TECHNICAL REPORT

The Lexile® Framework for Reading

Linking the Classic Learning Test of Verbal Reasoning

Bringing Meaning to Measurement

800 Taylor Street,
Suite 102
Durham, NC 27701

MetaMetricsInc.com

November 2023
Redacted

Linking the Classic Learning Test of Verbal Reasoning to the Lexile® Framework for Reading

Linking Study Report
Redacted

Prepared by MetaMetrics for Classic Learning Initiatives under License Agreement, signed September 28th, 2022.

MetaMetrics
800 Taylor Street, Suite 102
Durham, NC 27701
https://lexile.com/

November 2023

Table of Contents

List of Appendices

List of Tables

List of Figures

Introduction

It is often desirable to convey more information about student test performance than what can be gleaned from a raw score, percentage correct, or scale score. It can be hard to explain what a student can read based on the results of a reading test alone. Students may ask, "Based on my test results, what can I read and how well?" Auxiliary score scales can be used to "convey additional normative information, test-content information, and information that is jointly normative and content based" (Petersen, et al., 1989, p. 222).

One such auxiliary scale is The Lexile® Framework for Reading, which was developed to appropriately match students with texts that provide challenge but not frustration. Linking assessments to the Lexile scale can provide context for understanding the results of an assessment in terms of student reading ability. Once a linkage is established with an assessment, the results of the linked assessment can be explained and interpreted in the context of the specific books and titles that a student can read.

Linking assessment results with the Lexile Framework provides a mechanism for matching each student's reading ability with text on a common scale. It serves as an anchor to which texts and assessments can be connected, allowing parents, teachers, and administrators to speak the same language regarding test results. 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.

Current Study. The current study was conducted by MetaMetrics® for Classic Learning Initiatives under License Agreement, signed September 28, 2022, to determine a mechanism to provide reading levels to students so that they can be matched with text based on their performance on the Classic Learning Test (CLT) Levels 3–6 and 8.

The primary purposes of this study were to:

- link the theta scales for CLT3–CLT6 and CLT8 Verbal Reasoning to the Lexile Framework;
- develop tables for converting CLT3–CLT6 and CLT8 Verbal Reasoning scale scores to Lexile reading measures;
- present a solution for matching students with text; and
- produce a report that describes the linking analysis procedures.

By linking the CLT3–CLT6 and CLT8 Verbal Reasoning with the Lexile Framework, educators and parents will be able to better answer the questions posed above and will be better able to use CLT Verbal Reasoning results to improve instruction and to develop each student's level of reading comprehension.

Classic Learning Test Verbal Reasoning–Lexile Framework Linking Process

Description of the Assessments

The Classic Learning Test (CLT). The CLT is a comprehensive assessment suite intended to provide an alternative college entrance exam (Grades 11 and 12) (Tyler, 2018), the CLT10 is a college preparatory exam (Grades 9 and 10) (Tyler, 2018), and the CLT8 is a high school readiness exam (Grades 7 and 8) (Gardner, 2021). The CLT3 through CLT6 are under development (Classic Learning Initiatives, 2022) and are intended to provide diagnostic and summative measurement for students in Grades 3 through 6. The objective of all of the CLT assessments is to provide a more meaningful and positive test-taking experience for students.

The CLT3–6 assessments each contain two Verbal Reasoning and two Quantitative Reasoning sections and each comprises 140 total questions (Classic Learning Initiatives, 2022). The CLT8 contains one Verbal Reasoning, one Grammar/Writing, and one Quantitative Reasoning section, and comprises 120 total questions (Gardner, 2021). The Verbal Reasoning and Grammar/Writing sections are designed to assess student ability in reading comprehension and analysis as well as grammar skills, including sentence structure, spelling, and punctuation. The Quantitative Reasoning sections are designed to measure student ability in arithmetic, computation, and mathematical reasoning. For this study, six forms were administered for each level CLT 3–6 and one form for CLT8, with the goal of assembling two operational forms for each level from the best performing items in the forms.

The blueprints for the CLT3–6 and CLT8 assessments contain the following reporting categories:

- CLT3–6
    - Grammar (Orthography, Parts of Speech, and Sentence Structure and Diagramming);
    - Reading Comprehension and Writing (Analysis, Reading Comprehension, and Writing Concepts and Skills).
- CLT8
    - Verbal Reasoning (Comprehension and Analysis);
    - Grammar/Writing (Grammar and Writing).

For each level from CLT3–CLT6, all of the items on the six forms were calibrated on the same logit scale. Consequently, all of the students in the sample taking one of the six forms could be assigned a theta measure of their verbal reasoning ability that also lay on that same logit scale, and their thetas could therefore be combined when constructing a link between the thetas and the Lexile reading measures. This, in turn, means that when the items from the six forms were winnowed down to two operational forms, the thetas from those forms would also be on the same logit scale, and the theta to Lexile link based on the six forms would also be appropriate for these two forms. For each of the operational forms, CLT created a separate linear formula that converted the thetas to scales scores having a range from 150 to 300.

The Lexile Framework. The Lexile Framework is a tool that helps teachers, parents, and students locate appropriate reading materials. Text complexity (difficulty) and reader ability are measured in the same unit—the Lexile. Text complexity is determined by examining such characteristics as word frequency and sentence length. Items and text are calibrated using the Rasch model as implemented in the Winsteps Rasch Analysis and Rasch Measurement program (Linacre, 2022). The typical range of the Lexile Scale is from 200L to 1600L, although actual Lexile reading measures can range from below BR400L (BR = Beginning Reader) to above 1600L.

The Lexile linking items consist of multiple-choice items focused on the skills readers employ when studying written materials sampled from various content areas, including both literary and informational text. Lexile items do not require prior knowledge of ideas outside of the passage, vocabulary taken out of context, or formal logic. Each test item consists of a passage, a cloze (a statement that is added at the end of the passage with a missing word or phrase), and four options (one correct choice and three distractors). The skills measured by these items include referring to details in the passage, drawing conclusions, and making comparisons and generalizations.

Lexile linking item pools were developed for administration to students in Grades 3 through 6, and a 40-item test form was developed for Grades 7 and 8. For Grades 3 through 6, MetaMetrics provided CLT with 30 linking items per Grade level divided into 8 subsets each, featuring common items across subsets. Each linking item pool and test form contained an array of items varying in complexity within the typical range of complexity established by Lexile grade-level norms. The range of item complexity for each grade-level item pool and test form and the grade-level item Lexile reading measure mean were determined by examining test information provided by CLT, as well as national normative data and information from previously administered English Language Arts Lexile linking tests. The mean Lexile reading measures for the item pools were CLT3, 714L; CLT4, 912L; CLT5, 972L; CLT6, 1033L; and CLT7 and 8, 1124L.

Common items were included to provide connectivity across grades, resulting in a total of 97 unique items. The items were embedded in the CLT assessment for online administration.

Study Design

A non-equivalent anchor test design was chosen for this study (Dorans & Holland, 2000). 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 (Kolen & Brennan, 2014, pp. 16–17).

The Lexile linking items for levels CLT3–CLT6 were embedded in the six different forms administered at each level between May 8 and June 2, 2023. A spiral administration of the six forms within a classroom meant that even though the forms were fixed and administered as booklets, the forms and their items were being assigned to students in a approximately random manner. Each of the six forms had four five-item blocks of linking items, with a total of 30 linking items being spread across the six forms. The Lexile linking items for CLT8 were appended to the end of the single form which was also administered between May 8 and June 2, 2023. There were common linking items across all the levels (CLT3-CLT6 and CLT8) and forms.

Results

A total of 14,432 student records were provided to MetaMetrics containing student demographic information, responses to the CLT tests, and responses to the embedded/appended Lexile linking items.

Evaluation of Lexile Reading Linking Items. After administration, the performance of all Lexile linking items was reviewed and evaluated for use in the linking study based on the following criteria:

- item difficulty (i.e., extreme *p*-values less than 0.10 or greater than 0.90);
- construct validity of the item as evidenced by alignment of student performance on the item to the student's ability measure (i.e., point-measure correlation greater than 0.10);
- misfit to the Lexile theory (i.e., substantial theory-observed difference in the study); and
- misfit to the Rasch model (i.e., infit statistic greater than 1.5 or an outfit statistic greater than 2.0; Linacre, 2022).

Based on the evaluative criteria, no linking items were flagged for removal from *Table 1*. The total number of examinees that encountered individual linking items varied considerably with some items seen by as few as 770 students taking CLT8 to as many as 7,669 students taking the same item that was common to levels CLT3–CLT5.

Table 1. Item statistics from the administration of the Lexile linking items

Test Level	*N* Persons* (Range)	*N** Items	Percent Correct Mean (Range)	Point-Measure Mean (Range)
3	1,541 – 7,669	30	0.79 (0.41 – 0.95)	0.46 (0.33 – 0.57)
4	1,526 – 7,669	30	0.71 (0.32 – 0.95)	0.48 (0.36 – 0.55)
5	1,469 – 7,669	30	0.70 (0.29 – 0.93)	0.46 (0.33 – 0.56)
6	1,384 – 6,540	30	0.74 (0.37 – 0.93)	0.47 (0.35 – 0.59)
8	770 – 6,540	40	0.74 (0.42 – 0.95)	0.46 (0.30 – 0.59)
All	770 – 7,669	97	0.73 (0.29 – 0.95)	0.46 (0.30 – 0.59)

* N *(Persons) reflects the removal of 875 misfitting persons.*

Description of the Student Samples. Subsequent to item evaluation, three student samples were used for the linking process:

- An initial sample was established by removing invalid student records.
- A calibration sample was established to evaluate the performance of the Lexile reading linking items, to calibrate and place the CLT items on the Lexile scale, and to express student results in the Lexile reading metric.
- Finally, a linking sample was established to link the CLT thetas with the Lexile reading measures.

Counts of students for each of these samples are presented in *Table 2* and *Table 3*.

Initial Sample. The initial sample was established by removing 269 students due to a missing grade or off-grade testing, 279 due to missing test sections, and 325 due to having a read-aloud

accommodation, leaving 14,432 students as the initial sample after undergoing data cleaning and de-identification.

Calibration Sample. The linking item responses of the initial sample of student records were submitted to Rasch analysis using the Winsteps program (Linacre, 2022). This process helped identify student responses that exhibited misfit to the Rasch model, indicated by an infit statistic greater than 1.5 or an outfit statistic greater than 2.0 (Linacre, 2022). Additional misfitting students were identified when the linking and CLT items were combined in order to calibrate the CLT item difficulties.

Records misfitting the Rasch model represented about 7% of the total data, ranging from 7% to 9% across test levels – well below the threshold for misfit established by MetaMetrics (i.e., 15%). Overall the calibration sample included 12,609 students (approximately 93% of the initial sample).With limited and similar exclusion of records across test levels, the calibration sample may be considered sufficiently large and representative for the purposes of this study.

