

Calculation of Lexile Word Measures Using a Corpus-Based Model and Student Performance Data

Jeff Elmore, Steve Lattanzio, A. Jackson Stenner, Ph.D. and Eleanor E. Sanford-Moore, Ph.D.

AUGUST 2016

OBJECTIVE

As part of a larger project related to modeling multiple aspects of vocabulary knowledge, word difficulty measures denominated in Lexile® units were calculated in a manner similar to the development of The Lexile® Framework for Reading (Stenner, Horabin, Smith, & Smith, 1988; Stenner, Burdick, Sanford, & Burdick, 2007). A Lexile word measure is an estimate of the challenge a particular word will present, on average, to a particular reader during independent reading. Lexile word measures were calculated using a corpus-based, machine-learning model that was developed using student performance data from several reading tasks. The model was then used to calculate Lexile word measures for tens of thousands of words for which student performance data was unavailable.

Two additional sets of analyses were performed to assess the validity and utility of Lexile word measures: 1) the measures were compared with existing measures of word difficulty including EDL Grade, PPVT rank-order difficulty, and age-of-acquisition; and 2) the percentage of running words and total unique words with Lexile word measures was calculated for a large sample of texts to assess the comprehensiveness of the dataset.

METHODS

Data Sources:

Student Performance Data

Three sources of student performance data were used to calculate empirical Lexile word measures: 1) measures derived from automatically generated, inline cloze performance data from EdSphere® (Lattanzio, Elmore, & Stenner, 2016); 2) measures derived from native-Lexile items (short passages with human-created, embedded completion statements and four responses) used in previous Lexile research studies ($n = 361$); and 3) measures derived from picture items (picture and four responses) for low-level words that were also used in previous Lexile research studies ($n = 129$).

A total of 35,452 empirically derived measures of word difficulty, denominated in Lexile units, were calculated based on large numbers of students answering reading comprehension items on the EdSphere platform (Lattanzio, et al., 2016). Lexile word measures, along with Lexile text measures accurately predict student performance on inline cloze items. Optimal measures were determined to occur under the following conditions: when a 500L reader encounters a 500L word in a 500L text, he or she is expected to answer correctly an item generated for that word 75% of the time. As word measures diverge from text measures, their impact on expected comprehension is about 50% the magnitude of changing both the text and word measures. So, a 700L word found in a 500L text results in an overall predicted item difficulty of approximately 600L (Lattanzio, et al., 2016).

Empirical Lexile word measures from the three data sources were combined into a single dataset. Measures from the EdSphere dataset with standard errors greater than 100L were removed ($n = 23,739$). No measures were removed from the other datasets. Because of the relative paucity of lower-Lexile items, in cases in which the same word appeared in more than one dataset and thus had more than one measure, the lowest measure was selected. In total, 11,836 empirically derived Lexile word measures were used in the present study. Table 1 shows descriptive statistics for each dataset.

Table 1: Descriptive statistics for empirically derived Lexile word difficulty measures.

	N	Mean	SD	Min	25%	50%	75%	Max
EdSphere	11,713	1023L	393L	-147L	731L	1005L	1299L	2368L
Lexile Native Items	361	848L	424L	-604L	581L	919L	1146L	1765L
Picture Items	129	-164L	201L	-819L	-294L	-162L	-32L	366L
Combined	11,836	1012L	410L	-819L	726L	1001L	1293L	2368L

Lexile Word Frequency Profiles

Lexile Word Frequency Profiles were used to develop a corpus-based theoretical model that can accurately predict empirically derived Lexile word measures and calculate Lexile word measures for words without an empirically derived measure. A Lexile Word Frequency Profile for a word includes a total of 624 word-frequency measures conditional on the text Lexile measure in which they occurred. These profiles provide a description of not just how often a word occurs, but a developmental progression of its use in a corpus of over 1.4-billion words drawn from books used in K-12 classrooms (Elmore, 2016). Lexile Word Frequency Profiles provide a greater source of information about likely exposure to

a word since they account for the developmental nature of word use. For example, so-called academic words such as *analyze* and *elaborate* tend to appear more frequently in higher Lexile texts whereas concrete nouns such as the names of animals and foods tend to occur in lower Lexile texts. Lower ability readers are less likely to have encountered words that appear primarily in higher Lexile texts.

