standard deviation for the overall equation when the scale is vertically aligned. The means and standard deviations are the same for the linking test and the target test when calculated across grades. Linear linking using an SD-line approach is preferable to linear regression because the tests are not perfectly correlated. With less than perfectly reliable tests, linear regression is dependent on which way the regression is conducted: predicting scores on test *X* from scores on test *Y* or predicting scores on test *Y* from scores on test *X*. The SD line provides the symmetric linking function that is desired.

The final linking equation between STAAR scale scores and Lexile measures can be written as:

Lexile measure = Slope(STAAR scale score) + intercept   Equation (4)

where the slope is the ratio of the standard deviations of the STAAR assessment scale scores and the Lexile Linking Test Lexile measures.

Using the final sample data described in *Table 13*, the linear linking function relating the STAAR Reading, English I, and English II assessment scale scores and the Lexile Linking Test Lexile measures was established. Because the STAAR Reading reports scores on a vertical scale, one linking function was developed for Grades 3 through 8. STAAR English I and English II employ horizontal scales, so separate links were developed for each assessment. The slope and intercept for the linking equations are shown in *Table 13*.

Table 13. *Linear linking equation coefficients used to predict Lexile measures from the STAAR Reading, English I, and English II scale scores.*

Test Level	Slope	Intercept
3 through 8		
English I		
English II		

Conversion tables were developed to express the STAAR Reading, English I, and English II assessment scale scores in the Lexile metric and were delivered to ETS in electronic format.

Recommendations about reporting Lexile measures. 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). All Lexile measures should be rounded to the nearest 5L to avoid over interpretation of the measures. As with any test score, uncertainty in the form of measurement error is present.

Lexile measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is research (e.g., to measure growth at the student, grade, school, district, or state level), then actual measures should be used at all score points, rounded to the nearest integer. A computed Lexile measure of 772.5L would be reported as 773L. If the purpose

is instructional, then the Lexile measures should be capped at the upper bound of measurement error (e.g., at the 95th percentile of the national Lexile norms) to ensure developmental appropriateness of the material. MetaMetrics expresses these as "Reported Lexile Measures" and recommends that these measures be reported on individual score reports. The grade level caps used for reporting Lexile measures are shown in *Table 14*.

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 either BR (Beginning Reader) or "BRxxxL." No student should receive a negative Lexile measure on a score report. The lowest reported value below 0L is BR400L.

Table 14. Maximum reported Lexile measures, by grade.

Grade	Lexile Caps
3	1200L
4	1300L
5	1400L
6	1500L
7	1600L
8	1700L
English I	1725L
English II	1750L

Some assessments report a Lexile range for each student, which is 50L above and 100L below the student's actual Lexile measure. This range represents the boundaries between the easiest kind of reading material for the student and the level at which the student will be more challenged, yet can still read successfully.

Validity of the STAAR Reading, English I, and English II assessments–Lexile Link

Percentile Rank Distributions. Table 15 contains the percentile ranks of the Lexile Linking Test Lexile measures and the STAAR Reading, English I, and English II assessment Lexile measures based on the final sample. The criterion of a half-standard deviation (100L) on the Lexile scale was used to determine the size of the difference. In examining the values, the measures for the two tests are most similar in the middle of the score ranges and generally fall within the 100L criterion within the interquartile range. As the scores move toward the extreme ends of the distributions, in some grades, more variation is observed. In addition, Grade 5 results show

differences between the Lexile Linking Test and STAAR Reading Lexile measures throughout the distribution. Grade 5 student performance was higher than the typical performance patterns observed for the other grades. All grades were samples of convenience, and Grade 5 in particular was a higher ability group when compared to the state averages for Grade 5 students (See Note 2). When placed on the Lexile scale with the other grades in the study, the higher scores observed for the Grade 5 sample were aligned more closely to the overall trajectory of the vertical scale.

Table 15.   Comparison of Lexile measures for selected percentile ranks from the Lexile Linking Test and the STAAR Reading, English I, and English II.

Grade 3				Grade 4		
Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure		Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure
1	302L	186L		1	359L	367L
5	371L	330L		5	492L	513L
10	435L	401L		10	568L	577L
25	540L	542L		25	660L	705L
50	708L	730L		50	801L	846L
75	895L	917L		75	968L	988L
90	1167L	1129L		90	1113L	1125L
95	1167L	1279L		95	1192L	1191L
99	1167L	1279L		99	1323L	1432L

Grade 5				Grade 6		
Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure		Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure
1	554L	405L		1	506L	460L
5	680L	544L		5	626L	539L
10	774L	606L		10	696L	603L
25	888L	743L		25	818L	741L
50	1044L	922L		50	977L	922L
75	1207L	1094L		75	1165L	1085L
90	1394L	1249L		90	1308L	1187L
95	1526L	1343L		95	1387L	1286L
99	1526L	1492L		99	1518L	1441L

Table 15 (continued). Comparison of Lexile measures for selected percentile ranks from the Lexile Linking Test and the STAAR Reading, English I, and English II.

Grade 7				Grade 8		
Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure		Percentile Rank	Linking Test Lexile Measure	STAAR Reading Lexile Measure
1	590L	509L		1	501L	635L
5	682L	613L		5	679L	761L
10	731L	683L		10	728L	846L
25	873L	827L		25	850L	959L
50	1059L	994L		50	1030L	1085L
75	1265L	1153L		75	1229L	1227L
90	1407L	1286L		90	1379L	1361L
95	1486L	1381L		95	1517L	1405L
99	1616L	1447L		99	1647L	1524L

English I				English II		
Percentile Rank	Linking Test Lexile Measure	STAAR EOC Lexile Measure		Percentile Rank	Linking Test Lexile Measure	STAAR EOC Lexile Measure
1	701L	664L		1	716L	630L
5	850L	791L		5	808L	761L
10	901L	847L		10	883L	867L
25	991L	1014L		25	1001L	1006L
50	1155L	1178L		50	1218L	1216L
75	1331L	1329L		75	1423L	1452L
90	1485L	1497L		90	1608L	1597L
95	1589L	1555L		95	1740L	1727L
99	1799L	1657L		99	1962L	1906L

Lexile Framework Norms. Figures 9 and *10* show the Lexile measures from the STAAR Reading (*Figure 9)* and English I and English II (*Figure 10*) assessments compared to the norms that have been developed for use with The Lexile Framework for Reading. The normative information for The Lexile Framework for Reading is based on linking studies conducted with the Lexile Framework and the results of assessments that report directly in the Lexile metric ($N = 3,535,123$ students). The sample included students in Grades 1 through 12 from 51 states, districts, or territories and who were tested from 2010 to 2016. Of the students with gender information (45%), 51.6% of the students were male and 48.4% of the students were female. Of the students with race or ethnicity information (39%), the majority of the students in the norming sample were White (46.2%), with 4.8% African-American, 1.6% American Indian/Alaskan Native, 12.3% Hispanic, 12.5% Asian, and 4.4% Other. Of the students with data, 5.7 percent of the students were classified as "limited English proficient"; and 8.9 percent of the students were classified as "Needing Special Education Services." Approximately 45 percent of the students were eligible for the free or reduced-price lunch program. The 2017 Lexile norms have been validated in relation to a longitudinal sample of students across Grades 3 through 11 ($N = 101,610$).

At each percentile being examined (i.e. 25th, 50th, and 75th) in *Figure 9*, the STAAR Reading assessment scale scores reported in the Lexile metric fall well above the Lexile norm values, with the STAAR Reading assessment 50th percentile slightly higher than the Lexile 75th percentile for Grades 3 and 4 and below the Lexile 75th percentile in Grades 6 through 8. The STAAR Reading assessment 75th percentile is far above the Lexile 75th percentile value for all grades. The STAAR English I and English II Lexile measures in *Figure 10* are all slightly lower than the Lexile norms, except for the English II 75th percentile Lexile measure, which is slightly higher than the Lexile norms. The STAAR Lexile measures reflect the performance of a sample of students. Another sample could perform higher or lower on STAAR Reading, English I, or Engiish II; their corresponding Lexile measures would potentially be even higher or lower than the norm values.

