



The Vocabulary of Elementary Grades Disciplinary Textbooks

Jeff Elmore, Senior Research Engineer, Jill Fitzgerald, Ph.D., MetaMetrics Distinguished Research Scientist and The University of North Carolina at Chapel Hill Research Professor and Professor Emerita, Michael Graves, Ph.D., University of Minnesota Professor Emeritus, & Kimberly Bowen, Ph.D., Research Associate

October 2015

OBJECTIVE & RATIONALE

As part of a multi-phase research endeavor to identify the words that U.S. students are likely to encounter in grades K–12, the vocabulary contained in current, best-selling elementary grades (first through fifth grade) disciplinary textbook series (social studies, science, and mathematics) was analyzed. Word and morphological word family frequencies were tallied. Lists of words and word families by grade level and series were created and are available from the first author on request to researchers, curriculum developers, policy makers, and publishers to improve vocabulary curricula and instruction. Based on initial (Version 1) analyses, the present brief summarizes the word and family frequencies by discipline and series and elaborates selected points about elementary grade children’s exposure to words and word families in disciplinary textbooks.

Student exposure to words and word families in domain-specific texts is likely to be especially significant because of the concept-building capacity of disciplinary materials (cf. Adams, 2009; Bravo & Cervetti, 2008). However, disciplinary vocabulary can be abstract and conceptually dense (Nagy, Anderson, & Herman, 1987)—features that can create lexical overload for students. Knowledge of the words and word families that appear in such current-day texts is extremely valuable for instructional purposes. As well, related to the Common Core State Standards for English Language Arts (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) emphasis on academic words, no comprehensive contemporary resources exist to inform educators, researchers, and publishers about academic words contained in elementary-grades disciplinary textbooks. (The widely cited Coxhead [2000] Academic Word List was drawn from university textbooks.) The present research addressed this void.

METHODS

Data Sources: Four current, best-selling (Education Market Research, 2014) disciplinary textbook series were identified in each of three disciplines—science, social studies, and mathematics. Five publishers were represented. (Series are noted as A, B, etc. in the present brief, and the publishers are identified in the “Literature Cited” section of the References.)

The textbooks were digitized and edited according to a standardized protocol with the overarching goal of including all the words that children were intended to read and excluding words that were in the texts primarily for teachers’ benefit. For example, callouts that provided Common Core State Standards were excluded. To reduce the presence of potential non-words arising from Optical Character Recognition scanning errors and other irregularities, character strings not appearing in the 450-million-word *Corpus of Contemporary American English* (<http://corpus.byu.edu/coca/>) were removed from further analyses.

The resulting corpus contained 3,453,554 running words.

Units of Analysis and their Identification: One unit of analysis was a word (a “type”). A word was operationally defined as a set of connected letters with surrounding white space (cf. Carroll, Davies, & Richman, 1971).

Morphological families consisted of three units: base word, inflected form, and derived form. Before defining the three units, two additional morphological terms are required. A *morpheme* is a meaningful unit of language that cannot be further divided (“Morpheme,” 2002) (e.g., *unladylike* consists of three morphemes, *un*, *lady*, and *like*); a *free morph* is a unit that can stand on its own as a word (e.g., *lady*). To return to defining the units, a *base word* is a free morph and when it is used in combination with affixes, the base meaning remains unchanged from its original use in the word (cf. Bauer & Nation, 1993). A *morphological inflection* is “the modification of a word to express different grammatical categories such as tense, mood, voice, aspect, person, number, gender, and case” (“Inflection,” 2015) (e.g., for the base, *determine*—*determines*, *determining*, *determined*). A *morphological derivation* is “. . . the process of forming a new word on the basis of an existing word. . . It often involves the addition of a morpheme in the form of an affix . . .” (“Morphological Derivation,” 2015) (e.g., *unhappy* from the base *happy*; *determination* from the base *determine*).

Finally, morphological families were divided into four units called levels based on a modification of Bauer and Nation’s (1993)

and White and colleagues' (White, Sowell, & Yanagihara, 1989) morphological levels: Level 1—every word form; Level 2—base words and their inflected forms and derived forms with the suffixes such as *-ed past*, *-en past participle*, and *-ing present participle*; Level 3—all the forms of Levels 1 and 2 plus the 10 most frequent prefixes and suffixes (that are not in Level 2:); and Level 4—all the forms in Levels 1 through 3 plus 107 prefixes and 108 suffixes listed in the *English Lexicon Project (ELP)* database (Balota et al., 2007).