Table 2. Number of records in the initial and calibration samples.

Test Level	*N* Initial Sample*	*N* Removed Based on Misfit Person on Linking Test Only	*N* Linking Item Calibration Sample	Percent of Initial Sample	*N* Removed Based on Misfit Person on Linking and Target Test	*N* Target Item Calibration Sample	Percent of Initial Sample
3	3,377	253	3,124	92.51	31	3,093	91.59
4	3,286	210	3,076	93.61	9	3,067	93.34
5	3,141	202	2,939	93.57	7	2,932	93.35
6	2,934	159	2,775	94.58	24	2,751	93.76
8	821	51	770	93.79	4	766	93.30
All	13,559	875	12,684	93.55	75	12,609	92.99

* 873 students were removed prior to the initial sample. 269 removed for off-level or missing enrolled grade, 279 removed for missing test section(s), and 325 removed for read-aloud accommodation.

Linking Sample. The sample used to link the CLT levels to the Lexile reading scale was derived from the initial sample, with records removed according to the following criterion:

- Cases with perfect scores (i.e., raw scores of 0% or 100% correct).

The resulting linking sample contained a total of 13,557 records, or almost 100% of the initial sample (see *Table 3*).

Table 3. Number of records in the initial and linking samples.

Test Level	*N* Initial Sample	Perfect Raw: 0% and 100%	*N* Linking Sample	Percent of Initial Sample
3	3,377	1	3,376	99.97
4	3,286	0	3,286	100.00
5	3,141	1	3,140	99.97
6	2,934	0	2,934	100.00
8	821	0	821	100.00
All	13,559	2	13,557	99.99

Demographic Information. A summary of the demographic information provided to MetaMetrics along with the response data is provided in *Table 4*. Recall that the calibration and linking samples are subsamples of the initial sample. As can be seen, the demographic characteristics of these subsamples were comparable to the initial sample. This demonstrates that the records removed for the various reasons stated had minimal effect on the demographic characteristics of the subsamples.

Table 4. Percentage of initial, calibration, and linking sample for selected demographic characteristics.

Student Characteristic	Value	Initial Sample *N* = 13,559	Linking Item Calibration Sample *N* = 12,684	Target Item Calibration Sample *N* = 12,609	Linking Sample *N* = 13,557
Grade	3	24.91	24.63	24.53	24.90
	4	24.23	24.25	24.32	24.24
	5	23.17	23.17	23.25	23.16
	6	21.64	21.88	21.82	21.64
	7	2.77	2.77	2.75	2.77
	8	3.28	3.30	3.32	3.28
Gender	Female	49.60	49.94	50.04	49.60
	Male	49.07	48.75	48.65	49.07
	Not Provided	1.33	1.31	1.32	1.33
Race/ Ethnicity	American Indian/Alaskan Native	0.44	0.45	0.45	0.44
	Asian	3.74	3.78	3.79	3.74
	Black or African American	2.57	2.54	2.55	2.57
	Hispanic or Latino	7.66	7.55	7.53	7.66
	Native Hawaiian or Other Pacific Islander	0.30	0.27	0.27	0.30
	Other	3.35	3.36	3.36	3.35
	White	56.57	56.78	56.86	56.58
	Not Provided	25.37	25.27	25.19	25.36

Item Calibration and Scoring

Three steps were performed prior to the linking analysis for each CLT test level.

- A calibration of all Lexile reading linking items was conducted to evaluate how well student responses had adhered to the Rasch model, and to assess the appropriateness of using the theoretical difficulties assigned to them by the Lexile Framework to place the CLT Verbal Reasoning item difficulties on the Lexile reading scale.
- A concurrent calibration of the CLT Verbal Reasoning items, with the difficulties of the Lexile reading linking items serving as anchors, was conducted to place the CLT Verbal Reasoning items on the Lexile reading scale.
- A scoring run was conducted using only the CLT Verbal Reasoning items on the Lexile scale to express student performance on the CLT Verbal Reasoning test in the Lexile reading metric (referred to as "calibrated Lexile reading measures").

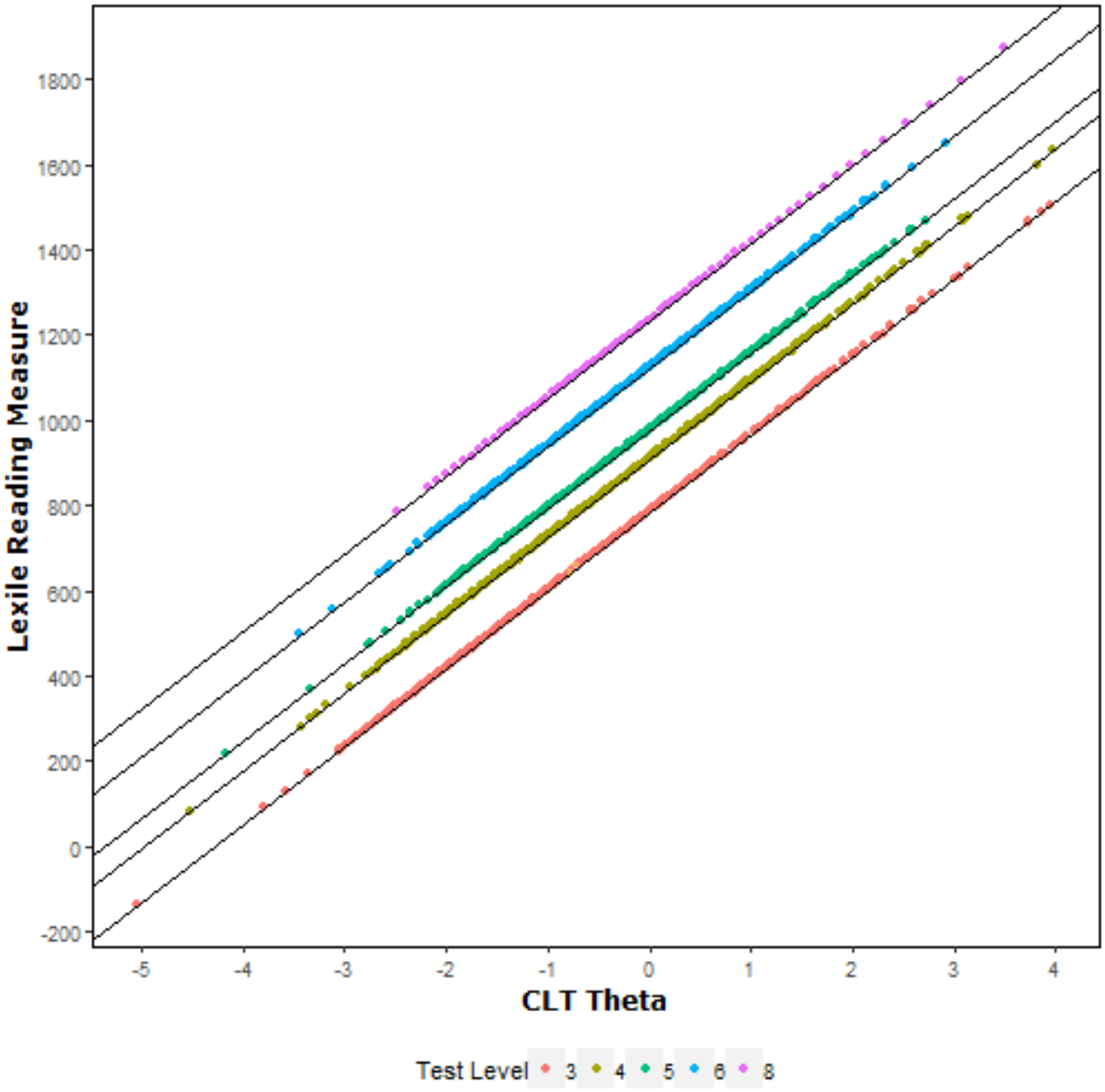
Table 5 provides the descriptive statistics from the CLT Verbal Reasoning thetas and their calibrated Lexile reading measures. Test level means for Calibrated Lexile reading measures increase by test level, demonstrating the vertical scale of the Lexile Framework, as expected. The correlation between scales was approximately 1.00 for each test level, indicating that student performance on the CLT is consistent with performance based on the calibrated CLT item difficulties derived from anchoring them on the Lexile linking test items.

Table 5. Descriptive statistics for the CLT Verbal Reasoning[i] thetas and the calibrated Lexile reading measures for the linking sample, by test level.

Test Level	*N*	CLT Theta Mean (SD)	Calibrated Lexile Measure Mean (SD)	*r*
3	3,376	[redacted]	[redacted]	1.00
4	3,286	[redacted]	[redacted]	1.00
5	3,140	[redacted]	[redacted]	1.00
6	2,934	[redacted]	[redacted]	1.00
8	821	[redacted]	[redacted]	1.00

Figure 1 shows the relationship between student theta values on *CLT Verbal Reasoning* test levels 3 through 6 and 8 and the corresponding calibrated Lexile reading measures for the linking sample. As might be expected from the correlations in *Table 5*, the relationship between the CLT Verbal Reasoning thetas and the calibrated Lexile reading measures is essentially linear, with the slopes appearing to be nearly parallel and the intercepts increasing as a function of level.

*Figure 1. CLT3–CLT6 and CLT8 Verbal Reasoning thetas and their calibrated Lexile reading measures, linking sample (*N = *13,557).*



Linking the CLT Verbal Reasoning thetas with the Lexile Reading Scale

Linking in general means "putting the scores from two or more tests on the same scale" (National Research Council, 1999, p.15). Two score scales can be linked using linear equating when (1) items have similar difficulties and (2) simplicity in developing conversion tables or equations, in conducting analyses, and in describing procedures are desired (Kolen & Brennan, 2014).

In scale alignment which uses the same methods as linear equating (Dorans, et al., 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, et al., 1989; Kolen & Brennan, 2014). Given scores *x* and *y* on tests *X* and *Y*, the linear relationship is:

$$\frac{(x-\mu_X)}{\sigma_X} = \frac{(y-\mu_Y)}{\sigma_Y} \qquad \text{Equation (1)}$$

and the linear transformation $l_Y(x)$ (called the SD line in this report) used to transform scores on test *X* to scores on test *Y* is:

$$y = l_Y(x) = \left(\frac{\sigma_Y}{\sigma_X}\right)X + \left(\mu_Y - \frac{\mu_X\sigma_Y}{\sigma_X}\right) \qquad \text{Equation (2)}$$

Linear linking by definition has the same mean and standard deviation for the linking equation because the means and standard deviations are the same for the tests being linked. Linear linking using an SD-line approach is preferable to linear regression because the tests are not perfectly correlated. With less than perfectly correlated 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.