Figure 9.   Selected percentiles (25th, 50th, and 75th) plotted for the STAAR Reading Lexile measure for the final sample (N = 5,856)
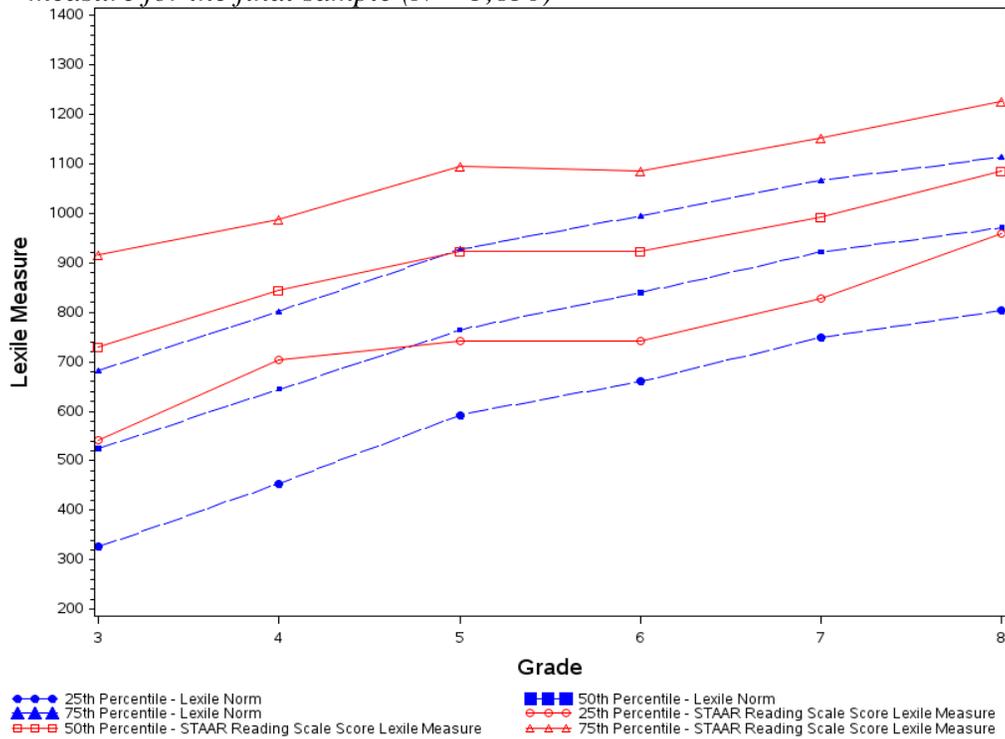
Figure 10. Selected percentiles (25th, 50th, and 75th) plotted for the STAAR EOC English I and English II Lexile measure for the final sample (N = 1,683)
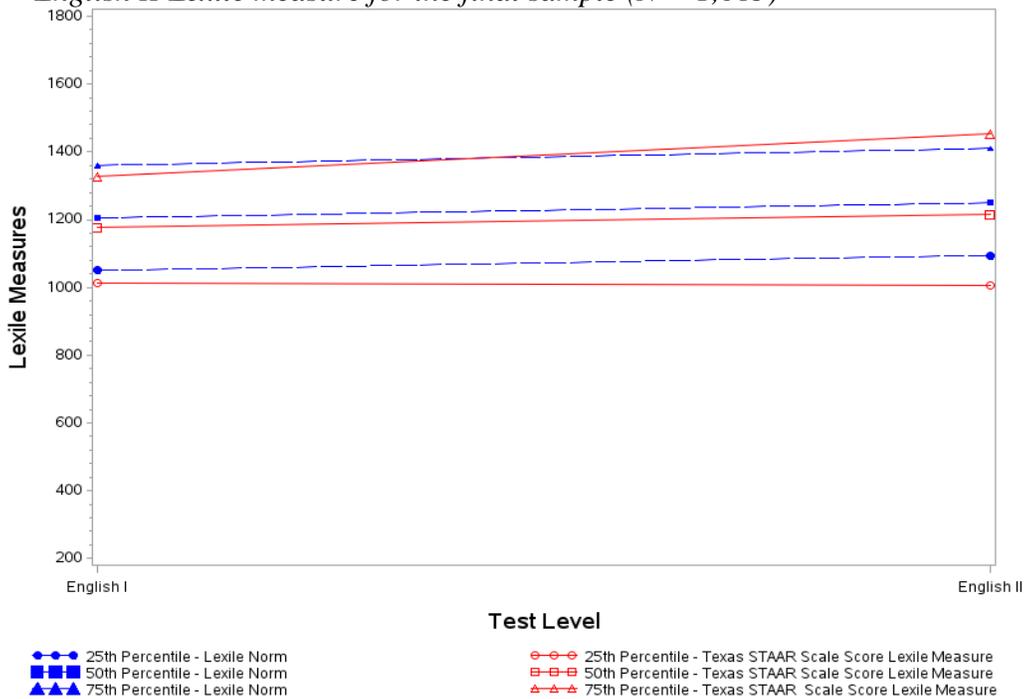
Grade Level Progressions. The following box-and-whisker plots (*Figures 11* through *13*) show the progression of STAAR Reading assessment scale scores and the Lexile measures from test level to test level. For each test level, the box refers to the interquartile range. The line within the box indicates the median. A solid line connects the means across test levels. The end of each whisker represents the 5th and 95th percentile of scores (the *y*-axis).

Both the STAAR Reading assessment scale scores and the Lexile measures are on vertical scales. As shown in *Figures 11* and *12*, the Grade 5 student sample had higher STAAR Reading scale scores and Lexile Linking Test Lexile measures than the Grade 6 student sample. This result is sample dependent and does not affect the vertical articulation of the conversion from STAAR Reading scale scores to Lexile measures.

Figure 11. Box-and-whisker plot of the STAAR Reading assessment scale scores by grade, final sample (N = 5,856).
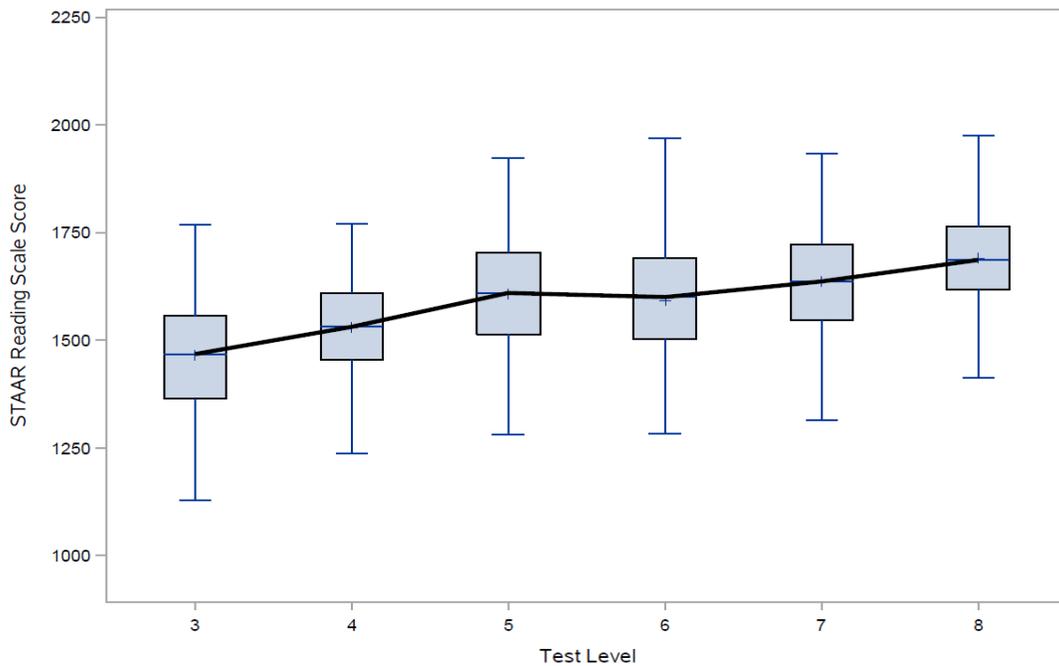
Figure 12. Box-and-whisker plot of the STAAR EOC English I and English II assessment scale scores, final sample (N = 1,683).
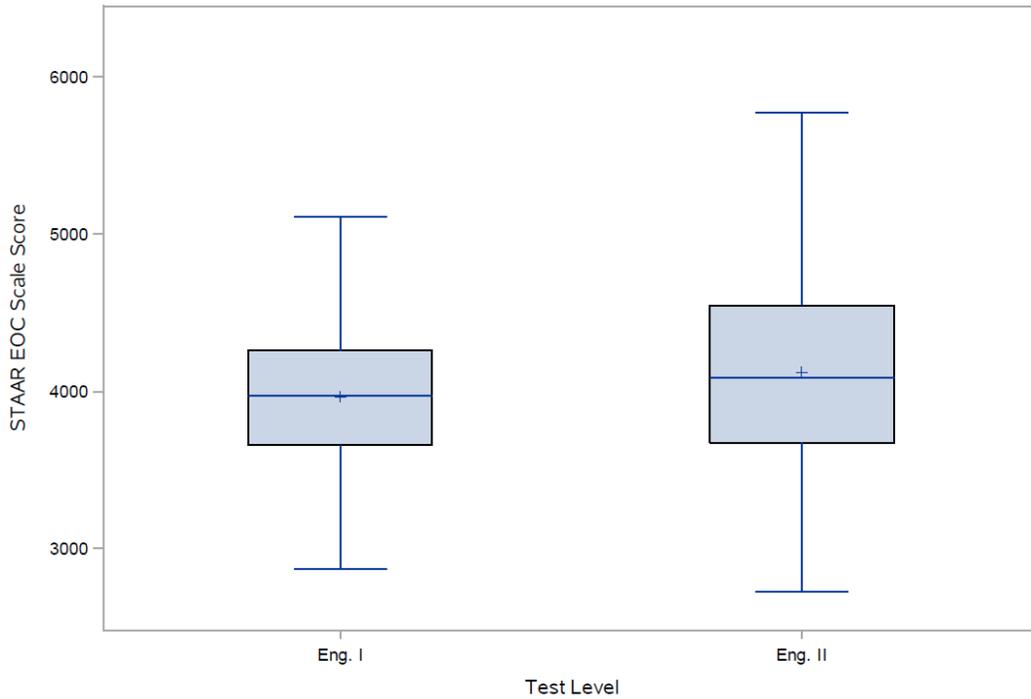
Figure 13. Box-and-whisker plot of the STAAR Reading, English I, and English II Lexile measures by grade, final sample (N = 7,538).
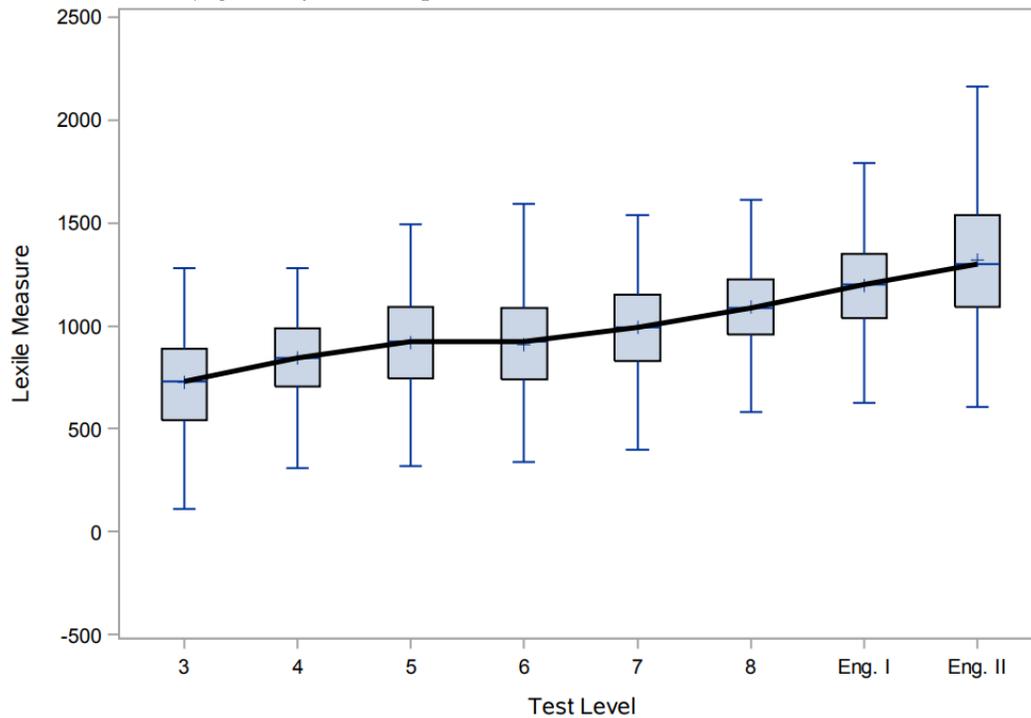
STAAR performance standards. STAAR performance standards relate levels of test performance to the expectations defined in the TEKS. Performance standards (or cut scores) established by the Texas Education Agency distinguish between four performance levels: Did Not Meet, Approaches, Meets, and Masters.