ANALYSES

All analyses were accomplished computationally. To determine base words and family levels, a computer-based two-stage process was used. First, the *ELP* (Balota et al., 2007) dataset was used. Second, if a word was not in the *ELP*, then a computer-based morphological analysis was accomplished using a modified form of the software, *Morfessor 2.0* (Virpioja, Smit, Grönroos, & Kurimo, 2013) to segment words into morphemes, and each word form was assigned to one of the four levels of word families. The reliability of the computational determinations was checked by comparing them to the decisions of two judges with extensive backgrounds in English, language, and literacy education and research. For a random selection of words stratified by frequency (with 300 high frequency and 300 other words), after approximately two hours of training, the two judges' agreements with the computationally determined base word were 84% for each judge. For determination of family level, agreements were 85% for each judge.

RESULTS & DISCUSSION

Tokens and Types: *Substantial variation in total number of words (tokens for all grades combined) existed across series within discipline.* Table 1 shows that in each discipline, children who received a cumulatively long series through the elementary grades (e.g., series B in social studies) were exposed to 1.5 to 2 times as many words as children who received a short series (e.g., series D in social studies). Word exposure through wide reading is one means of vocabulary expansion (Nagy et al., 1987), so series length could play a role in disciplinary vocabulary development.

Table 1
N of Tokens by Discipline and Series, across All Grades

Social Studies				
A	B	C	D	Total
326,734	398,987	200,185	234,505	1,160,411
Science				
E	F	G	H	
292,285	397,866	303,309	352,874	1,346,334
Mathematics				
I	J	K	L	
133,724	367,465	307,772	137,848	946,809

For unique word exposure (types), for series within discipline, there was considerable variation in total number of unique words (across all grades). Table 2 shows that for social studies and mathematics, by the end of fifth grade, children in series with the most types (e.g., series B in social studies) were exposed to roughly 1.5 to 2 times as many types as peers in series with the fewest types (e.g., series C in social studies). The differences across the four science series were less dramatic. Again, quantity of unique word exposure through text reading is important for children's vocabulary growth, so series exposure could play a role in students' disciplinary vocabulary development.

Table 2

N of Types by Discipline and Series, across All Grades

Social Studies			
A	B	C	D
10,912	12,722	8,505	9,377
Science			
E	F	G	H
8,480	8,716	9,343	9,864
Mathematics			
I	J	K	L
5,093	7,400	7,156	3,525

Word Families: For all three disciplines, when morphological families were collated at Level 2, there was a substantial reduction in vocabulary load for children, depending on series from 27% to 35% (see Table 3). Further collocation of types at Level-3 and -4 families made little difference in load reduction, ranging from 3% to 10%. The result supports the contention that knowing morphological affixes of Level-2 type could be quite productive for expanding children's vocabulary knowledge.

Table 3

Word Families by Series within Content Area and Morphological Level (Counted across All Grades) (Using Unique Types)

	Level-1 Unique Types	Level-2 Unique Types (% Reduction from Level-1)	Level-3 Unique Types (% Reduction from Level-2)	Level-4 Unique Types (% Reduction from Level-3)
<u>Social Studies</u>				
A	10,912	7,528 (31%)	6,881 (9%)	6,508 (5%)
B	12,722	8,638 (32%)	7,869 (9%)	7,385 (6%)
C	8,505	5,893 (31%)	5,422 (8%)	5,146 (5%)
D	9,377	6,549 (30%)	6,047 (8%)	5,739 (5%)
<u>Science</u>				
E	8,480	5,561 (34%)	5,041 (9%)	4,776 (5%)
F	8,716	5,876 (33%)	5,344 (9%)	5,005 (6%)
G	9,343	6,058 (35%)	5,519 (9%)	5,220 (5%)
H	9,864	6,449 (35%)	5,822 (10%)	5,508 (5%)
<u>Mathematics</u>				
I	5,093	3,699 (27%)	3,563 (4%)	3,444 (3%)
J	7,400	5,367 (27%)	5,153 (4%)	4,978 (3%)
K	7,156	5,075 (29%)	4,849 (4%)	4,683 (3%)
L	3,525	2,415 (31%)	2,305 (4%)	2,245 (3%)

Repetition of Unique Word Types: The median repetition of unique word types was computed and summed across all five grades, for each series within each discipline. *The median repetition ranged consistently from three to five.* There is strong evidence that repetition of words aids during reading student vocabulary learning and remembrance (Hiebert & Fisher, 2006). Whether three to five repetitions is sufficient for student vocabulary learning from disciplinary textbooks is unknown, but the results of the present research raise the question of its adequacy.