MetaMetrics and Classic Learning Initiatives conducted this linking study to provide information that could be used to match 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 CLT Verbal Reasoning test. To achieve this, a linear linking function between CLT Verbal Reasoning thetas and Lexile reading measures was constructed according to the following equation:

$$\text{Lexile reading measure} = \text{Slope}_g(\text{CLT Verbal Reasoning theta}) + \text{intercept}_g \qquad \text{Equation (3)}$$

where the slope is the ratio of the standard deviations of the CLT Verbal Reasoning thetas and the Lexile reading measures and *g* represents the test levels. Separate linear linking functions between CLT Verbal Reasoning thetas and Lexile reading measures were constructed for CLT3–CLT6 and CLT8.

Table 6 provides the slopes and intercepts used to transform CLT Verbal Reasoning thetas to Lexile reading measures.

Table 6. Linear linking equation slopes and intercepts used to transform CLT Verbal Reasoning thetas to Lexile reading measures.

Test Level	Slope	Intercept
3	[redacted]	[redacted]
4	[redacted]	[redacted]
5	[redacted]	[redacted]
6	[redacted]	[redacted]
8	[redacted]	[redacted]

Table 7 provides the slopes and intercepts that were developed by Classic Learning Initiatives to transform CLT Verbal Reasoning thetas and forms to CLT Verbal Reasoning scale scores for their new operational test levels. CLT7 appears in this table because CLT used a subset of the CLT8 items to create a CLT7 level test. As the items in this new test retained the difficulties they had when the CLT8 was administered, CLT could create a linear formula to convert the CLT7 thetas to scale scores (150-300), which could then be combined with the CLT8 theta to Lexile conversion formula to create a scale score to Lexile conversion formula.

Table 7. Linear linking equation slopes and intercepts used to transform CLT Verbal Reasoning thetas to CLT Verbal Reasoning scale scores.

Test Level	Slope	Intercept
3	[redacted]	[redacted]
4	[redacted]	[redacted]
5	[redacted]	[redacted]
6	[redacted]	[redacted]
7	[redacted]	[redacted]
8	[redacted]	[redacted]

The linear functions in *Table 6* and *Table 7* were combined so that CLT Verbal Reasoning scale scores could be transformed into Lexile reading measures. *Table 8* provides the slopes and intercepts of the resulting linear functions for each new operational form and level which can then be used by Classic Learning Initiatives to express the CLT Verbal Reasoning scale scores in the Lexile reading metric.

Table 8. Linear linking equation slopes and intercepts used to transform CLT Verbal Reasoning scale scores to Lexile reading measures.

Test Level	Slope	Intercept
3	[redacted]	[redacted]
4	[redacted]	[redacted]
5	[redacted]	[redacted]
6	[redacted]	[redacted]
7	[redacted]	[redacted]
8	[redacted]	[redacted]

Validity of the CLT Verbal Reasoning–Lexile Links

This section provides multiple sources of validity evidence for the link between the CLT Verbal Reasoning scale and the Lexile scale.

1. The consistency of scores between the calibrated Lexile reading measures and the linked Lexile reading measures is examined to evaluate the generalizability of the link.
2. The consistencies of the distributions based on the calibrated and the linked Lexile reading measures are compared for selected percentiles.
3. The linked Lexile reading measures are compared across grades to the 25th, 50th, and 75th percentiles of the Lexile reading user norms.
4. The linked Lexile reading measures distributions are compared across grades.

The sample used for these presentations is the initial sample unless otherwise stated (N = 13,559).

Generalizability of Linking Study Results. Table 9 provides the descriptive statistics from the CLT Verbal Reasoning calibrated and linked Lexile reading measures, based on the theta to Lexile reading conversion formulas. Because of the essentially perfect correlations between the thetas and the calibrated Lexile reading measures, the two scoring methods yielded nearly identical Lexile reading measures between the calibrated Lexile reading measures and the linked Lexile reading measures. These results, therefore, provide evidence to support the use of the linear linking functions.

Table 9. Descriptive statistics for the calibrated Lexile reading measures and the linked Lexile reading measures by test level, linking sample.

Test Level	*N*	Calibrated Lexile Reading Measure Mean (SD)	Linked Lexile Reading Measure Mean (SD)	*r*
3	3,376	[redacted]	[redacted]	1.00
4	3,286	[redacted]	[redacted]	1.00
5	3,140	[redacted]	[redacted]	1.00
6	2,934	[redacted]	[redacted]	1.00
8	821	[redacted]	[redacted]	1.00

Percentile Rank Distributions. Table 10 presents a comparison of the student Lexile reading measures for selected percentiles based on the CLT Verbal Reasoning-calibrated and linked Lexile reading measures. All six grades from Grades 3 through 8 are reflected in this table, as Grades 7 and 8 both used the linking function that had been based on the CLT8. Differences that were essentially zero were observed throughout the distributions for each grade. These results provide evidence that the calibrated Lexile reading measures and the linear linked Lexile reading measure score students nearly identically and support the use of the linked Lexile reading measures.

Table 10. Comparison of the Lexile reading measures for selected percentile ranks from CLT Verbal Reasoning-calibrated Lexile reading measures and linked Lexile reading measures.

Grade 3		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	320L	323L
5	457L	458L
10	533L	535L
25	664L	663L
50	794L	794L
75	915L	915L
90	1010L	1010L
95	1071L	1072L
99	1196L	1198L

Grade 4		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	509L	509L
5	625L	627L
10	686L	687L
25	790L	791L
50	910L	913L
75	1027L	1028L
90	1131L	1129L
95	1189L	1189L
99	1310L	1309L

Grade 5		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	626L	627L
5	713L	713L
10	776L	775L
25	874L	873L
50	976L	975L
75	1083L	1080L
90	1170L	1171L
95	1227L	1226L
99	1346L	1349L

Grade 6		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	762L	765L
5	867L	867L
10	924L	925L
25	1021L	1021L
50	1123L	1123L
75	1226L	1225L
90	1316L	1314L
95	1380L	1378L
99	1476L	1476L

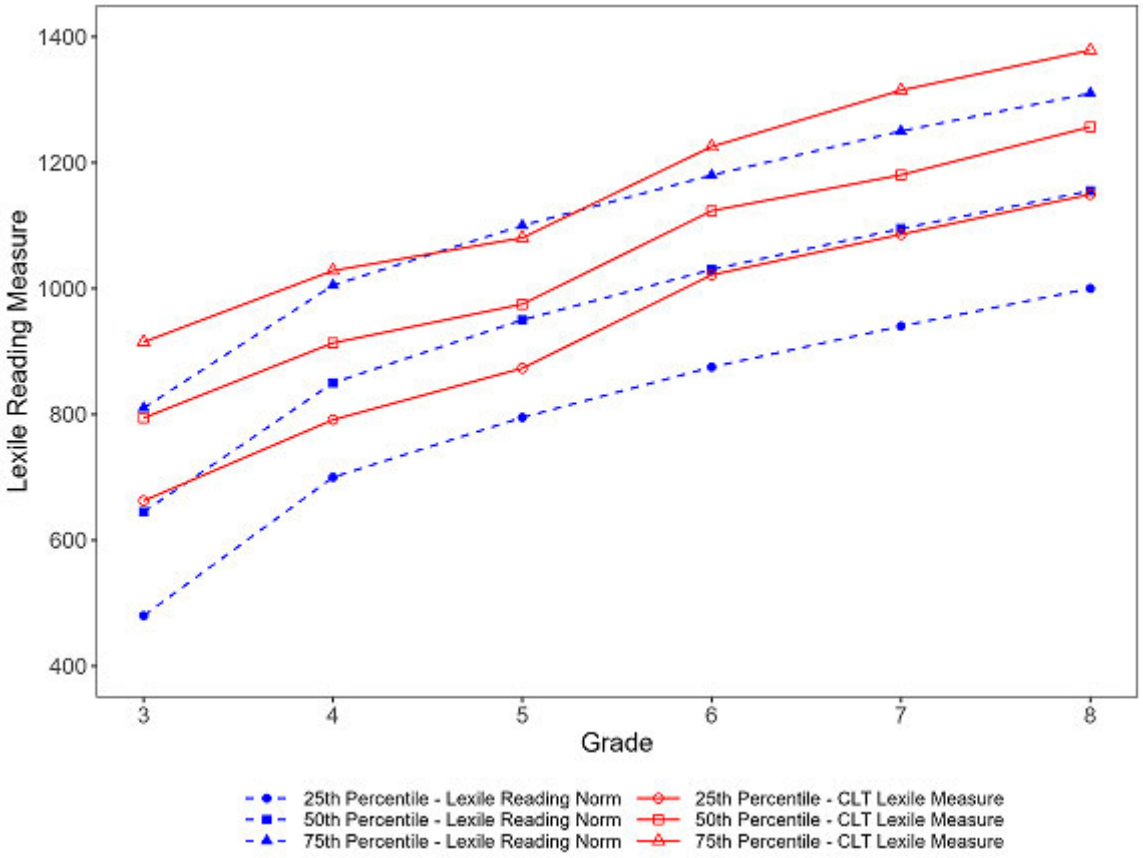
Grade 7		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	881L	881L
5	945L	946L
10	995L	995L
25	1086L	1086L
50	1181L	1180L
75	1314L	1315L
90	1414L	1413L
95	1457L	1457L
99	1666L	1666L

Grade 8		
Percentile Rank	**Calibrated Lexile Measure**	**Linked Lexile Measure**
1	904L	904L
5	983L	984L
10	1042L	1042L
25	1150L	1149L
50	1257L	1257L
75	1378L	1378L
90	1487L	1488L
95	1596L	1597L
99	1795L	1797L

The Lexile Framework Norms. Figure 2 shows the student performance in linked Lexile reading measures from the initial sample compared to the user norms developed for use with the Lexile Framework. Selected percentiles (i.e., 25th, 50th, and 75th) are presented for both the linked Lexile reading measures and the Lexile Framework user norms[ii].

The general pattern demonstrates that the student sample is consistently above the Lexile reading user norms, though Grade 5 dips slightly. In addition, the 25th, 50th, and 75th percentile lines are all roughly parallel to each other. The steady, monotonic increases in the CLT Verbal Reasoning measures as a function of grade across all three percentile curves is indicative that, even though separate links were created for each CLT level, the system of links yield a continuum of reading comprehension ability as anticipated by the Lexile Framework.

*Figure 2. Selected percentiles (25th, 50th, and 75th) plotted for the CLT Verbal Reasoning Lexile reading measures (*N = *13,559) in relation to the Lexile reading user norms.*



Grade-Level Progressions. The box-and-whisker plots in *Figures 3* and *4* show the distributions of the calibrated and linked Lexile reading measures across the Grades 3 through 8 of the initial sample. For each grade/test level, the box refers to the interquartile range, the line within the box indicates the median, the plus symbol indicates the mean, and the trend line connects each box at the median. The end of each whisker represents the 5th and 95th percentile values of the scores (the *y*-axis).