A phase-in period was implemented for STAAR Reading, English I, and English II performance standards to provide school districts with time to adjust instruction. A phase-in period for Level II is in place for all general STAAR assessments. Phase-in 1 performance standards were in effect from 2012 to 2015. Phase-in 2 standards will be in effect from 2016 and beyond (B. Xiang, personal communication, March 1, 2017).

Table 16 provides the STAAR Reading, English I, and English II assessment scale scores associated with the current performance levels associated cut points and corresponding Lexile measures.

Table 16.   STAAR Reading, English I, and English II Lexile measures for each performance level cut point.

Grade	Did Not Meet	Lexile Measure	Approaches	Lexile Measure	Meets	Lexile Measure	Masters	Lexile Measure
3	<1345	<500L	1345-1467	505L to 725L	1468-1554	730L to 885L	≥1555	≥890L
4	<1434	<665L	1434-1549	670L to 875L	1550-1632	880L to 1025L	≥1633	≥1030L
5	<1470	<730L	1470-1581	735L to 935L	1582-1666	940L to 1090L	≥1667	≥1095L
6	<1517	<815L	1517-1628	820L to 1020L	1629-1717	1025L to 1180L	≥1718	≥1185L
7	<1567	<905L	1567-1673	910L to 1100L	1674-1752	1105L to 1245L	≥1753	≥1250L
8	<1587	<945L	1587-1699	950L to 1150L	1700-1782	1155L to 1300L	≥1783	≥1305L
Eng. I	<3775	<1095L	3775-3999*	1100L – 1210L	4000–4690	1215L – 1570L	≥4691	≥1575L
Eng. II	<3775	<1140L	3775-3999**	1145L – 1255L	4000-4830	1260L – 1675L	≥4831	≥1680L

* For years 2012 – 2015, the cut score for the English I Approaches was 3750, which corresponds to 1085L.
** For years 2012 – 2015, the cut score for the English II Approaches was 3750, which corresponds to 1130L.

The Lexile Framework and Forecasted Comprehension Rates

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 is carved into item-sized slices of approximately 125-140 words with a question embedded in each slice. A reader who answers three-fourths of the questions correctly has a 75-percent comprehension rate.

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

If all the item-sized slices in the 350L text had the same calibration, the 250L difference between the 600L reader and the 350L text could be determined using the Rasch model equation. 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 5 is an unbiased approximation:

$$Rate = \frac{e^{ELD+1.1}}{1+e^{ELD+1.1}}$$
Equation (5)

where ELD is the "effective logit difference" given by

$$ELD = (\text{Reader Lexile measure} - \text{Text Lexile measure}) \div 225.$$
Equation (6)

Figure 14 shows the general relationship between reader-text discrepancy and forecasted comprehension rate. When the reader measure and the text calibration are the same (difference of 0L) then the forecasted comprehension rate is 75 percent. In the example in the preceding paragraph, the difference between the reader measure of 600L and the text calibration of 350L is 250L. Referring to *Figure 14* and using +250L (reader minus text), the forecasted comprehension rate for this reader-text combination would be 90 percent.

Figure 14. Relationship between reader-text discrepancy and forecasted comprehension rate.
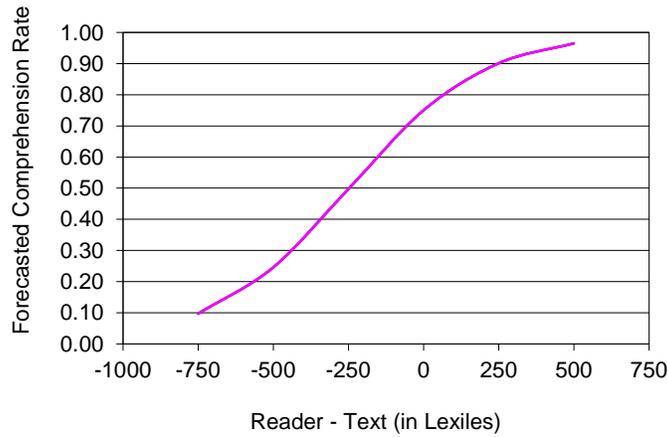
Tables 17 and *18* show comprehension rates calculated for various combinations of reader measures and text calibrations.

Table 17. Comprehension rates for the same individual with materials of varying comprehension difficulty.

Person Measure	Text Calibration	Sample Titles	Forecasted 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 18.   Comprehension rates of different person abilities with the same material.

Person Measure	Calibration for a Grade 10 Biology Textbook	Forecasted Comprehension Rate
500L	1000L	25%
750L	1000L	50%
1000L	1000L	75%
1250L	1000L	90%
1500L	1000L	96%

The subjective experience of 50-percent, 75-percent, and 90-percent comprehension as reported by readers varies greatly. A 1000L reader reading 1000L text (75-percent 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 appear to comprehend what they are reading. A 1000L reader reading 1250L text (50-percent 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. Finally, a 1000L reader reading 750L text (90-percent comprehension) reports total control of the text, reads with speed, and experiences automaticity during the reading process.

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

Relationship between Linking Error and Forecasted Comprehension Rate. Using Equation 5 with different combinations of reader measure and text difficulty, the effect of linking error on forecasted comprehension rate can be examined. *Table 19* shows the changes in the forecasted comprehension rate for different combinations of reader and text interactions. When the linking error is small, 5–10L, then the effect on forecasted comprehension rate is a minimal difference (1 to 2 percent) increase or decrease in comprehension.