Conclusion: To gain breadth and depth of disciplinary knowledge, students must attain control over domain-specific vocabulary (Armbruster, 1992; Schleppegrell, 2007). With recent renewed emphasis on the integration of literacy in the disciplines (e.g., the Common Core emphasis), including the elementary grades (Moss, 2011), better understanding of the vocabulary in disciplinary textbooks is an essential foundation for enhanced instruction.

REFERENCES

- Armbruster, B. B. (1992). Science and reading. *The Reading Teacher*, 46, 346–347.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). *The English Lexicon Project. Behavior Research Methods*, 39, 445–459.
- Bravo, M. A., & Cervetti, G. N. (2008). Teaching vocabulary through text and experience in content areas. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 130–149). Newark, DE: International Reading Association.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 68, 253–279.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. New York: American Heritage.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Education Market Research. (2014). *Reading market 2012*. New York: Author.
- Hiebert, E. H., & Fisher, C. W. (2006). Fluency from the first: What works with first graders. In T. Rasinski, C. L. Z. Blachowicz, & K. Lems (Eds.), *Teaching reading fluency: Meeting the needs of all readers* (pp. 279–294). New York: Guilford.
- “Inflection.” (2015). In *Wikipedia*. Retrieved October 16, 2015. <https://en.wikipedia.org/wiki/Inflection>
- “Morpheme.” (2002). In *Oxford English Dictionary*. Third Edition.
- “Morphological Derivation.” (2015) in *Wikipedia*. Retrieved October 16, 2015. https://en.wikipedia.org/wiki/Morphological_derivation
- Moss, B. (2011). Making a case and a place for effective content area literacy instruction in the elementary grades. *The Reading Teacher*, 59, 46–55.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237–270.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/Appendix_A.pdf.
- Schleppergrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly*, 28, 35–54.
- Virpioja, S., Smit, P., Grönroos, S-A., & Kurimo, M. (2013). *Morfessor 2.0: Python Implementation and Extensions for Morfessor Baseline. Aalto University publication series SCIENCE + TECHNOLOGY*, 25/2013. Helsinki: Aalto University.
- White, T. G., Sowell, J., & Yanagihara, A. (1989). Teaching elementary students to use word-part clues. *The Reading Teacher*, 42, 302–308.

Literature CitedMathematics

Fusion, K. C. (2013). *Math expressions Common Core* [Series]. Orlando, FL: Houghton Mifflin Harcourt.
Kendall Hunt. (2015). *Math trailblazers Common Core State Standards* (4th ed.) [Series]. Chicago, IL: Author.
McGraw-Hill Education. (2012). *Everyday mathematics* [Series]. Columbus, OH: Author.
Pearson Education. (2011/2012). *enVisionMath Common Core* [Series]. Glenview, IL: Author.

Science

DiSpezio, M. A., Frank, M., Heithaus, M., & Ogle, (Consulting Authors). (2015). *Science fusion* [Series]. Orlando, FL: Houghton Mifflin Harcourt.
Macmillan/McGraw-Hill. (2008/2011). *Science: A closer look* [Series]. Columbus, OH: Macmillan/McGraw-Hill.
Pearson Education. (2010). *Scott Foresman Science: Diamond Edition* [Series]. Glenview, IL: Author.
Pearson Scott Foresman. (2010). *Interactive science* [Series]. Upper Saddle River, NJ: Author.

Social Studies

Banks, J. A., Colleary, K. P., Greenow, L., Parker, W. C., Schell, E. M., & Zike, D. (2009). *Macmillan/McGraw-Hill Timelinks* [Series]. Columbus, OH: MacMillan/McGraw-Hill.
Bennett, L., Cummins, J., Kracht, J. B., Tatum, A. W., & White, W. E. (2013). *Pearson my world social studies*. [Series]. Boston, MA: Pearson Education.
Berson, M. J., Howard, T. C., & Salinas, C. (2010/2012). *Harcourt social studies* [Series]. Orlando, FL: Houghton Mifflin Harcourt.
Boyd, C. D., Gay, G., Geiger, R., Kracht, J. B., Pang, V. O., Risinger, C. F., Sanchez, S. M. (2011). *Scott Foresman social studies* [Series]. Boston, MA: Pearson Education.

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.