Figures 3 and *4* demonstrate the vertical nature of the Lexile reading scale for the calibrated and the linked Lexile reading measures, respectively. The scores in *Figures 3* and *4* increase as grade level increases and the score distributions for adjacent grades overlap. The "overlap across grades" is characteristic of vertical scales (Kolen & Brennan, 2014). The comparability of the observed grade-over-grade or across course patterns provides evidence that the calibration process and the linking function maintain the observed patterns in the CLT Verbal Reasoning sample.

*Figure 3. CLT Verbal Reasoning-calibrated Lexile reading measures (*N = *13,559).*

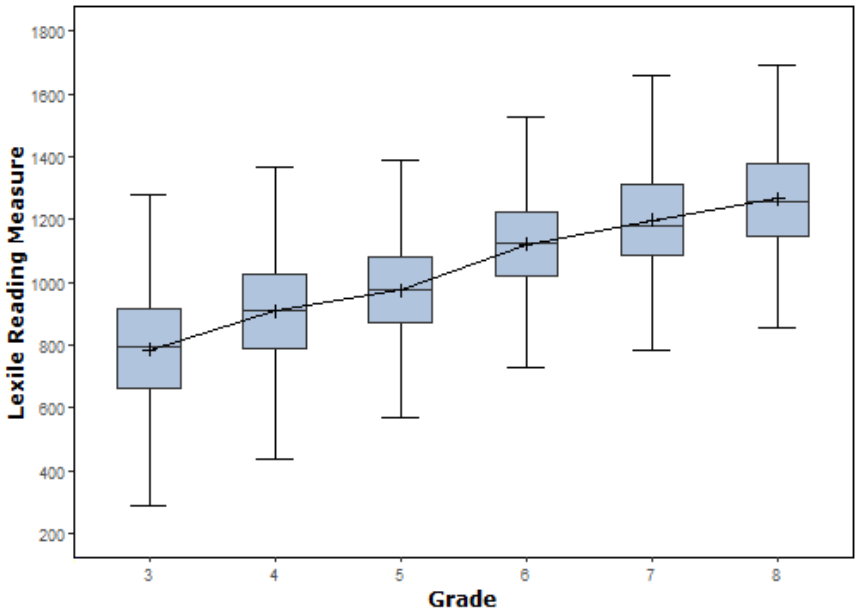
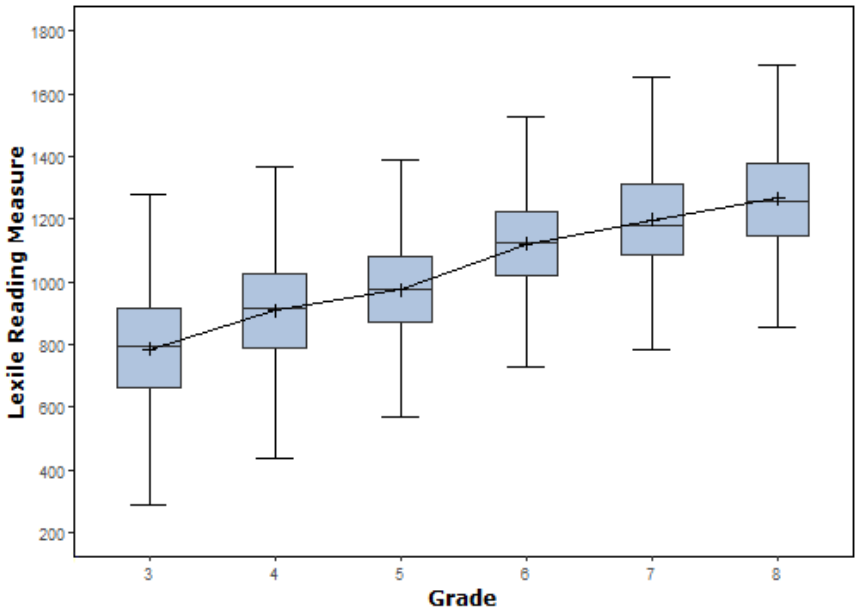


*Figure 4. CLT Verbal Reasoning-linked Lexile reading measures (*N = *13,559).*



Conclusions

The purpose of this study was to establish and provide validity evidence for a linkage between the CLT Verbal Reasoning thetas and scale scores in CLT levels 3 through 6 and level 8 and the Lexile scale. This linking procedure included processes to ensure that a similar construct was measured by the CLT Verbal Reasoning assessment and the Lexile Framework.

A non-equivalent anchor test design was employed because it was logistically possible to administer two sets of test items to the same group of students (Kolen & Brennan, 2014). The linking study was conducted through three major phases: (i) evaluating the linking procedure, (ii) linking two score scales using linear equating, and (iii) providing validity evidence for the linking.

Lexile reading linking items with difficulty levels in the same range as those of the CLT Verbal Reasoning items were selected to enhance the validity of the linking procedures. The Lexile linking items exhibited similar psychometric properties with those of CLT Verbal Reasoning items, including fit to the Lexile theory and sufficient classical item statistics. Strong correlations between the CLT Verbal Reasoning thetas and the calibrated Lexile reading measures were observed, indicating that both scales yield consistent scores.

To evaluate the linking procedures, scatter plots between the Lexile reading measures and the CLT Verbal Reasoning thetas were examined to reinforce the appropriateness of using linear linking methodologies. Linear functions were constructed to transform CLT Verbal Reasoning scale scores to Lexile reading measures based on the calibrated Lexile reading measures. Finally, conversion tables were developed in order to express CLT Verbal Reasoning scale scores in the Lexile metric.

To support the generalizability of the reported Lexile reading measures, the differences between percentile ranks of the CLT Verbal Reasoning calibrated and linked Lexile reading measures were evaluated. Minimal differences were observed throughout the distributions for each grade. These results provide evidence that the calibrated and linked Lexile reading measures score students similarly and support the use of the linear function to link with the Lexile Framework.

Finally, the linking procedures yielded test-level specific constants which reflect the intended interpretations of both the CLT Verbal Reasoning scores and Lexile scores.

Caveats

Study Design Limitations: The slopes and intercepts of the linear linking equations were based on student ability estimates ("thetas") since the operational scale score for CLT Verbal Reasoning was under development at the time. However, linking Lexile reading scores to operational CLT Verbal Reasoning scale scores can be achieved by modification of the linking formulas, provided that the final transformations of thetas to scale scores preserves the linear function of the relationship. Additionally, linking to a non-administered CLT7 test form was made possible by designating a subset of CLT8 items for use on this test form. All these links are valid, therefore, under the assumption that the operational test forms are based on the same scales as those reflected in the test forms employed in this study.

Lexile Reading Measures and Grade Levels. Lexile reading measures do not translate specifically to grade levels. Within any grade, there will be a range of students and a range of materials to be read. In a sixth-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 sixth graders assumes that all sixth graders are reading at the same level. The Lexile Framework can be used to match students with texts at whatever level the student 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 reading measure for a grade indicates that the student can read grade-appropriate materials at a higher comprehension level (90%, for example).

Maintenance of the CLT Verbal Reasoning scales. Maintenance of the focal scales (i.e., CLT3 through CLT8 scales) is critical to the validity of any link with an auxiliary scale (i.e., Lexile scale). If an update occurs to a focal scale, the integrity of the link should be re-evaluated and additional linking studies may be needed to accommodate fundamental changes to the focal scale. Such updates may include, but are not limited to, incorporating new item types into the assessment; revising item calibrations; or revising the assessment program and the reported scale scores.

Linking error. Error in estimating the linking relationship of two scales is present whenever linking is conducted. Not all error associated with a study can be accounted for, however error should be continually investigated to ensure scores are as accurate and reliable as possible. The two sources of error present are random error and systematic error. Random linking error occurs when directly estimating the linking relationship because a sample is collected to perform the study. Systematic error occurs when estimation methods introduce bias, statistical assumptions for the methods are not met, improper sampling techniques were used to collect the data for the linking study, or different placement of items impacts scale scores. To the extent possible, MetaMetrics and Classic Learning Initiatives worked to minimize systematic error through the design of the linking study.

Sample representativeness. To the extent that the targeted population is sufficiently represented by the study sample, the generalizability of the study will extend to future examinees. Classic

Learning Initiatives selected the sample for the study and it is assumed that this sample is representative of the CLT target population during the study window. A substantial difference between the sample and the target population (e.g., the target population has a much broader range of reading abilities) may result in inappropriate estimates of the linking functions' parameters. By extension, should the nature of the targeted population change, then the study may need to be reexamined.

Summary

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:

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

This report shows how a link has been established between the CLT Verbal Reasoning thetas and scale scores and the Lexile reading measures, permitting students to be matched with books and texts that provide an appropriate level of challenge while avoiding frustration. Students can be matched with texts that they are forecasted to read with 75% comprehension. It is anticipated that as a result of this purposeful match, students will read more, and thereby, read better. Wherever the student may be in the development of his or her reading skills, the Lexile Framework can be used to examine their growth. As a student grows, they can be matched with more demanding texts, thus facilitating additional growth.

To utilize the results from this study, Lexile reading measures need to be incorporated into the CLT Verbal Reasoning results processing and interpretation frameworks. This information can then be used in a variety of areas within the educational system—instruction, assessment, and communication, to name a few.

Now that a linkage is established between the CLT Verbal Reasoning scale and the Lexile scale, educators are able to utilize the assessment results, reported in Lexile measures, to inform classroom instruction.

References

Classic Learning Initiatives, LLC. (2022). *CLT3-6 Assessment Framework Brief for Grades 3-6*. Classic Learning Initiatives, LLC.

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, *37*, 281–306.

Dorans, N. J., Moses, T. M., & Eignor, D. R. (2010). Principles and practices or test score equating. *ETS RR-10-29*. Princeton, NJ: ETS.

Gardner, T. (2021). *2021 TECHNICAL REPORT The Classic Learning Test SEVENTH AND EIGHTH GRADE*. Classic Learning Initiatives, LLC.

Kolen, M.J. & Brennan, R.L. (2014). *Test equating, scaling, and linking: Methods and practices*. 3rd edition. New York: Springer Science + Business Media, LLC.

Linacre, J.M. (2022). WINSTEPS (Version 5.2.2.0) [Computer Program]. Chicago: 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.

Tyler, N. (2018). *2018 TECHNICAL REPORT The Classic Learning Test*. Classic Learning Initiatives, LLC.

[i] CLT thetas are horizontally scaled for each level.