Table 19.   Effect of reader-text discrepancy on forecasted comprehension rate.

Reader Lexile Measure	Text Lexile Measure	Difference	Forecasted Comprehension Rate
1000L	970L	30L	77.4%
1000L	975L	25L	77.0%
1000L	980L	20L	76.7%
1000L	985L	15L	76.3%
1000L	990L	10L	75.8%
1000L	995L	5L	75.4%
1000L	1000L	0L	75.0%
1000L	1005L	⁻5L	74.6%
1000L	1010L	⁻10L	74.2%
1000L	1015L	⁻15L	73.8%
1000L	1020L	⁻20L	73.3%
1000L	1025L	⁻25L	72.9%
1000L	1030L	⁻30L	72.4%

Conclusions, Caveats, and Recommendations

Forging a link between scales is a way to add value to one scale without having to administer an additional test. Value can be in the form of any or all of the following:

- increased *interpretability* (e.g., "Based on this test score, what can my child actually read?"), or
- increased *instructional use* (e.g., "Based on these test scores, I need to modify my instruction to include these skills.").

The link that has been established between the STAAR Reading, English I, and English II assessment scales and the Lexile scale permits readers to be matched with books and texts that provide an appropriate level of challenge while avoiding frustration. The result of this purposeful match may be that students will read more, and, thereby read better. 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-percent comprehension. As a reader grows, he or she can be matched with more demanding texts. In addition, as the texts become more demanding, then the reader grows.

The concordance provides a link between the two scales. All linking functions are statistical estimates based on data collected from specific samples. These linking results may be somewhat different with a different sample. Other factors may affect the students' performance during the test administrations such as health conditions, surrounding environment, or motivation.

Recommendations about reporting Lexile measures for readers. 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). All Lexile measures should be rounded to the nearest 5L to avoid over interpretation of the measures. As with any test score, uncertainty in the form of measurement error is present.

Lexile measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is research (e.g., to measure growth at the student, grade, school, district, or state level), then actual measures should be used at all score points, rounded to the nearest integer. A computed Lexile measure of 772.51 would be reflected as 773L. If the purpose is instructional, then the Lexile measures should be capped at the upper bound of measurement error (e.g., at the 95th percentile of the national Lexile norms) to ensure developmental appropriateness of the material. MetaMetrics expresses these as "Reported Lexile Measures" and recommends that these measures be reflected on individual score reports. 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 either BR or "BRxxxL." No student should receive a negative Lexile measure on a score report. The lowest reported value below 0L is BR400L.

Some assessments report a Lexile range for each student, which is 50L above and 100L below the student's actual Lexile measure. This range represents the boundaries between the easiest kind of reading material for the student and the level at which the student will be more challenged, yet can still read successfully. A reader with a Lexile measure of 1000L would have a Lexile range of 900L–1050L. This range represents the recommended guideline to select reading materials for instructional purposes. Understanding the impact of selecting reading materials on a student's comprehension is important for student success. It should be noted that material above or below the reader's Lexile range may be used for specific instructional purposes. As in any academic setting, the teachers and parents know the student best. The Lexile range is best viewed as a tractable guideline where teachers or parents selecting reading materials outside of the Lexile range may seem more appropriate.

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* (Summer 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. *Figure 15* shows the Lexile ranges of reading materials in the domains investigated by Williamson. 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 15. A continuum of text difficulty for the transition from high school to postsecondary experiences (box plot percentiles: 5th, 25th, 50th, 75th, and 95th).[1]
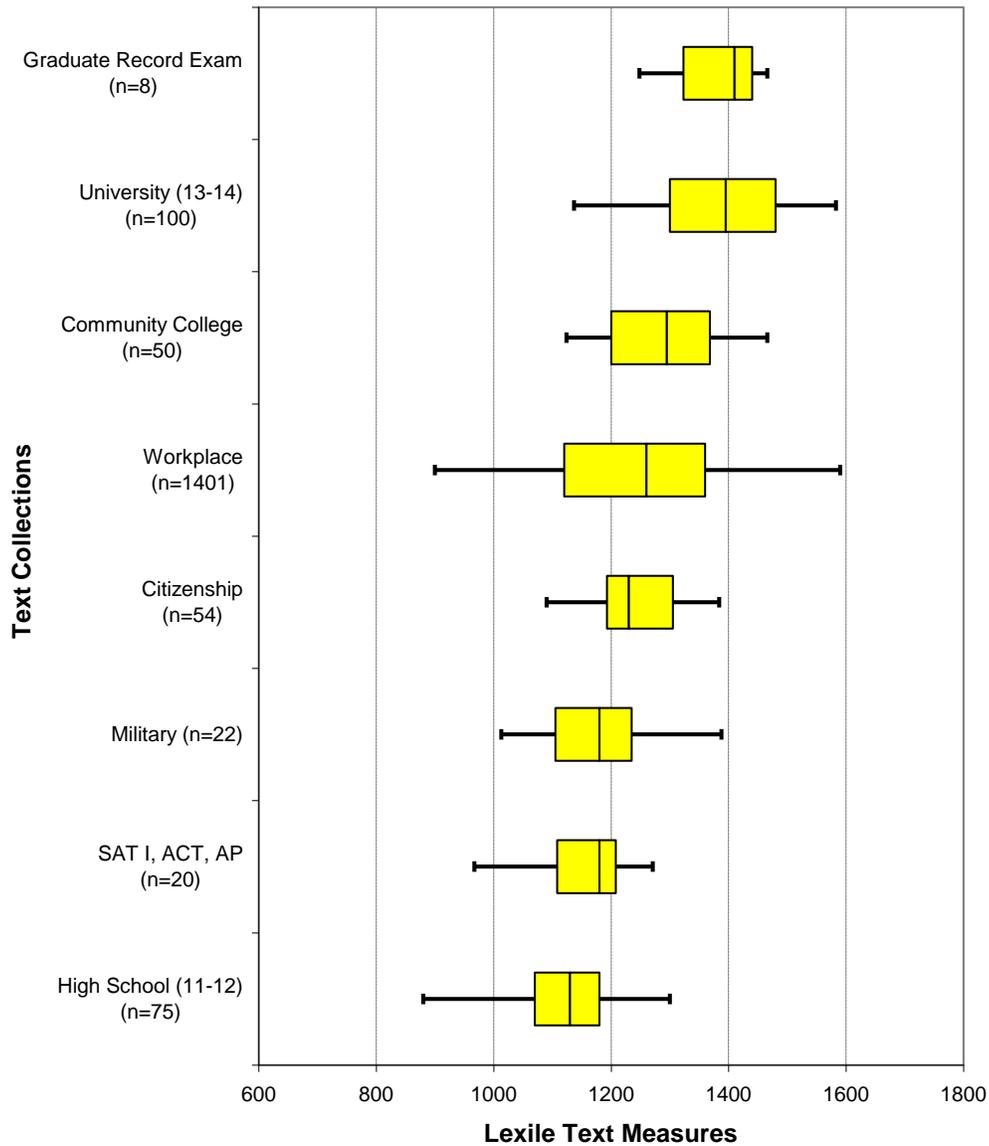
In Texas, two studies examined the reading demands in various postsecondary options – technical college, community college, and 4-year university programs. Under Commissioner Raymond Paredes, THECB conducted a research study in 2007 (and extended in 2008) which addressed the focal question of "how well 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. The text complexity of these books was measured using The Lexile Framework for Reading. Since

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

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. (For a complete description of this report, visit www.thecb.state.tx.us/index.cfm?objectid=31BFFF6B-BB41-8A43-C76A99EDA0F38B7D.)

Since 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.

Expanding on Williamson's work, Stenner, Sanford-Moore, and Williamson (2012) aggregated the 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."

The question for educators becomes how to determine if a student is "on track" for college and career. Implementing the higher standards 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.

Between 2004 and 2008, MetaMetrics (Williamson, Koons, Sandvik, and Sanford-Moore, 2012) conducted research to describe the typical reading demands and develop a text continuum of reading materials across Grades 1-12. The grade-by-grade text distributions are presented in *Figure 16*.