Existing Word Difficulty Measures

Three datasets of word difficulty measures were collected for the purpose of comparison with and validation of the newly calculated Lexile word measures:

- EDL Grade Levels for 8,000 words (Taylor, Frackenpohl, & White, 1979),
- Peabody Picture Vocabulary Test (PPVT) rank order difficulty for 402 words (Dunn & Dunn, 2012), and
- Age-of-acquisition ratings for 51,000 words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012).

PROCEDURES & ANALYSES

To calculate Lexile word measures for words without student data, much as we estimate the text complexity of texts with the Lexile Analyzer®, a random forest regression model (Breiman, 2001) was developed using the 624 Lexile Word Frequency Profile measures to predict empirically derived Lexile word measures and calculate measures for words without empirical data. The performance of the model was evaluated using out-of-bag (OOB) estimates of R^2 . Function words such as *the*, *a*, *to*, *and*, *therefore* were marked with the code F instead of a Lexile measure, since these words tend to not carry specific meanings and are difficult to assess (Hartmann & Stork, 1972).

To reduce the number of strange or erroneous words receiving measures, three trimming steps were performed: 1) words that occurred fewer than 30 times in the MetaMetrics® corpus— corresponding with just one occurrence in 50,000,000 running words on average— were excluded; 2) words without a definition in WordNik (www.wordnik.com) were excluded; and 3) words with large standard errors from the random forest model were also excluded. Words with large standard errors were often highly polysemous or otherwise ambiguous words outside of their immediate context such as proper nouns. Next, inflected word-families (Elmore, Fitzgerald, Graves, & Bowen, 2015) were combined and each word in the family was assigned the lowest Lexile level in the family. For example, *play*, *plays*, *playing*, and *player* were all assigned a measure of 60L, however *playful* has a separate measure of 680L. Finally, measures were capped at 1700L due to limited empirical data from high-ability readers.

The final set of Lexile word measures was then compared to existing measures of word difficulty using Spearman ρ rank-order correlations. To assess the comprehensiveness of the final dataset, the average percentage of unique words (types) and running words (tokens) with Lexile measures were calculated for a random sample of texts spanning the full range of the Lexile scale. Descriptive statistics for coverage percentages were calculated. A brief qualitative analysis of the words without Lexile measures found in the sample of texts was performed.

RESULTS

A random forest regression model with 100 trees, mtry set to 25 (which is approximately the square root of the total number of variables), and unpruned regression trees (Breiman, 2001) was fit using the 11,836 words with empirical measures. The out-of-bag R^2 for the model was 0.69, suggesting accurate modeling of word difficulty with the Lexile Word Frequency Profile model.

Of the 261,794 unique words in the MetaMetrics corpus, 155,624 words (59.4%) occurred at least 30 times. Of these, 82,500 words (31.5%) were found to have definitions in WordNik. Of those, a further 6,678 words with high standard errors from the model were also removed, leaving 75,822 words (29.0%) with Lexile measures. Table 2 includes a sample of words across the range of the Lexile scale. Words with lower Lexile word measures include common place concrete nouns and familiar concepts; whereas, words with higher Lexile word measures include abstract, unfamiliar, and academic words. Table 3 shows correlations with other word difficulty measures. The high correlations suggest Lexile word measures are measuring the same underlying construct of word familiarity.

Table 2: A sample of theoretically derived Lexile word measures.

garden	BR	primary	900L
snake	100L	intimidate	1100L
dream	300L	abandonment	1300L
active	500L	gyre	1500L
access	700L	inexorably	1700L

Table 3: Relationship of theoretical Lexile word measures with other word difficulty measures.

	<i>N</i>	Spearman ρ
EDL Grade	8,641	0.85
PPVT Rank	390	0.83*
Age-of-acquisition	44,465	0.75*

*Significant at $p < 0.01$

Next, to assess the comprehensiveness of the dataset, a random sample of 1,000 books from the MetaMetrics corpus from 0L to 1600L with a mean Lexile measure of 730L was selected for the analysis of coverage of the Lexile word measures. Table 4 shows descriptive statistics for the percentage of running words (tokens) and unique words (types) with Lexile measures. Mean coverage was 94% for tokens and 95% of types.