[ii] The normative information for the Lexile Framework is based on linking studies conducted with the Lexile Framework and the results of assessments that report directly in the Lexile metric (N = 3,888,110). The sample included students in Kindergarten through Grade 12 from 51 states, districts, or territories who were tested from 2010 to 2016 (Grades 1–12) and 2016 to 2019 (Kindergarten). Of the students with gender information (45.1%), 51.6% of the students were male and 48.4% of the students were female. Of the students with race or ethnicity information (30.2%), the majority of the students in the norming sample were White (56.3%), with 5.8% African American, 2% American Indian/Alaskan Native, 14.7% Hispanic, 16% Asian, and 5.2% Other. Information on limited English proficiency (LEP) status was available for 2.9% of students, with 7% of the students classified as LEP. Special needs status was available for 2.8% of students, with 9.1% of the students classified as "Needing Special Education Services." Free and reduced-price lunch status was available for 2.9% of students, with 45.9% of the students eligible for free and reduced-priced lunch. The 2020 Lexile norms have been validated in relation to a longitudinal sample of students across Grades 3 through 11 (N = 101,610).

Appendix A:
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 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 for Reading focuses primarily on two features: reader ability and reading text complexity.

The Lexile Framework measures for both texts and readers typically range from 200L to 1600L. When matching readers with texts, all Lexile reading measures below 0L should be reported as "BRxxxL." Lexile text measures can be below 0L for beginning reader materials (e.g., BR150L) to above 1600L 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.

Reading 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 to measure reading text complexity by examining the characteristics of word frequency and sentence length. In addition, when measuring early reader texts, the Lexile Framework utilizes characteristics found to be important to the complexity of early reader text such as word decodability, patterning, and repetition.

Variables that Impact the 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 & Gildea, 1987; Stenner et al., 1983). Klare (1963) hypothesized that the semantic component varied along a familiarity-to-rarity continuum. This concept was further developed by Carroll et al. (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 et al. (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 & 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 et al., 1971). For example, the word "accident" appears 176 times in the 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 et al. (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 & 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 Impact the 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 reading development. For all readers, making meaning of a text 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 et al., 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 et al., 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 et al., 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 et al., 2012) and has identified the importance of considering the impact of interaction between the features (Biber, 1988; Merlini Barbaresi, 2003).

In order to determine which text characteristics had the greatest impact on reading text complexity for early readers, MetaMetrics identified 22 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, 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 substudies 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 et al., 2015; Fitzgerald et al., 2016).

The Lexile Scale

In developing the Lexile scale, the Rasch model (Wright & 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 people (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.

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:

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

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 Framework for Reading Map, a graphic representation of the Lexile scale from 200L to 1500L+, provides a context for understanding reading comprehension (see Appendix C).

Calibration of Difficulty of Upper Level Texts

The research study on semantic units (Stenner et al., 1983) was extended to examine the relationship of word frequency and sentence length to reading comprehension. Stenner et al. (1987a) 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 & 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 eight other standardized tests. The resulting correlation between the observed logit difficulties and the theoretical calibrations across the nine 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.

Calibration of Difficulty of Early Reader Texts

To bring the observed difficulties (logit scores) of early reader texts from the two studies previously described (Fitzgerald et al., 2015; Fitzgerald et al., 2016) 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 reading measure was large, so 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 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 et al., 1989; Kolen & Brennan, 2014). The result of the linear linking function was that the early reader observed difficulties were transformed to Lexile text measures while still maintaining the relative ordering of the difficulty of the texts derived from educator judgments and student performances.

Once observed Lexile reading measures were calculated, a random forest regression technique was employed to evaluate the importance of the 238 operationalizations of characteristics that research suggests affect reading 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 (2015). The first step in the analysis was to set baseline performance. Eighty percent of the texts were selected for this training process and 20% 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 recalculated. This process determined whether fewer variables could predict reading text complexity as well or nearly as well as the 238-variable model. The result was a set of nine variables:

- Word-level variables—monosyllable decoding, syllable count, age of acquisition, word rareness, and abstractness
- Within-sentence and across-sentence/discourse level variables—intersentential complexity, phrase diversity, non-compressibility, and text density

Lastly, 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 root mean square error of 9.68. The final model can now be used to calibrate texts intended for early readers.

The nine variables have been grouped into four Early Reading Indicators based on the linguistic level addressed:

- Decoding Demand (Decoding)—syllable count and monosyllable decoding demand
- Semantic Demand (Vocabulary)—abstractness, word rareness, and age of acquisition
- Syntactic Demand (Sentences)—intersentential complexity
- Structure Demand (Patterns)—non-compressibility, phrase diversity, and text density

The Lexile Text Analyzer®

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

Reporting Lexile Reading Measures

Lexile reading 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 reading measures that are being used for instructional purposes should be rounded to the nearest 5L to avoid overinterpretation of the measures. As with any test score, uncertainty in the form of measurement error is present.

Lexile reading 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 reading norms) to ensure developmental appropriateness of the material. MetaMetrics expresses these as "Reported Lexile Reading Measures" and recommends that these measures be reported on individual score reports. The Grade/Level Caps used for reporting Grades K–12 Lexile reading measures are shown in *Table A. 1.*

In instructional environments where the purpose of the Lexile reading measure is to appropriately match readers with texts, all scores below 0L should be reported as "BRxxxL." No student should receive a negative Lexile reading measure on a score report. The lowest reported value below 0L is BR400L.

Some assessments report a Lexile reading range for each student, which is 50L above and 100L below the student's actual Lexile reading 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.

Table A. 1. Maximum reported Lexile reading measures, by grade.

Grade/Level	Lexile Cap
Kindergarten	850L
Grade 1	900L
Grade 2	1100L
Grade 3	1200L
Grade 4	1300L
Grade 5	1400L
Grade 6	1500L
Grade 7	1600L
Grade 8	1700L
Grade 9	1725L
Grade 10	1750L
Grade 11	1800L
Grade 12	1825L

Validity Evidence for the Lexile Framework

The 2014 *Standards for Educational and Psychological Testing* (America Educational Research Association, American Psychological Association, and National Council on Measurement in Education) state that "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). In applying this definition to the Lexile Framework, the question that should be asked is "What evidence supports the use of the Lexile Framework to describe reading 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 can be evaluated by examining how well Lexile reading measures relate to other measures of reading ability and reading comprehension.

Relationship of Lexile Reading Measures to 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 et al., 1989, p. 222). Linking the Lexile Framework with other measures of reading comprehension produces a common frame of reference: the Lexile reading measure.

Table A. 2 presents the results from linking studies conducted with the Lexile Framework. In these studies, students were administered a Lexile reading 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 reading 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 individual norm-referenced test information, the student can receive a reading list consisting of texts (books and articles) targeted to his or her specific reading level.

Table A. 2. Results from linking studies conducted with The Lexile Framework.

Standardized Test	Grades in Study	*N*	Correlation Between Test Score and Lexile Measure
ACT Aspire	3, 5, 7, and EHS	1,264	0.85
PreACT	10	376	0.80
ACT	11 – 12	297	0.79
Arizona's Instrument to Measure Standards (AIMS)	3, 5, 7, and 10	5,599	0.89
ERB Comprehensive Testing Program (CPT4)	2, 4, 6, and 8	644	0.88
Gates-MacGinitie Reading Tests	2, 4, 6, 8, and 10	4,644	0.90
Georgia Milestones EOG/EOC Assessments	3 – 9, and AME	12,415	0.82 to 0.86*
ISIP Early Reading assessment	1 – 3	5,471	0.87
Advanced Reading assessment	4, 6, and 8	6,479	0.65
Kentucky Performance Rating for Educational Progress (K-PREP)	3 – 8	6,480	0.71 to 0.79*
Metropolitan Achievement Test (8th ed.)	2, 4, 6, and 8	2,713	0.92
North Carolina ACT	11	2,675	0.84
North Carolina READY End-of-Grade/End-of-Course Tests (NC READY EOG/EOC)	3, 5, 7, and 8	7,709	0.92
	English II	2,068	0.89
Oklahoma Core Competency Tests (OCCT)	3 – 8	8,437	0.81 to 0.86*
Oregon Reading/Literature Knowledge and Skills Test	3, 5, 8, and 10	3,180	0.87
Proficiency Assessment for Wyoming Students (PAWS)	3, 5, and 8	2,293	0.91
	11	442	0.84
South Carolina READY Reading	3 – 8	10,951	0.94
Stanford Achievement Test Series (Tenth Edition)	2, 4, 6, 8, and 10	3,064	0.93
State of Texas Assessments of Academic Readiness (STAAR™)	3 – 8	5,856	0.86
	English I	620	0.87
	English II	1,063	0.87
The Iowa Assessments (formerly Iowa Test of Basic Skills and Iowa Test of Educational Development)	3, 5, 7, 9, and 11	4,146	0.91
TOEFL iBT	NA	2,867	0.65
TOEIC	NA	2,770	0.74
West Virginia SAT School Day (Reading)	11	4,637	0.79

*Notes: * Tests were not vertically scaled; separate linking equations were derived for each grade/course.*

The 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 reading 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 reading measures. In a study conducted by Stenner et al. (1987b) Lexile reading 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 reading calibration of each unit in each series. After correction for range restriction and measurement error, the average disattenuated correlation between the Lexile reading calibration of text comprehensibility and the rank order of the basal units was 0.995 (see *Table A. 3*).

Table A. 3. 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	0.93	0.98	1.00
Harcourt Brace Jovanovich Eagle Series (1983)	70	0.93	0.98	1.00
Scott Foresman Focus Series (1985)	92	0.84	0.99	1.00
Riverside Reading Program (1986)	67	0.87	0.97	1.00
Houghton Mifflin Reading Series (1983)	33	0.88	0.96	0.99
Economy Reading Series (1986)	67	0.86	0.96	0.99
Scott Foresman: An American Tradition (1987)	88	0.85	0.97	0.99
HBJ Odyssey Program (1986)	38	0.79	0.97	0.99
Holt Basic Reading Series (1986)	54	0.87	0.96	0.98
Houghton-Mifflin Reading Series (1986)	46	0.81	0.95	0.98
Open Court Headway Program (1985)	52	0.54	0.94	0.97
Total/Means*	660	0.839	0.965	0.995

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 A. 3*, the Lexile reading 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.

