



WHITE PAPER

Text Complexity of English International Newspapers in a World Preparing for College and Career

Carl W. Swartz, Sean T. Hanlon, A. Jackson Stenner, Hal Burdick and Donald S. Burdick,
Learning Science and Technology, MetaMetrics

Abstract

English is the unofficial technical and business language of the world. Estimates suggest that more than 1 billion people worldwide use English to varying degrees of understanding and expression. A common second language like English enables the internet to function as a digital passport allowing those whose first language might be Russian, Arabic, Cantonese, French, Spanish, or Hindi to cross international borders and share understanding of local, national, and international events and cultures. The purpose of this study was to investigate the text complexity of online English language newspapers sampled from around the world. The results of this study suggest that the text complexity of online English newspapers is commensurate with the complexity of text encountered by readers in two and four year universities, colleges and the workplace and is slightly higher than the text demands of domestic newspapers. Text at this high level may prove to be a barrier to understanding across borders and cultures. But, the level of text complexity sets an implicit aspirational goal for those who desire to be educated in or work in the United States. Our goal is not to advocate for lowering the text complexity of online English newspapers, but to enhance the reading ability of all English language learners who desire to access the information and knowledge contained in college and career texts.

Keywords: text complexity, on-line English newspapers, Common Core State Standards, Lexile

Text Complexity of English International Newspapers in a World Preparing for College and Career

English may be the unofficial global language of the modern era (Graddol, 2000). The approximately 1 million words found in the English language comprise the third most spoken native language. Estimates of the total number of native and non-native users of English vary from 450 million to more than 1 billion people (Central Intelligence Agency [CIA], 2009; Wyne, 2009). Mydans (2007) estimated that more than 400 million people are fluent in English as a first language with another 300-500 million people fluent in its use as a second language. Crystal (2003) estimated that non-native speakers of English outnumbered native speakers 3:1. It is interesting to note that while the United States has approximately 309 million people within its borders, more than 300 million people in China and more than 350 million people in India use English on a daily basis (Wyne, 2009).

The potential for English to be a shared language that can advance understanding among people without regard to borders or cultures is confirmed by its increasing ubiquity in the workplace and in higher education. English is the official language of the United Nations, international treaties, as well as airborne and nautical communication. English is the primary or secondary language spoken around the world and is used in more than 90% of the world's science journals (CIA, 2009; Mydans, 2007, Wyne, 2009). The world's top graduate and undergraduate business schools are moving towards teaching courses in English. The number of master's degree programs in business offered in both English and the language native to the university has more than doubled to 3,300 programs in 1,700 universities; South Korean universities offer almost 30% of their courses in English (Carvajal, 2007).

But not all individuals who learn English or want to access information written in English do so while pursuing university educations or working for multi-national corporations. Online information written in English may play an increasingly pivotal role in advancing understanding among people regardless of first language, borders, educational attainment, or educational aspirations. Almost every country around the world publishes at least one online newspaper in the language native to its citizens as well as in English (Wyne, 2009). Access to relatively lower cost devices (e.g., smartphones, laptops) is helping people to overcome issues related to accessing technical and business data found on the internet especially for young people who increasingly use smartphones for more than voice and text messages (Nielsen, 2010).

An examination of world-wide Internet usage statistics revealed that, over the last decade, the amount of Internet users has increased 445% (see Table 1 on the following page). The greatest increase in usage occurred in Africa and the Middle East, followed by Latin America/Caribbean and Asia. This increased growth coincides with improved economic conditions, increased international trade, increased access to technology, and greater openness by governments. Even though a smaller percent of people in Asia have access to the Internet compared to Europe, North America, and Oceania/Australia, the total number of individuals who have access represents double the number of users.

Table 1. Internet Usage Statistics by World Region for 2010¹

World Regions	Estimated Population	Number of Internet Users	Percent of Population	Change in Percent of Users (2000-2010)	Percent of World Population Using Internet
Africa	1,013,779,050	110,931,700	10.9%	2,357.3%	5.6%
Asia	3,834,792,852	825,094,396	21.5%	621.8%	42.0%
Europe	813,319,511	475,069,448	58.4%	352.0%	24.2%
Middle East	212,336,924	63,240,946	29.8%	1,825.3%	3.2%
North America	344,124,450	266,224,500	77.4%	146.3%	13.5%
Latin America/Caribbean	592,556,972	204,689,836	34.5%	1,032.8%	10.4%
Oceania/Australia	34,700,201	21,263,990	61.3%	179.0%	1.1%
World Total	6,845,609,960	1,966,514,816	28.7%	444.8%	100.0%

¹Adapted from Internet World Stats (January 1, 2011; <http://www.internetworldstats.com>)

Common Core Standards: Setting an International Bar for College and Career Readiness