Figure 16. Text complexity distributions, in Lexile units, by grade (whiskers represent 5th and 95th percentiles).
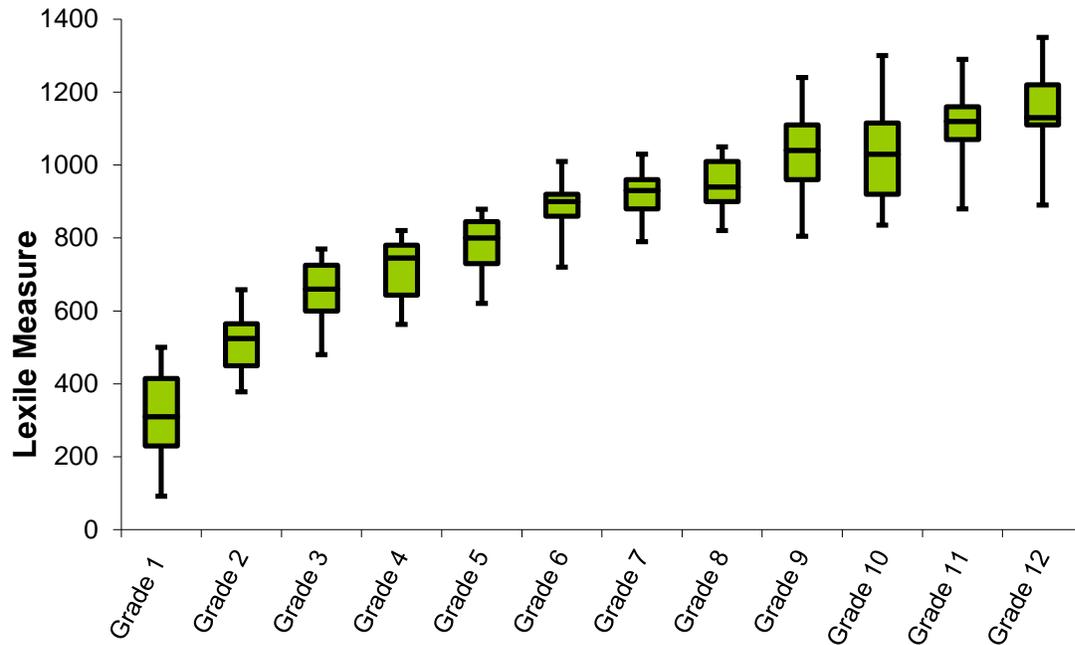
This continuum can be "stretched" to describe the reading demands expected of students in Grades 1-12 who are "on track" for college and career (Sanford-Moore and Williamson, 2012). This information can provide a basis for defining at what level students need to be able to read to be ready for various postsecondary endeavors such as further education beyond high school and entering the work force.

Table 20.   Lexile ranges aligned to college- and career-readiness reading expectations, by grade.

Grade	2012 "Stretch" Text Measure
1	190L to 530L
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

MetaMetrics' research on the typical reading demands of college and careers contributed to the Lexile-based grade bands in *Figure 17*. *Figure 17* shows the relationship between the Meets score range for each test level established on the STAAR Reading, English I, and English II assessments and the "stretch" reading demands based on college and career readiness text complexity research conducted by MetaMetrics. At each grade, the lowest score in the Meets range is the cut point and the highest score in the Meets range is last score before the Masters cut point.

Figure 17. Comparison of the STAAR Reading, English I, and English II assessment scale score ranges and the college and career reading.
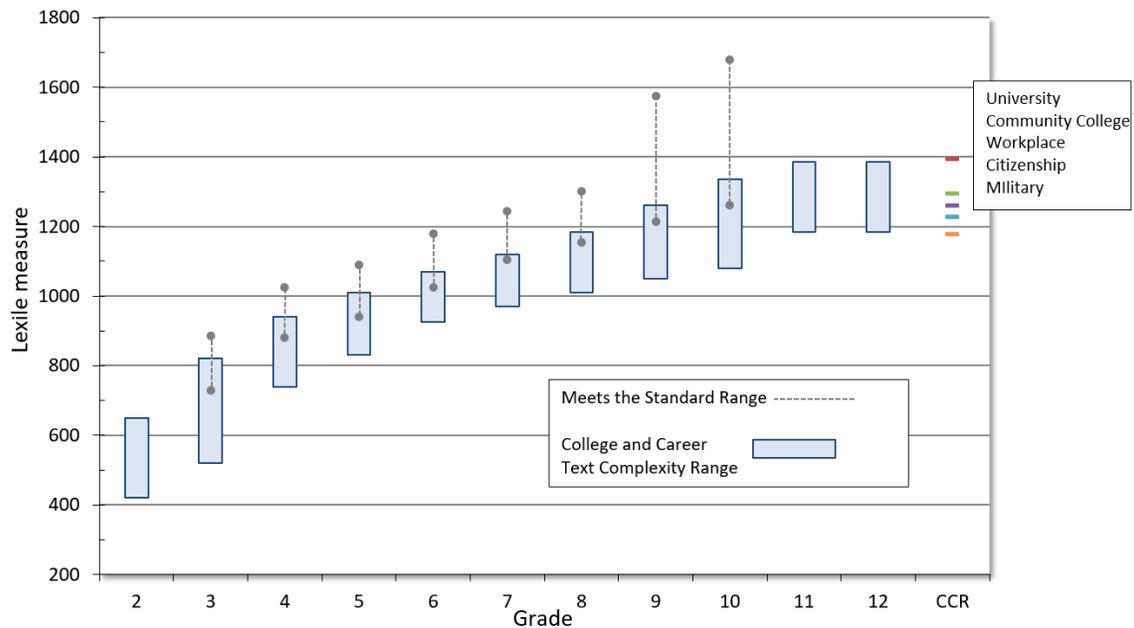
Figure 18 shows the linking study sample student performance on the STAAR Reading, English I, and English II assessments expressed as Lexile measures at each test level. For each test level, the box refers to the interquartile range. The line within the box indicates the median. The end of each whisker represents the 5th percentile at the low end and the 95th percentile at the high end values of the scores on the *y*-axis. The boxes indicated by dashed lines represent the recommended college and career text complexity ranges. Students can then be matched with reading materials that are at or above the recommendations for college and career readiness for each grade level.

Figure 18. STAAR Reading, English I, and English II assessment student performance expressed as Lexile measures compared to college and career readiness ranges (N = 7,539).
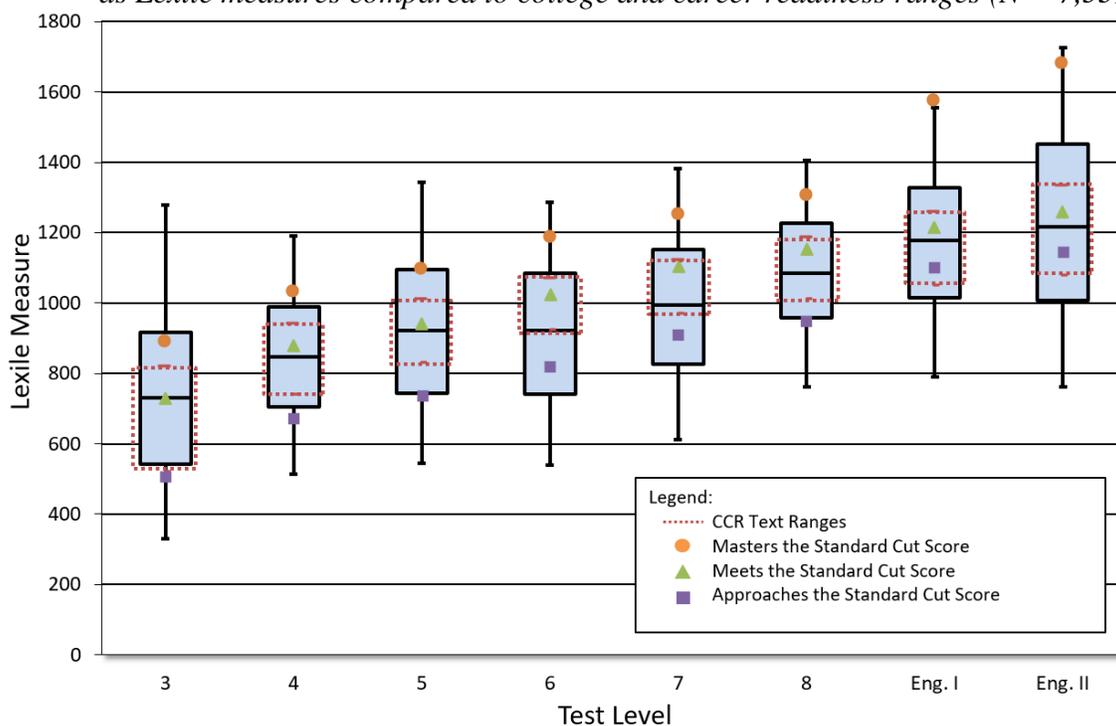
Next Steps. To utilize the results from this study, Lexile measures need to be incorporated into the STAAR Reading assessment results processing and interpretation frameworks. When that occurs, the measures and the educators can use the tools available through The Lexile Framework for Reading to link the assessment results with subsequent instruction.

Within the *instructional area*, suggested book lists can be developed for ranges of readers. Care must be taken to ensure that the books on the lists are also developmentally appropriate for the readers. The Lexile measure is one factor related to comprehension and is a good starting point in the selection process of a book for a specific reader. Other factors such as student developmental level, motivation, and interest; amount of background knowledge possessed by the reader; and characteristics of the text such as illustrations and formatting also need to be considered when matching a book with a reader.

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).