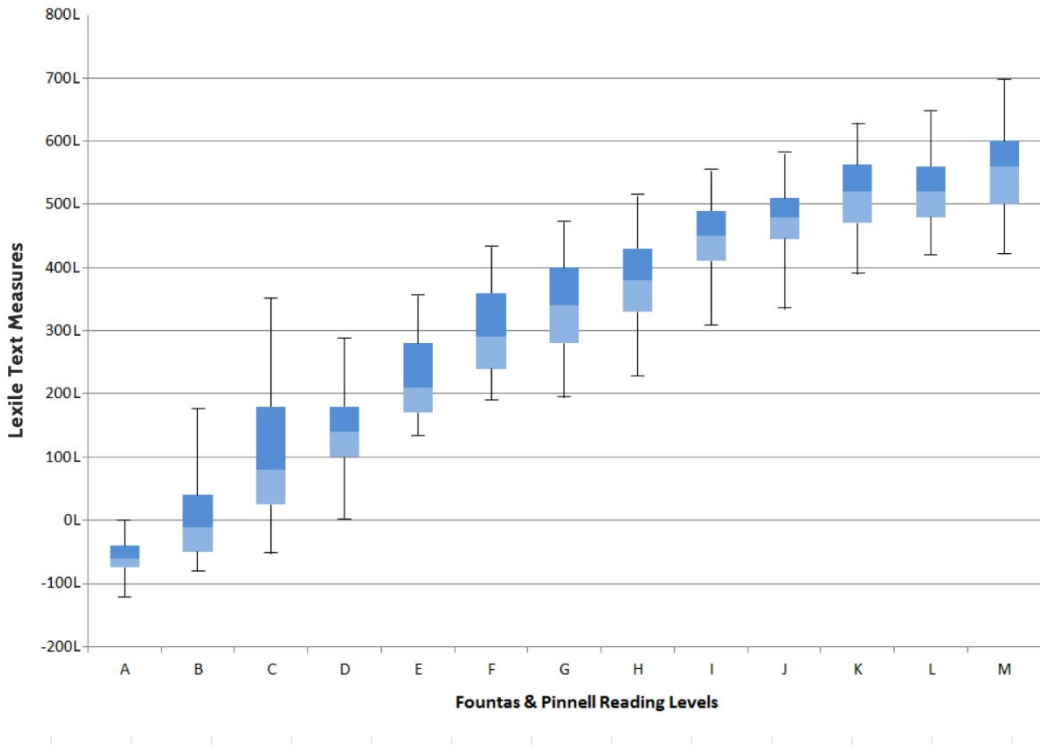
The Lexile Framework and Fountas & Pinnell Reading Levels

Koons et al. (2017) explored the relationship between Fountas & Pinnell reading levels for a set of texts A through M (i.e., Kindergarten through Grade 2) and their corresponding Lexile reading measures to obtain construct validity evidence for the measurement of early reader texts. The Spearman correlation coefficient between the two text sets was 0.84, indicating a strong positive relationship. Because Fountas & Pinnell reading levels are "larger grained" than the Lexile reading measures, some variation of Lexile reading measures within each Fountas & Pinnell reading level was expected.

Figure A. 1 shows a series of box and whisker plots of the results. The box in each box and whisker plot depicts the interquartile range (IQR) with the bottom of the box at the 25th percentile of the distribution of Lexile reading 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 A. 1 shows steadily increasing Lexile text reading 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 A. 1. Progression of Lexile text measures and Fountas & Pinnell reading levels, A through M.



The Lexile Framework and the Difficulty of Reading Test Items

Additional construct validity evidence was obtained by exploring the relationship between Lexile reading calibrations of item difficulties and actual item difficulties of reading comprehension tests. In a study conducted by Stenner et al. (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 reading 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, & 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 A. 4* contains the results of this analysis.

Table A. 4. 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	0.95	0.97	1.00
CAT-E	418	74	789	258	1339	212	1551	0.91	0.95	0.98
Lexile	262	262	771	463	1910	−304	1606	0.93	0.95	0.97
PIAT	66	66	939	451	1515	242	1757	0.93	0.94	0.97
CAT-C	253	43	744	238	810	314	1124	0.83	0.93	0.96
CTBS	246	50	703	271	1133	173	1306	0.74	0.92	0.95
NAEP	189	70	833	263	1162	169	1331	0.65	0.92	0.94
Battery	26	26	491	560	2186	−702	1484	0.88	0.84	0.87
Mastery	85	85	593	488	2135	−586	1549	0.74	0.75	0.77
Total/ Mean	1780	722	767	343	1441	50	1491	0.84	0.91	0.93

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 A. 4* show the raw correlation between observed (*O*) item difficulties and theoretical (*T*) item calibrations, with the correlations corrected for restriction in range and measurement error. The Fisher *Z* mean of the raw correlations (*rOT*) is 0.84. When corrections are made for range restriction and measurement error, the Fisher *Z* mean disattenuated correlation between theory-based calibration and empirical difficulty in an unrestricted group of reading comprehension items (*R′OT*) is 0.93. These results show that most attempts to measure reading comprehension (no matter what the item form used, type of skills or objectives assessed, or item type used) measure a common comprehension factor specified by the Lexile reading theory.

Text Measure Error Associated with the Lexile Framework

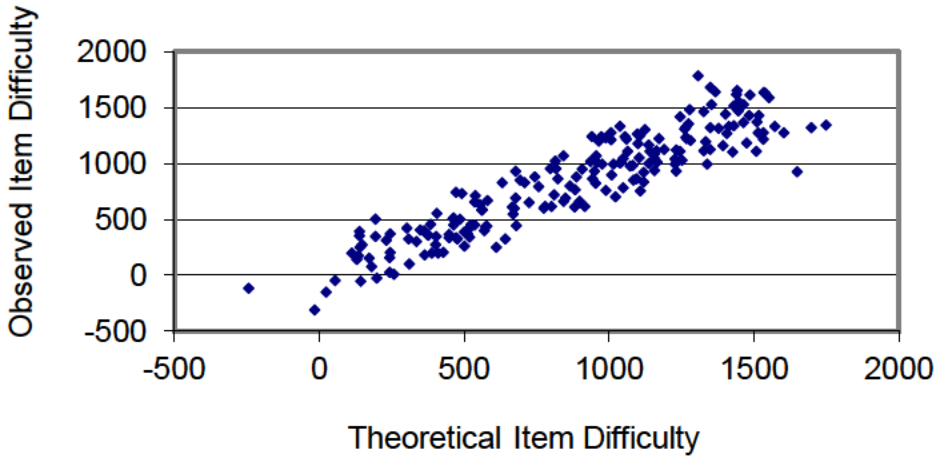
To determine a Lexile reading measure for a text, the standard procedure is to process the entire text. All pages in the work are concatenated into an electronic file that is processed by the Lexile Reading Analyzer software (developed by MetaMetrics, Inc.). The analyzer slices the text file into as 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 reading 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 Reading 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 Framework for Reading Map. The Lexile reading 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 A. 2*.

Figure A. 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 A.* 5).

Table A. 5. Standard errors for selected values of the length of texts.

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 on First Philosophy	206	1720L	12
The Metaphysics of Morals	209	1620L	12
The Adventures of Pinocchio	294	780L	10
The Red Badge of Courage	348	900L	10
The Scarlet Letter	597	1420L	7
Pride and Prejudice	904	1100L	6
The Decameron	2431	1500L	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 the 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 tool (hub.lexile.com/find-a-book/search) contains information about each book analyzed: author, Lexile reading measure, awards, ISBN, and developmental level/age range as determined by the publisher. For some books, Find a Book also provides Lexile text measures by chapter along with selected vocabulary words.

Lexile Item Bank

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

Passage selection. The passages used for item development are excerpted from authentic text, authored by MetaMetrics' staff, or commissioned by MetaMetrics' staff. Excerpted authentic text passages are selected from real-world reading materials that students 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. Passages authored or commissioned by MetaMetrics staff are created to model real-world reading materials.

The following criteria are used to select passages from authentic and authored passages:

- The passage consists of one main idea or contains one complete piece of information.

- Understanding the passage is independent of the information that comes before or after the passage in the source text.
- Understanding the passage is independent of prior knowledge not contained in the passage.

When writing items based on published text, item writers examine blocks of text that have Lexile reading measures within 100L of the source text (source targeting). Item writers select four to five source-targeted text blocks for potential item development. If it is necessary to shorten or lengthen a 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 to within 100L of the complete text. Items are then developed in conjunction with their associated passages.

When writing original passages, MetaMetrics staff who are experienced in item development and have experience with the everyday reading ability of students at various levels write original content calibrated to specific Lexile reading zones. Please see "Item Writer Training" in the next section for a detailed description of MetaMetrics' item development process.

Item format. The native Lexile reading 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 125 words in length. The passages are shorter for early 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 options presented, the reader is asked to select the best option to complete 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 based on information in the passage: paraphrasing, making an inference, or making a generalization. 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.

Items used to assess the reading ability of early readers include picture items, picture/word audio enhanced items, one-sentence items, and two-sentence items. These items are designed using Lexile appropriate vocabulary, sight words, images, and other text characteristics typically associated with early reading. More information on foundational reading items is provided in the next section.

The components of the Lexile Item Bank reading comprehension items and their descriptions are included below.

> *Passage*—the ancillary text for which an item is written. For most items, the Lexile reading measure of the passage is considered the Lexile reading measure of the item.

Each passage is used for only one item. For picture items, an image is in place of the passage. For one-sentence items, the passage consists of the stem only. And for two-sentence items, one sentence acts as the passage.

Stem—the question or embedded completion statement. For embedded completion statements, they should appear as if they were written as part of the passage. The statement portion of the embedded completion item can assess a variety of skills related to reading comprehension: paraphrasing information in the passage, making an inference based on the information in the passage, identifying a supporting detail, or making a generalization based on the information in the passage. The statement is written to ensure that by reading and comprehending the passage the reader is able to select the correct option.

Correct answer—the correct response. The correct answer (key) typically has a Lexile reading measure similar to the measure of the passage.

Distractors—the three wrong responses that are semantically and syntactically correct. These should be attractive responses if the reader has not read the passage. The distractors have similar Lexile reading measures as the correct answer.

Foundational reading items. Early in their pathway to reading, students develop foundational reading skills which are associated with improved reading outcomes in later stages of reading development and ultimately reading comprehension (National Governors Association & CCSSO, 2010; National Institute of Child Health and Human Development, 2000). To support teachers with evaluating the foundational reading skills of students during their early literacy development and inform instruction, appropriate assessment items are needed. In 2019 and 2021, MetaMetrics conducted research to expand the Lexile Item Bank to include items on the Lexile scale that measure Kindergarten and Grade 1 foundational reading skills (Webb et al., 2022). This research led to the development of a foundational reading framework consisting of three primary domains — Print and Alphabet Knowledge, Phonological Awareness, and Phonics. Each domain is further divided into two or more subdomains (see *Table A. 6*).

Table A. 6. Foundational reading domains and subdomains, by grade.

Domain	Subdomain	Grade	
		K	G1
Print and Alphabet Knowledge	Concepts of Print	x	
	Alphabetic Awareness	x	
	Letter Sequence	x	
Phonological Awareness	Words in a Sentence	x	
	Rhyme		x
	Syllables	x	
	Onsets and Rimes	x	
	Phoneme Isolation	x	x
	Phoneme Blending	x	x
	Phoneme Segmenting		x
	Phoneme Manipulation		x
Phonics	Letter Sound Correspondence	x	
	Consonant Sounds	x	x
	Word Families		x
	Consonant Blends and Digraphs		x
	Vowel Sounds	x	x

Table A. 6. MetaMetrics conducted two rounds of item development (Summer 2019 and Summer 2021). A total of 270 items were developed which were reviewed by subject matter experts, teachers, and test development researchers. The items were field-tested in Fall 2019 and Fall 2021. The participants in the field-test studies included a total of 3,859 students in Pre-K (n = 626), Kindergarten (n = 1,914) and Grade 1 (n = 1,319) across 36 U.S. states representative of all geographical regions. The students were from 247 classrooms in 166 different schools. Analysis of the resulting data placed each item on the Lexile scale.