Table 4: Percentage of running words (tokens) and unique words (types) in a sample of 1,000 texts from the MetaMetrics corpus.

	<i>N</i>	Mean	SD	Min	25%	50%	75%	Max
% coverage of tokens	1,000	94	3	74	92	94	96	100
% coverage of types	1,000	95	2	81	93	95	97	100

Finally, to determine what kinds of words account for the small percentage of words without theoretical Lexile word measures, a random sample of words without Lexile word measures from the texts was examined. The sample reveals mostly proper nouns, with some misspellings, or rare compound words such as *warehouseman* and *megaspore*. Table 5 shows a random sample of 30 words without Lexile measures from texts in the coverage analysis.

Table 5: random sample of words without Lexile measures found in random sample of books.

superiore	tsion	Rosies	citta	Kostoses	kindergartner
Volvos	whirlie	Schloss	Vuitton	toueiros	Belezzan
Jose	Panama	pombal	gustan	voiture	Anh
Tazawa	wifeman	drabbest	ain	ratingbone	bronya
woodhull	pukeful	temnodontosaurus	teatro	cataplasms	Ehrlich

CONCLUSION

This study represents an important first step in developing a theoretical measure of word difficulty that can connect word, text, and reader by placing them on the same scale. Lexile word measures derived from student reading performance and directly connected to the Lexile scale makes the word difficulty measures a useful measure for tasks such as identifying challenging words in a particular text or collection of texts.

Although Lexile word measures correlated highly with other measures of word difficulty, there are always improvements to be made. No other dataset was as comprehensive as the Lexile word measures dataset, so the extent of their relationship is unknown for some words. In particular, very rare words such as technical terms, compound words, and unusual derivations of other relatively common words should be more closely examined. In addition, theoretical models that take into account compounding and other forms of derivation than just inflections and plurals should be considered.

REFERENCES

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Dunn, L. M., & Dunn, D. M. (2012). Peabody Picture Vocabulary Test, (PPVT™-4). Johannesburg: Pearson Education Inc.
- Elmore, J., Fitzgerald, J., Graves, M., & Bowen, K. (2015). The Vocabulary of Elementary Disciplinary Texts. Durham, NC: MetaMetrics, Inc.
- Elmore, J. (2016). Lexile Word Frequency Profiles. Durham, NC: MetaMetrics, Inc.
- Hartmann, R. R. K., & Stork, F. C. (1972). *Dictionary of language and linguistics*. London: Applied Science Publishers.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978-990.
- Lattanzio, S., Elmore, J., & Stenner, A. J. (2016). Empirical Lexile Measures for Words. Durham, NC: MetaMetrics, Inc.
- Stenner, A.J., Horabin, I., Smith, D.R., & Smith, M. (1988). The Lexile Framework. Durham, NC: MetaMetrics.
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). The Lexile Framework for Reading Technical Report. Durham, NC: MetaMetrics, Inc.
- Taylor, S. E., Frackenpohl, H., & White, C. E. (1979). *EDL core vocabularies in reading, mathematics, science, and social studies*. New York: McGraw-Hill.

For more information, visit www.MetaMetricsInc.com.

MetaMetrics® is focused on improving education for students of all ages. The organization develops scientific measures of academic achievement and complementary technologies that link assessment results with instruction. For more than twenty years, MetaMetrics' work has been increasingly recognized worldwide for its distinct value in differentiating instruction and personalizing learning. Its products and services for reading, mathematics and writing provide valuable insights about academic ability and the potential for growth, enabling students to achieve their goals at every stage of development.

METAMETRICS®, the METAMETRICS® logo and tagline, LEXILE®, LEXILE® FRAMEWORK, LEXILE ANALYZER® and the LEXILE® logo are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad. The trademarks and names of other companies and products mentioned herein are the property of their respective owners. Copyright © 2016 MetaMetrics, Inc. All rights reserved.