In this era of student-level accountability and high-stakes assessment, differentiated instruction—the attempt "on the part of classroom teachers to meet students where they are in the learning process and move them along as quickly and as far as possible in the context of a mixed-ability classroom" (Tomlinson, 1999)—is a means for all educators to help students

succeed. Differentiated instruction promotes high-level and powerful curriculum for all students, but varies the level of teacher support, task complexity, pacing, and avenues to learning based on student readiness, interest, and learning profile. One strategy for managing a differentiated classroom suggested by Tomlinson is the use of multiple texts and supplementary materials.

The Lexile Framework is an objective tool that can be used to determine a student's readiness for a reading experience; the Lexile Framework "targets" text (books, newspapers, periodicals) for readers at a 75 percent comprehension level—a level that is challenging, but not frustrating (Schnick and Knickelbine, 2000).

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

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

Communicate With Parents Meaningfully to Include Them in the Educational Process. Teachers can make statements to parents such as, "Your child 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.

Lexile Measures 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.

Communicating with Lexile measures. Lexile measures can be used to communicate with students, parents, teachers, educators, and the community by providing a common language to use to talk about reading growth and development. By aligning all areas of the educational system, parents can be included in the instructional process. With a variety of data related to a student's reading level a more complete picture can be formed and more informed decisions can be made concerning reading-group placement, amount of extra instruction needed, and promotion/retention decisions.

It is much easier to understand what a national percentile rank of 50 means when it is tied to the reading demands of book titles that are familiar to adults. Parents are encouraged to help their children achieve high standards by expecting their children to succeed at school, communicating with their children's teachers and the school, and helping their children keep pace and do homework.

Through the customized reading lists and electronic database of titles, parents can assist their children in the selection of reading materials that are at the appropriate level of challenge and monitor the reading process at home. A link can be provided to the "Find a Book with Lexiles" website. This site provides a quick, free resource to battle "summer slide" – the learning losses that students often experience during the summer months when they are not in school. Lexiles make it easy to help students read and learn all summer long and during the school year. This website can help build a reading list of books at a young person's reading level that are about subjects that interest him or her. This website can be viewed at http://www.lexile.com/findabook/.

In one large school district, the end-of-year testing results are sent home to parents in a folder. The folder consists of a Lexile Map on one side and a letter from the superintendent on the other side. The school district considers this type of material as "refrigerator-friendly." They encourage parents to put the Lexile Map on the refrigerator and use it to monitor and track the reading progress of their child throughout the school year.

The community-at-large (business leaders, citizens, politicians, and visitors) sees the educational system as a reflection of the community. Through the reporting of assessment results (after all, that is what the community is most interested in—results), people can understand what the community values and see the return for its investment in the schools and its children.

One way to involve the community is to work with the public libraries and local bookstores when developing reading lists. The organizations should be contacted early enough so that they can be sure that the books will be available. Often books can be displayed with their Lexile measures for easy access.

Many school districts make presentations to civic groups to educate the community as to their reading initiatives and how the Lexile Framework is being utilized in the school. Conversely, many civic groups are looking for an activity to sponsor, and it could be as simple as "donate-a-book" or "sponsor-a-reader" campaigns.

Notes

1. A T-parallel test is a test that is designed to be "theoretically parallel" to another test in that it has the same number of items/points, the same overall level of difficulty in terms of raw score means and standard deviations, and assesses the same construct domain (MetaMetrics, Inc. 1998).

2. Grade 5 Texas STAAR Reading average scale score for Spring 2016 was 1562 compared to the observed STAAR Reading average scale score for the sample of 1607 (TEA, 2016e).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bormuth, J.R. (1966). Readability: New approach. *Reading Research Quarterly*, *7*, 79-132.

Carroll, J.B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.

Carver, R.P. (1974). Measuring the primary effect of reading: Reading storage technique, understanding judgments and cloze. *Journal of Reading Behavior*, *6*, 249-274.

Chall, J.S. (1988). "The beginning years." In B.L. Zakaluk and S.J. Samuels (Eds.), *Readability: Its past, present, and future*. Newark, DE: International Reading Association.

Crain, S. & Shankweiler, D. (1988). "Syntactic complexity and reading acquisition." In A. Davidson and G.M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Erlbaum Associates.

Davidson, A. & Kantor, R.N. (1982). On the failure of readability formulas to define readable text: A case study from adaptations. *Reading Research Quarterly*, *17*, 187- 209.

Dunn, L.M. & Dunn, L.M. (1981). *Manual for Forms L and M of the Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L.M. & Markwardt, F.C. (1970). *Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.

Efron, B. (1981). Nonparametric estimates of the standard error: The Jackknife, the Bootstrap, and other resampling techniques. *Biometrika*. *68*, 589-599.

Grolier, Inc. (1986). *The electronic encyclopedia*. Danbury, CT: Author.

Haladyna, T. M., (1994). *Developing and Validating Multiple-Choice Test Items*. Hillsdale, NJ. Lawrence Erlbaum Associates.

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

Kolen, M.J. & Brennan, R.L. (2014). *Test equating, scaling, and linking: Methods and practices*. [rd] edition. New York: Springer Science + Business Media, LLC.

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.

Linacre, J.M. (2011). WINSTEPS (Version 3.70) [Computer Program]. Chicago: Author.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum Associates.

McGraw-Hill Book Company. (1983). *Guidelines for bias-free publishing*. Monterey, CA: Author.

MetaMetrics, Inc. (2008). The text continuum in 2008. Presented at the Lexile National Conference, San Antonio, TX.

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

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects, Appendix A. Washington, DC: Author.

National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, D.C.: National Academy Press.

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.

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

Schnick, T. & Knickelbine, M. (2000). *The Lexile Framework: An introduction for educators*. Durham, NC: MetaMetrics, Inc.

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

Sitter, R.R. (1992). Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics*, *20*(2), 135-154.

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, February). *Not so common: Comparing Lexile® Measures with the standards' other text complexity tools*. MetaMetrics White Paper. Durham, NC: MetaMetrics, Inc.

Stenner, A.J. (1990). Objectivity: Specific and general. *Rasch Measurement Transactions*, *4*, 111.

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., Horabin, 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., Horabin, I., & Smith, M. (1987b). Fit of the Lexile Theory to sequenced units from eleven basal series. Durham, NC: MetaMetrics, Inc.

Texas Education Agency. (2010a) *Grade 3 Reading Test Design Schematic (English and Spanish).* Retrieved September, 2, 2015 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2010b) *Grade 4 Reading Test Design Schematic (English and Spanish).* Retrieved September, 2, 2015 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2010c) *Grade 5 Reading Test Design Schematic (English and Spanish).* Retrieved September, 2, 2015 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2015). Chapter 4: State of Texas Assessment of Academic Readiness (STAAR). In *Technical digest: For the academic year 2014-2015* (pp. 77- 134). Retrived December 8, 2015 from http://tea.texas.gov/Student_Testing_and_Accountability/Testing/Student_Assessment_Overview/Technical_Digest_2014-2015/

Texas Education Agency. (2016a) *Grade 6 Reading Test Design Schematic.* Retrieved December, 2, 2016 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2016b) *Grade 7 Reading Test Design Schematic.* Retrieved December, 2, 2016 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2016c) *Grade 8 Reading Test Design Schematic.* Retrieved December, 2, 2016 from http://tea.texas.gov/student.assessment/staar/reading/

Texas Education Agency. (2016d) *STAAR Writing and English I, II, III Resources.* Retrieved December, 12, 2017 from https://tea.texas.gov/student.assessment/staar/writing/

Texas Education Agency. (2016e). *Statewide summary reports 2015-2016.* Retrieved December 2, 2016 from http://tea.texas.gov/Student_Testing_and_Accountability/Testing/State_of_Texas_Assessments_of_Academic_Readiness_(STAAR)/STAAR_Statewide_Summary_Reports_2015-2016/

Texas Education Agency. (2017). *Statewide summary reports 2016-2017.* Retrieved December 2, 2016 from http://tea.texas.gov/Student_Testing_and_Accountability/Testing/State_of_Texas_Assessments_of_Academic_Readiness_(STAAR)/STAAR_Statewide_Summary_Reports_2016-2017

Tomlinson, C.A. (1999). *The differentiated classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

Williamson, G. L. (2011, March 16-17). *Growth and growth norms: Standards for academic growth.* Presentation at the first semi-annual meeting of MetaMetrics' Technical Advisory Committee, held at MetaMetrics, Inc., Durham, NC.

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.

Williamson, G. L., Thompson, C. L., and Baker, R. F. (2007, April). *North Carolina's growth in reading and mathematics.* AERA Distinguished Paper Presentation at the 2007 annual meeting of the American Educational Research Association (AERA), Chicago, IL.

Wright, B.D. & Linacre, J.M. (1994, August). *The Rasch model as a foundation for the Lexile Framework*. Unpublished manuscript.

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

Appendix

The Lexile Framework for Reading Map

THE LEXILE® FRAMEWORK
FOR READING MAP

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 fab.lexile.com.