Item writer training. Item writers are professional writers, classroom teachers, and other educators who have had experience with the everyday reading ability of students at various levels. Experienced item writers help to ensure that all Lexile Item Bank reading comprehension items are valid measures of reading comprehension. New item writers practice item writing and reviewing over one to two months so that senior curriculum specialists can provide them with specific and individualized feedback to ensure proper training. 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 training materials also contain examples of poorly constructed items to illustrate the criteria used to evaluate items and corrections based on those criteria. Item writers are also provided vocabulary lists to use during statement and option development. The vocabulary lists were assembled from word lists compiled by MetaMetrics based on vocabulary research related to determining the Lexile reading measures (difficulty) of words (MetaMetrics, 2006). 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 given extensive training related to sensitivity issues. Item-writing training materials provide examples and identify areas to avoid when selecting or writing passages and

developing items. The following areas are covered: violence and crime, sources of common phobias, negative emotions surrounding death and family issues, offensive language, drugs/alcohol/tobacco, sex/attraction, race/ethnicity, class, gender, religion, supernatural/magic, parent/family, politics, animal cruelty and hunting, environmental issues, brand names, and junk food. These materials were developed based on material published by McGraw Hill (McGraw-Hill Book Company, 1983) related to universal design and fair access—the equal treatment of the sexes, the fair representation of minority groups, and the fair representation of disabled individuals.

Item review. All items are subjected to a multistage review process. First, items are read and edited by item writers and reviewers according to the 25 criteria identified in the item writing materials as well as 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 reviewed and edited by a group of specialists that represent various perspectives—curriculum specialists, content editors, fact-checkers, sensitivity reviewers, and test developers. These individuals examine each item for sensitivity issues, grammar and spelling, and item quality (stem, key, and distractors).

During the second stage of the item review process, items are either "approved as presented," "approved with edits," or "rejected." Approximately 90% of the items written are "approved as presented" and 10% are "approved with edits" or "rejected" at this stage. When necessary, item writers receive additional 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 reading measure] and theoretical Lexile reading measure), internal consistency (point-biserial or point-measure correlation), and bias (ethnicity and gender where possible). Where necessary, items are deleted from the bank or revised and recalibrated.

In addition to content and sensitivity reviews during the development process, Lexile Item Bank items are field-tested as part of MetaMetrics' ongoing research. These items may be field-tested as part of stand-alone research field tests or they may be embedded within research tests for concurrent projects. During Spring 1999, eight levels of a Lexile reading assessment were administered in a large urban school district to students in Grades 1 through 12. The eight 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 eight 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 reading measures by multiplying by 180 and anchoring each set of items to the mean theoretical difficulty of the items on the form.

MetaMetrics continues to add new items to its item bank and regularly evaluates items for potential use on linking studies. Each time items are administered, their empirical data are evaluated to determine whether they should be removed from the item bank, revised and retested,

or kept for future use on tests developed for MetaMetrics' partners, linking studies, and research studies.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Biber, D. (1988). *Variation across speech and writing.* Cambridge, England: Cambridge University Press.

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.

Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, *14*, 135-163.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychology*, *93*, 3-22.

Fitzgerald, J., Elmore J., Hiebert, E.H., Koons, H., Bowen, K., Sanford-Moore, E.E., & Stenner A.J. (2016). Examining text complexity in the early grades. *Phi Delta Kappan*, *97*, 60-65.

Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E. H., Bowen, K., Sanford-Moore, E. E., & Stenner, A.J. (2015). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, *107*, 4-29.

Graesser, A. C., & McNamara, D. S. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223-234.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*, 223-234.

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.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

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*. 3rd edition. New York: Springer Science + Business Media, LLC.

Koons, H., Elmore, J., Sanford-Moore, E., Stenner, A.J. (2017). The relationship between Lexile text measures and early grades Fountas & Pinnell reading levels. MetaMetrics: Durham, NC.

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.

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

Merlini Barbaresi, L. M. (2003). Towards a theory of text complexity. In L. Merlini Barbaresi (Ed.), *Complexity in language and text* (pp. 22-66). Pisa, Italy: Edizioni Plus.

Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, *47*, 235-258.

MetaMetrics, Inc. (2006). *Lexile Vocabulary Analyzer. Technical report*. Durham, NC: Author.

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

Muter, V., Hulme, C., Snowling, M. J., Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*, 665-681.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects (Appendix A).* Washington DC: Author.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, CD: U.S. Government Printing Office.

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.

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.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica: RAND Corporation.

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

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.

Vadasy, P. F., Sanders, E. A., Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities, 38*, 364-382.

Webb, X.J., Steinkamp, S., Koons, H., Saha, N., Sanford-Moore, E.E., Baker, R., & Hinson, A. (2022). Foundational reading skills: Item development and validation [Unpublished manuscript]. Durham, NC: MetaMetrics.

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 B:
Recommendations for Using the Lexile Framework for Reading

Recommendations about reporting Lexile reading 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 that are being used for instructional purposes 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 reading norms) to ensure developmental appropriateness of the material. MetaMetrics expresses these as "Reported Lexile Reading Measures" and recommends that these measures be reported on individual score reports. The grade level caps used for reporting Grades K–12 Lexile reading measures are shown in *Table B. 1*.

In instructional environments where the purpose of the Lexile reading measure is to appropriately match readers with texts, all scores below 0L should be reported as "BRxxxL." No student should receive a negative Lexile reading measure on a score report. The lowest reported value below 0L is BR400L.

Some assessments report a Lexile reading range for each student, which is 50L above and 100L below the student's actual Lexile reading 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.

Table B. 1. Maximum reported Lexile reading measures, by grade.

Grade	Lexile Cap
K	850L
1	900L
2	1100L
3	1200L
4	1300L
5	1400L
6	1500L
7	1600L
8	1700L
9	1725L
10	1750L
11	1800L
12	1825L

Use The Lexile Framework for Reading to Select Books. Teachers can use the tools provided by the Lexile® Framework for Reading to select materials to develop individualized reading lists that are tailored to individual students. 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. A student's Lexile reading measure can be leveraged to aid comprehension and is a good starting point in the selection process of a book for a specific reader.

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 & Knickelbine, 2000).

Another feature of the Lexile Framework is that it makes provisions for students who read below or beyond their grade level, because the reporting scale is not bounded by grade level. See The Lexile Framework for Reading Map for literary and informational titles, leveled reading samples, and approximate grade ranges. (Appendix C)

However, it is important to note that the Lexile reading measure should never be the only piece of information used when selecting a text for a reader. When matching a book with a reader, one must also consider other factors that may affect the relationship between a reader and a book. These factors include student developmental level, motivation, and interest; amount of background knowledge possessed by the reader; and suitability of the text and text difficulty. For

example, 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 reading range. If the student is less motivated or intimidated by a reading task, material at the lower end of his or her Lexile reading range can provide the basic comprehension support to keep the student from feeling overwhelmed.

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.

Target 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.

Similarly, teachers can use Lexile reading measures to guide a struggling student by selecting texts at the lower end of the student's Lexile reading range (e.g., 50L below his or her Lexile reading measure). At the same time, teachers can also motivate advanced students by challenging them with reading texts at the midpoint of their Lexile reading range or slightly above (i.e., 25L above to 100L above his or her Lexile reading measure).

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.

Apply Lexile Reading Measures Across the Curriculum. Over 600 publishers provide Lexile reading 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 reading 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 reading measures.

Use the Lexile Framework to facilitate communicating with stakeholders. Lexile reading 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 an appropriate level of challenge and monitor the reading process at home. The "Lexile Find A Book" website also 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. Lexile reading measures 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 https://hub.lexile.com/find-a-book/search.

In one large school district, the end-of-year testing results are sent home to parents in a folder. The folder consists of the Lexile Framework for Reading 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 Framework for Reading 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, stakeholders can understand what the community values and more readily 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 reading 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.

There are numerous ways to incorporate the Lexile Framework including:

- Building 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.
- Sequencing 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).
- Developing 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 reading 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.
- Selecting texts lower in the student's Lexile reading 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.
- Enhancing a student's experience with exposure to differentiated, challenging text at least once every two to three weeks.
- Leveraging the free Find a Book website (at https://hub.lexile.com/find-a-book/search) to support book selection and create booklists within a student's Lexile reading range to help the student make more informed choices when selecting texts.
- Utilizing database resources to infuse research into the curricula while tailoring reading selections to specific Lexile reading 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 https://metametricsinc.com/products/library-products/.
- Using Lexile® WordLists (https://hub.lexile.com/wordlists) to identify subsets of words that are relevant to the context or application. Lexile WordLists contain approximately 50,000 unique words from the top four best-selling textbook programs (published after 2011) in science, math, social studies, and reading/English language arts. Some common uses include: identifying grade appropriate words to target vocabulary instruction and

assessment; identifying words to include in instructional materials for domain-specific content; and selecting important academic words by grade and domain to highlight in reading passages, books or other instructional materials.

Use the Lexile Framework in the Library. Augmenting libraries provides even more ways to leverage the Lexile Framework including:

- Making the Lexile reading measures of books available to students to better enable them to find books of interest at their appropriate reading level.
- Enabling comparison of student Lexile reading levels with the Lexile reading levels of the books and periodicals in the library to analyze and develop the collection to more fully meet the needs of all students.
- Leveraging the database resources to search for articles at specific Lexile reading 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 https://metametricsinc.com/products/library-products/)
- Using the free Find a Book website (at https://hub.lexile.com/find-a-book/search) to support book selection and help students make informed choices when selecting texts.

Set and Monitor Reading Program Goals. 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 with improving their reading skills can use the Lexile Framework for evaluation purposes. Schools can use student-level and school-level Lexile reading information to monitor and evaluate interventions designed to improve reading skills. Progress tests throughout the year can be conducted to help monitor students' progress toward their goals.

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

Example Goal 1: At least half of the students will improve reading comprehension abilities by 100L after one year of use of an intervention.

*Example Goal 2: S*tudents' attitudes about reading will improve after reading 10 books at their 75% comprehension level.

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

References

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

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

Appendix C:
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.

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.