Pete: 490L

480L
Rally for Recycling
INFORMATIONAL

540L
Ron's Big Mission
LITERATURE

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 reader text 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 literature 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 website 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 fab.lexile.com to access the "Find a Book" tool. "Find a Book" enables users to search from over 275,000 books to build custom reading lists based on Lexile range and personal interests and to check the availability of books at the local library.

Kaitlyn: 840L

810L
Where the Mountain Meets the Moon
LITERATURE

IG860L
Animals Nobody Loves
INFORMATIONAL

Marisa: 1300L

1350L
The Secret Sharer
LITERATURE

1200L
The Dark Game: True Spy Stories
INFORMATIONAL
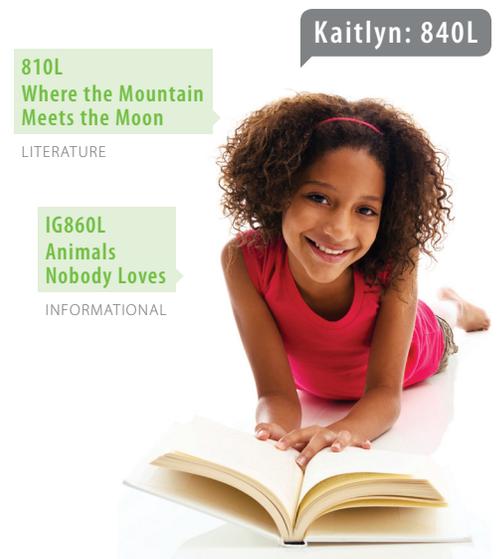
1500L+

1630L *Descartes: Philosophical Essays* LAFLEUR

But neither should we fall into the error of those who occupy their minds only with deep and serious matters, of which, after much effort, they acquire only a confused knowledge, while they hoped for a profound one. It is therefore in these easier matters that we should first exercise our minds, but methodically, so that we become accustomed to penetrate each time, by open and recognized paths and almost as in a game, to the inner truth of things. In this way, soon afterward, and in less time than one could hope, we will find ourselves able to deduce with equal ease and from self-evident principles, many propositions which appear very difficult and intricate. But perhaps some will be astonished that in this study, where we are inquiring how we can be made more competent to deduce some truths from others, we omit all the rules by which the logicians think they regulate human reason. These prescribe certain forms of argument which involve such necessary implications that the mind which relies upon this method, even though it neglects to give clear and attentive consideration to the reasoning, can nevertheless reach certain conclusions on the strength of the form of the argument alone.

SAMPLE TITLES

LITERATURE
- 1640L **The Plot Against America** (ROTH)
- 1530L **The Good Earth** (BUCK)
- 1520L **A Fable** (FAULKNER)

INFORMATIONAL
- 1650L **Twenty Years at Hull-House** (ADDAMS)
- 1600L **The U.S. Constitution and Other Key American Writings** (ASSORTED)
- 1600L **Sustaining Life: How Human Health Depends on Biodiversity** (CHIVIAN)
- 1590L **Captain John Smith: A Select Edition of His Writings** (SMITH)
- 1520L **Collapse: How Societies Choose to Fail or Succeed** (DIAMOND)
- 1510L **Original Meanings: Politics and Ideas in the Making of the Constitution** (RAKOVE)

1400L–1495L

1440L *Fordlandia* GRANDIN

As Ford biographer Robert Lacey put it, the "Five Dollar Day raised the pain threshold of capitalism." But beyond an incentive to make workers stay put, it also became a model for how to respond to another crisis that plagued industrialism. The mechanized factory production that took flight during America's Gilded Age had promised equality and human progress but in reality delivered deepening polarization and misery, particularly in sprawling industrial cities like Detroit. Ford, advised by farsighted company executives such as James Couzens and John Lee, understood that high wages and decent benefits would do more than create a dependable and thus more productive workforce; they would also stabilize and stimulate demand for industrial products by turning workers into consumers.

SAMPLE TITLES

LITERATURE
- 1460L **The Legend of Sleepy Hollow** (IRVING)
- 1450L **Billy Budd** (MELVILLE)
- 1420L **The Life All Around Me by Ellen Foster** (GIBBONS)
- 1420L **The Fall of the House of Usher** (POE)
- 1410L **Death in Venice** (MANN)

INFORMATIONAL
- 1490L **Rousseau's Political Writings** (ROUSSEAU)
- 1430L **America's Constitution: A Biography** (AMAR)
- 1410L **Profiles in Courage** (KENNEDY)
- 1400L **The Mysteries of Beethoven's Hair** (MARTIN & NIBLEY)
- 1400L **Life and Times of Frederick Douglass: His Early Life as a Slave, His Escape From Bondage, and His Complete History to the Present Time** (DOUGLASS)

1300L–1395L

1340L *Silent Spring* CARSON

The basic element, carbon, is one whose atoms have an almost infinite capacity for uniting with each other in chains and rings and various other configurations, and for becoming linked with atoms of other substances. Indeed, the incredible diversity of living creatures from bacteria to the great blue whale is largely due to this capacity of carbon. The complex protein molecule has the carbon atom as its basis, as have molecules of fat, carbohydrates, enzymes, and vitamins. So, too, have enormous numbers of nonliving things, for carbon is not necessarily a symbol of life.

SAMPLE TITLES

LITERATURE
- 1390L **The Yellow Wallpaper** (GILMAN)
- 1350L **The Secret Sharer** (CONRAD)
- 1330L **The Jungle** (SINCLAIR)
- 1330L **Silas Marner** (ELIOT)
- 1300L **Gulliver's Travels** (SWIFT)

INFORMATIONAL
- 1390L **In Defense of Food: An Eater's Manifesto** (POLLAN)
- 1360L **Anne Frank: The Book, the Life, the Afterlife** (PROSE)
- 1340L **Walden and Civil Disobedience** (THOREAU)
- 1330L **The Professor and the Madman: A Tale of Murder, Insanity, and the Making of the Oxford English Dictionary** (WINCHESTER)
- 1300L **Arctic Dreams: Imagination and Desire in a Northern Landscape** (LOPEZ)

1200L–1295L

1210L *The Tortilla Curtain* BOYLE

He didn't wake America, not yet. He made four trips up to the ledge and back, with the tools, the sacks of vegetables—they could use the empty sacks as blankets, he'd already thought of that—and as many wooden pallets as he could carry. He'd found the pallets stacked up on the far side of the shed, and though he knew the maintenance man would be sure to miss them, it could be weeks before he noticed and then what could he do? As soon as Qindido had laid eyes on those pallets an architecture had invaded his brain and he knew he had to have them. If the fates were going to deny him his apartment, well then, he would have a house, a house with a view.

SAMPLE TITLES

LITERATURE

1290L	An Old-Fashioned Girl (ALCOTT)
1280L	The House of the Spirits (ALLENDE)
1280L	The Castle (KAFKA)
1220L	The Silent Cry (ŌE)
1210L	Chronicle of a Death Foretold (GARCÍA MÁRQUEZ)

INFORMATIONAL

1290L	A Brief History of Time: From the Big Bang to Black Holes (HAWKING)
1280L	Black, Blue, and Gray: African Americans in the Civil War (HASKINS)
1230L	Stiff: The Curious Lives of Human Cadavers (ROACH)
1230L	Knowing Mandela: A Personal Portrait (CARLIN)
1200L	The Dark Game: True Spy Stories (JANECZKO)

1100L–1195L

1150L *A Room of One's Own* WOOLF

The reason perhaps why we know so little of Shakespeare—compared with Donne or Ben Jonson or Milton—is that his grudges and spites and antipathies are hidden from us. We are not held up by some "revelation" which reminds us of the writer. All desire to protest, to preach, to proclaim an injury, to pay off a score, to make the world the witness of some hardship or grievance was fired out of him and consumed. Therefore his poetry flows from him free and unimpeded. If ever a human being got his work expressed completely, it was Shakespeare. If ever a mind was incandescent, unimpeded, I thought, turning again to the bookcase, it was Shakespeare's mind.

SAMPLE TITLES

LITERATURE

1180L	Sense and Sensibility (AUSTEN)
1170L	The Amazing Adventure of Kavalier & Clay (CHABON)
1150L	Great Expectations (DICKENS)
1140L	Cold Mountain (FRAZIER)
1130L	Democracy (DIDION)

INFORMATIONAL

1160L	The Longitude Prize (DASH)
1160L	In Search of Our Mothers' Gardens (WALKER)
1150L	The Human Microbiome: The Germs That Keep You Healthy (HIRSCH)
1150L	In My Place (HUNTER-GAULT)
1100L	Something to Declare (ALVAREZ)

1000L–1095L

1070L *Geeks: How Two Lost Boys Rode the Internet out of Idaho* KATZ

Geeks were the first to grasp just how much information was available on the Web, since they wrote the programs that put much of it there—movie times and reviews, bus and train schedules, news and opinions, catalogues, appliance instructions, plus, of course, software and its upgrades. And of course, music, the liberation of which is considered a seminal geek accomplishment.