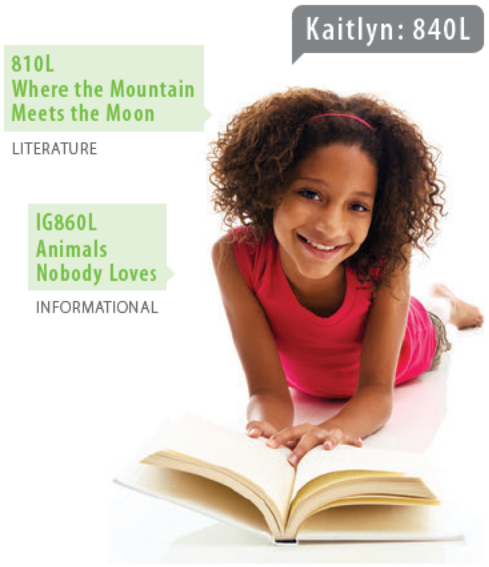
HOW TO USE IT

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.



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)

THE LEXILE® FRAMEWORK FOR READING

400L–695L LEXILE RANGE

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)

THE LEXILE® FRAMEWORK FOR READING

200L–395L+ LEXILE RANGE

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

Appendix D:
The Lexile Framework for Reading and Forecasted Comprehension Rates

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

The 75% comprehension rate for an examinee-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. An individual who answers three-fourths of the questions correctly has a 75% comprehension rate.

Suppose instead that the text and the examinee measures are not the same. It is the difference in Lexile reading measures between the examinee and text that governs comprehension. If the text measure is less than the examinee 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 individual reads a 350L text?

If all the item-sized slices in the 350L text had the same calibration, the 250L difference between the 600L examinee and the 350L text could be determined using the Rasch model equation. This equation describes the relationship between the measure of an examinee'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 D.1 is an unbiased approximation:

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

where ELD is the "effective logit difference" given by

$$\text{ELD} = (\text{Examinee Lexile measure} - \text{Text Lexile measure}) \div 225. \qquad \text{Equation (D.2)}$$

Figure D. 1 shows the general relationship between examinee-text discrepancy and forecasted comprehension rate. When the examinee 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 examinee measure of 600L and the text calibration of 350L is 250L. Referring to *Figure D. 1* and using +250L (examinee minus text), the forecasted comprehension rate for this examinee-text combination would be 90 percent.

Figure D. 1. Relationship between examinee-text discrepancy and forecasted comprehension

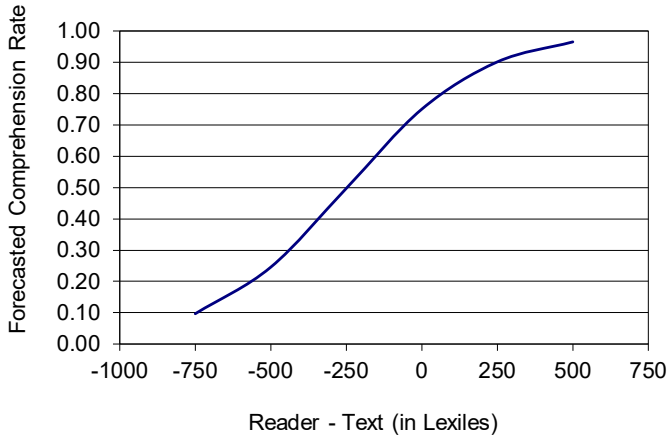


Table D. 1 and *Table D. 2* show comprehension rates calculated for various combinations of examinee measures and text calibrations.

Table 1. Comprehension rates for the same individual with materials of varying comprehension difficulty.

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

Table 2. Comprehension rates of different examinee abilities with the same material.

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

The primary utility of the Lexile® Framework for Reading is its ability to forecast what happens when examinees confront text. With every application by teacher, examinee, or librarian there is a test of the framework's accuracy. The Lexile Framework makes a point prediction every time a text is chosen for an individual. 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, examinee measures, and their difference modeled as forecasted comprehension. However, the error is sufficiently small that the judgments about examinees, texts, and comprehension rates are useful.

Examinee Forecasted Comprehension Rate. Using Equation D.2 with different combinations of examinee measure and text difficulty, a forecasted comprehension rate can be determined. *Table D. 3* shows the changes in the forecasted comprehension rate for different combinations of examinee and text interactions.

Table 3. Effect of examinee-text discrepancy on forecasted comprehension rate.

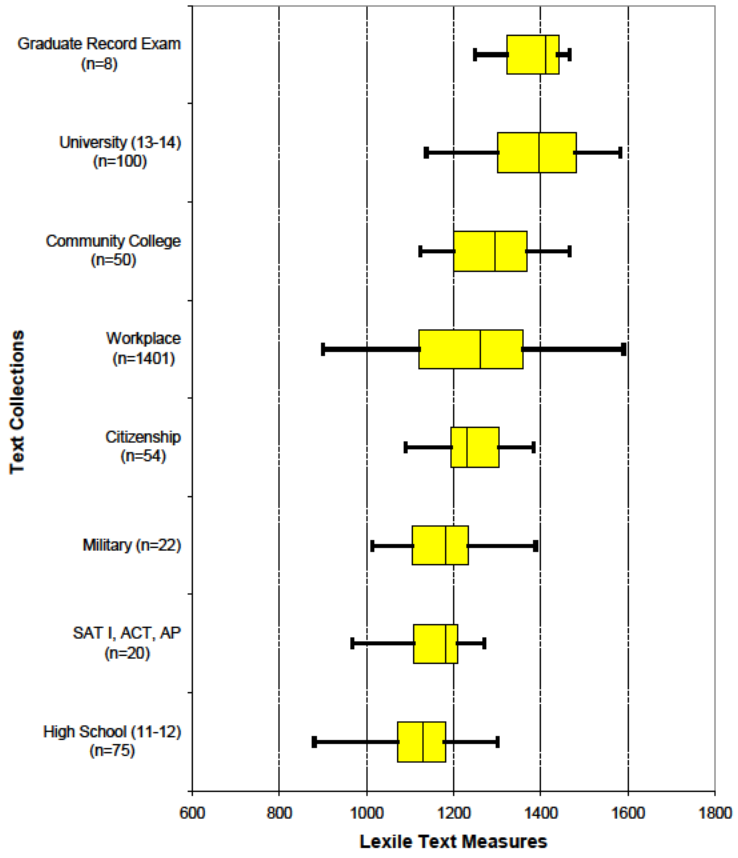
Examinee Lexile Reading Measure	Text Lexile Measure	Difference	Forecasted Comprehension Rate
1000L	970L	30L	77.4%
1000L	975L	25L	77.0%
1000L	980L	20L	76.7%
1000L	985L	15L	76.3%
1000L	990L	10L	75.8%
1000L	995L	5L	75.4%
1000L	1000L	0L	75.0%
1000L	1005L	⁻5L	74.6%
1000L	1010L	⁻10L	74.2%
1000L	1015L	⁻15L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

Appendix E:
College and Career Reading Demands

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.

Williamson (2008) investigated the gap between high school textbooks and various reading materials across several postsecondary domains. The resources Williamson used were organized into four domains that correspond to the three major postsecondary endeavors that students can choose—further education, the workplace, or the military—and the broad area of citizenship, which cuts across all postsecondary endeavors. Williamson discovered a substantial increase in reading expectations and reading 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). See *Figure E. 1.*

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

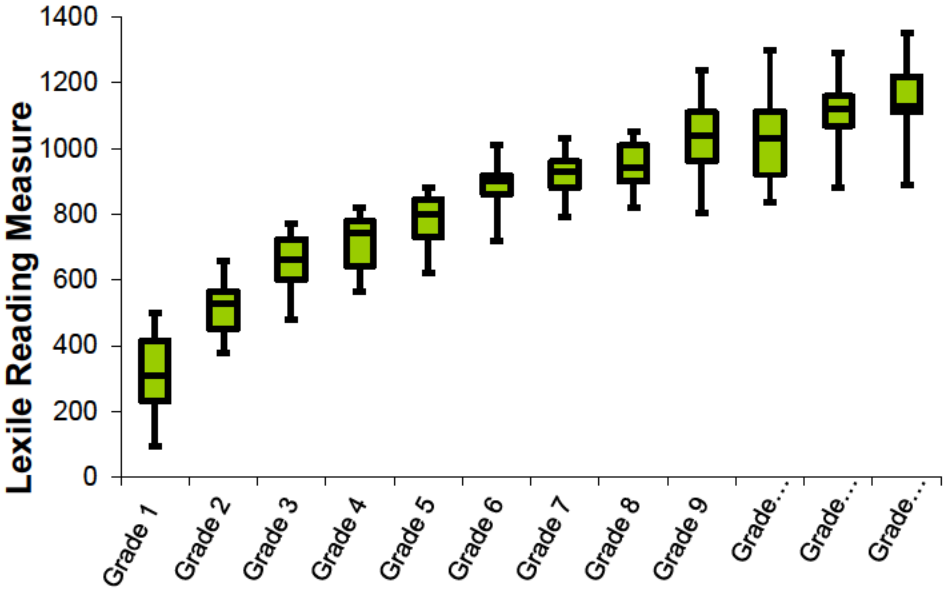


In Texas, two studies (MetaMetrics, 2007; MetaMetrics, 2008) were conducted to examine the reading demands in various postsecondary options—technical college, community college, and 4-year university programs. Under Commissioner Raymond Paredes, the Texas Higher Education Coordinating Board (THECB), in conjunction with MetaMetrics, 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 reading text complexity of these books was measured using the Lexile® Framework for Reading. Since the Texas Assessment of Knowledge and Skills (TAKS) had already been linked with Lexile reading measures for several years, the THECB study was able to overlay the TAKS cut scores onto the post high school reading requirements.

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 describe the reading demands individuals will likely encounter as they prepare for college and careers. 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 of postsecondary options and could be used to judge a student's "college-and-career readiness."

Between 2004 and 2008, MetaMetrics (Williamson et al., 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 E. 2*.

Figure E. 2. Reading text complexity distributions, in Lexile reading units, by grade (whiskers represent 5th and 95th percentiles).



This continuum can be "stretched" to describe the reading demands students will likely encounter in Grades 1–12 when "on track" for college and career (Sanford-Moore & 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 E. 1 provides the stretch text measure ranges for Grades 1 through 12. Combining student results with criterion referenced indicators provides information to reference when matching students with reading materials that are at or above the recommendations in Appendix A for each grade level.

Table 1. Lexile reading 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

References

MetaMetrics, Inc. (2007). *Texas Higher Education Coordinating Board: Text Measurement and Analysis. Technical report*. Durham, NC: Author.

MetaMetrics, Inc. (2008). *Texas Higher Education Coordinating Board: Text Measurement and Analysis (Update). Technical report*. Durham, NC: Author.

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

Smith, M. (2011). *Bending the reading growth trajectory: Instructional strategies to promote reading skills and close the readiness gap*. MetaMetrics Policy Brief. Durham, NC: MetaMetrics, Inc.

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

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

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