Virtually everything in a newspaper—and in many magazines—is now available online. In fact, some things, like the latest weather and breaking news, appear online hours before they hit print.

Yet while Jesse had gone through literally thousands of downloaded software applications, he'd never paid for any of them. He didn't even quite get the concept. The single cultural exception was books. Perhaps as a legacy of his childhood, Jesse remained an obsessive reader. He liked digging through the bins of used bookstores to buy sci-fi and classic literature; he liked books, holding them and turning their pages.

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	The Hobbit or There and Back Again (TOLKIEN)

INFORMATIONAL

1030L	Phineas Gage: A Gruesome but True Story About Brain Science (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	Harriet Tubman: Conductor on the Underground Railroad (PETRY)
1000L	Claudette Colvin: Twice Toward Justice (HOOSE)

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: An American Legend** (HILLENBRAND)

980L **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** (FINLAYSON)

920L **Talking With Artists** (CUMMINGS)

800L–895L

800L *Moon Over Manifest* VANDERPOOL

We tiptoed down the hall to the second classroom on the right. The heavy wooden door opened easily and we stepped in. There is an eerie, expectant feeling to a schoolroom in the summer. The normal classroom items were there: desks, chalkboards, a set of encyclopedias. The American flag with accompanying pictures of Presidents Washington and Lincoln. But without students occupying those desks and their homework tacked on the wall, that empty summer classroom seemed laden with the memory of past students and past learning that took place within those walls. I strained to listen, as if I might hear the whisperings and stirrings of the past. Maybe Ruthanne was right. Maybe there was more here than met the eye.

SAMPLE TITLES

LITERATURE

GN840L* **The Odyssey** (HINDS)

830L **Baseball in April and Other Stories** (SOTO)

820L **Maniac Magee** (SPINELLI)

810L **Where the Mountain Meets the Moon** (LIN)

800L **Homeless Bird** (WHELAN)

INFORMATIONAL

880L **Volcanoes** (SIMON)

880L **The Circuit: Stories From the Life of a Migrant Child** (JIMÉNEZ)

IG860L* **Animals Nobody Loves** (SIMON)

860L **Through My Eyes: Ruby Bridges** (BRIDGES)

830L **Quest for the Tree Kangaroo** (MONTGOMERY)

700L–795L

700L *The Miraculous Journey of Edward Tulane* DICAMILLO

Edward Tulane waited.

He repeated the old doll's words over and over until they wore a smooth groove of hope in his brain: *Someone will come; someone will come for you.*

And the old doll was right.

Someone did come.

It was springtime. It was raining. There were dogwood blossoms on the floor of Lucius Clarke's shop.

She was a small girl, maybe five years old, and while her mother struggled to close a blue umbrella, the little girl walked around the store, stopping and staring solemnly at each doll and then moving on.

When she came to Edward, she stood in front of him for what seemed like a long time. She looked at him and he looked back at her.

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** (MIOCHIZUKI)

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)

600L–695L

620L *The Year of Billy Miller* HENKES

His heart was pounding.

Once again, he forgot every word of his poem, including the title—but this time he didn't have a copy of it to read from.

He saw Ms. Silver in the fringes of his vision. She was smiling and nodding, urging him on with her wide eyes.

Should he walk over to her to get a copy of his poem? She seemed about a mile away. And he didn't think he could make his legs move.

What should he do?

The air felt weird all of a sudden. As if it had sprouted wings and was brushing against him. The air was fluttering against his arm.

How could that be?

He turned around and Mama was there with a copy of his poem, tapping it lightly against his elbow. "Here," she whispered. "You can do it."

SAMPLE TITLES

LITERATURE
- **690L** Firefly Hollow (MCGHEE)
- **680L** Charlotte's Web (WHITE)
- **670L** A Year Down Yonder (PECK)
- **660L** Holes (SACHAR)
- **610L** Mountain Bike Mania (CHRISTOPHER)

INFORMATIONAL
- **690L** Sadako and the Thousand Paper Cranes (COERR)
- **680L** An Eye for Color: The Story of Josef Albers (WING)
- **680L** The Moon (LANDAU)
- **660L** Remember: The Journey to School Integration (MORRISON)
- **620L** Crittercam (EINSPRUCH)

500L–595L

500L *The Curse of the Cheese Pyramid* STILTON

Trap winked at me and announced, "Grandfather has hired me to be his personal cook!"

This was ridiculous! I was getting hotter than a bag of cheese popcorn in a microwave. Who would help me run the paper?

At that moment, I felt a tug on the sleeve of my jacket. It was my young nephew Benjamin. "Uncle Geronimo, guess what?" he beamed. "Great-grandfather William has hired me to be his personal assistant!"

Grandfather stroked Ben's tiny ears.

"Ah, the family, there's nothing like the family! The Stilton Family, that is…" I snorted. I could see I was the workmouse of the family. It looked like I would be the only one doing any work!

SAMPLE TITLES

LITERATURE
- **590L** The Great Kapok Tree (CHERRY)
- **580L** Tops and Bottoms (STEVENS)
- **570L** Grace for President (DIPUCCHIO)
- **540L** Ron's Big Mission (BLUE & NADEN)
- **500L** Poppleton in Spring (RYLANT)

INFORMATIONAL
- **IG590L*** Claude Monet (CONNOLLY)
- **580L** What Magnets Can Do (FOWLER & BARKAN)
- **560L** Molly the Pony (KASTER)
- **550L** Martin Luther King, Jr. and the March on Washington (RUFFIN)
- **510L** A Picture for Marc (KIMMEL)

400L–495L

470L *Frog and Toad Are Friends* LOBEL

Toad said, "Frog, you are looking quite green."
"But I always look green," said Frog. "I am a frog."
"Today you look very green even for a frog," said Toad.
"Get into my bed and rest."
Toad made Frog a cup of hot tea.
Frog drank the tea, and then he said, "Tell me a story while I am resting."
"All right," said Toad.

SAMPLE TITLES

LITERATURE
- **480L** A Birthday for Frances (HOBAN)
- **470L** Tales of a Fourth Grade Nothing (BLUME)
- **450L** Amelia Bedelia (PARISH)
- **440L** Fox on the Job (MARSHALL)
- **420L** Hey, New Kid! (DUFFEY)

INFORMATIONAL
- **480L** Rally for Recycling (BULLARD)
- **480L** Grand Canyon (GILBERT)
- **470L** Life in China (CHUNG)
- **460L** Half You Heard of Fractions? (ADAMSON & ADAMSON)
- **440L** Abraham Lincoln (HANSEN)

300L–395L

330L *Seals* ARNOLD

Earless seals live in oceans.
Thick blubber keeps seals warm.
A seal's back flippers help it swim fast.
A seal on land is slow.
Its claws dig into rocks and ice.
Many seals have dark brown or gray fur.
Some have spots.
Seals molt every year.

SAMPLE TITLES

LITERATURE

370L	**Little Bear Book** (MINARIK)
350L	**To the Rescue!** (MAYER)
340L	**Snow** (SHULEVITZ)
GN320L*	**Spotlight Soccer** (SANCHEZ)
310L	**I Spy Fly Guy!** (ARNOLD)

INFORMATIONAL

370L	**Starfish** (HURD)
IG340L*	**We Can Be Friends** (JORDAN)
340L	**Fernando Exercises!: Tell and Write Time** (KAY)
340L	**Simple Machines** (RISSMAN)
310L	**Visiting the Beach in Summer** (FELIX)

200L–295L

220L *Put Me in the Zoo* LOPSHIRE

Look at this, now! One! Two! Three!
I can put them on a tree.
And now when I say "One, two, three"
All my spots are back on me!
Look, now!
Here is one thing more. I take my spots. I make them four.
Oh! They would put me in the zoo, if they could see what I can do.

SAMPLE TITLES

LITERATURE

290L	**The Class Pet From the Black Lagoon** (THALER)
280L	**Puddle** (YUM)
240L	**Are You My Mother?** (EASTMAN)
210L	**Green Eggs and Ham** (SEUSS)
200L	**Tiny Goes to the Library** (MEISTER)

INFORMATIONAL

280L	**Whales** (LINDEEN)
260L	**Leaves in Fall** (SCHUH)
220L	**Plants on a Farm** (DICKMANN)
210L	**Counting in the City** (STEFFORA)
210L	**The Tractor Race** (SCHUH)

* GN DENOTES GRAPHIC NOVEL, IG DENOTES ILLUSTRATED GUIDE

Please note:

The Lexile measure (text complexity) of a book is an excellent starting point for a student's book selection. It's important, though, 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 www.Lexile.com.

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	740–1010L
2–3	420L–820L
1	190L–530L

Common Core State Standards for English Language Arts, Appendix A (Additional Information), NGA and CCSSO, 2012

MetaMetrics

LINKING ASSESSMENT WITH INSTRUCTION