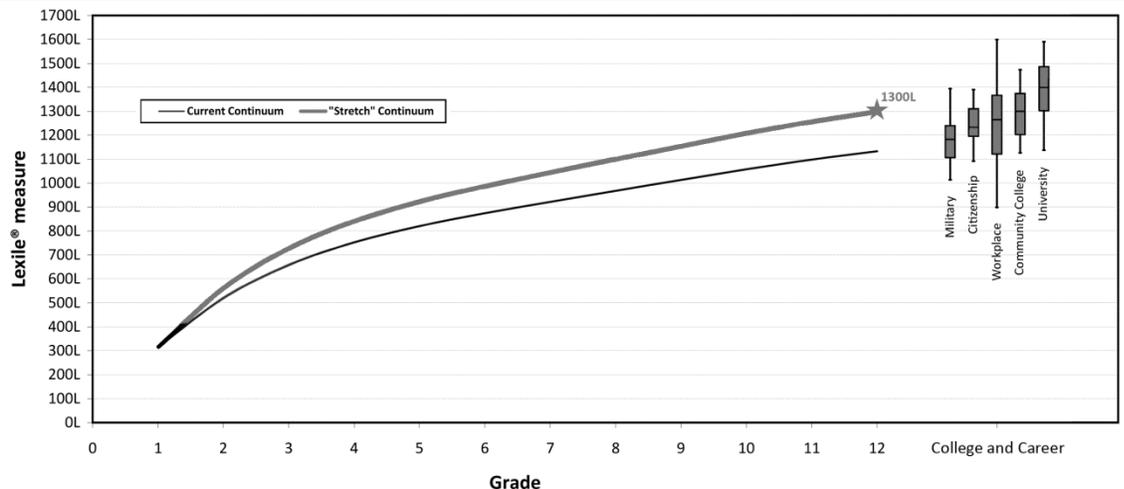
Access to the growing amount of information published in English on the Internet is improving each year. There is more to understanding and using information contained in online English newspapers or other content (e.g. Wikipedia) than easy access. Understanding and use is governed by comprehension and comprehension has been shown to be caused by the match between reader ability and text complexity. Readers who are matched to text will enjoy a higher degree of comprehension and self-engagement than readers who encounter text beyond their ability. Understanding the difficulty of text found in online English newspapers may impact current efforts to codify national and international benchmarks for 21st Century learners and promoting college and career readiness by increasing the amount of informational text contained in newspapers (Duke, 2010; Phillips & Wong, 2010).

Several recent studies have examined the complexity of reading material typically encountered in college and career (MetaMetrics, 2008a; Williamson, 2008). Additionally, MetaMetrics (2008b) examined the difficulty of textbooks in Grades 1 through 12 to extend the earlier investigations of Chall, Conrad, and Harris (1977) and Hayes, Wolfer, and Wolf (1996) into the complexity of text defining college and career readiness. Text for these studies were obtained from several sources. The text demands of commonly adopted textbooks used in grades 1-12 was measured. Textbooks were selected to represent multiple subject areas to better incorporate the full range of texts a student is likely to encounter during K-12 education. Included in the sample at each grade were health, language arts, literature, mathematics, science and social studies textbooks. The texts used to define college and career readiness were obtained from several state studies commissioned by Higher Education Coordinating Boards. The textbook analyses have been juxtaposed with military documents, citizenship documents and workplace materials (Williamson, 2008).

The results of these studies suggest that the general range defining college and career readiness in terms of text complexity encountered by adults is 1200L to 1400L (see Figure 1). The largest increases in text complexity occur in the early elementary years and correspond to the period when students are experiencing tremendous growth in reading ability. Growth in text complexity slows in the middle and high school years.

Figure 1. Observed text complexity at each grade and stretch text complexity at each grade given standard set at college and career readiness.

Adapted from Stenner, A. J., Koons, H. H., & Swartz, C. W. (2009). *Text Complexity, the text complexity continuum, and developing expertise in reading*. Durham, NC: MetaMetrics.



The purpose of this study was to further enhance the interpretation of “college and career readiness” by including on-line English newspapers published around the world and to compare the international with the domestic perspective. This is the first study conducted to estimate text complexity of text found in international newspapers that may be accessed without regard for first language and national boundaries. The results may inform the setting of international grade or grade band text-based standards.

METHOD

Sampling On-Line International English Newspapers

In an effort to provide an accurate representation of on-line international newspapers published in English, articles were sampled from 39 newspapers from 28 different countries. The newspapers were representative of various regions from around the world (i.e., Africa, Asia, Europe, Middle East, North America, Latin America/Caribbean, Oceania/Australia). Articles from each newspaper were extracted over a two-week period and focused on material found in the front section. Each text sample had at least 100,000 words. This minimum quantity was selected as it provides enough text to provide a highly reliable estimate of the text complexity of each newspaper. Articles were selected randomly from within each publication, with the only requirement being an article could not be chosen more than once.

After the articles were extracted, experts reviewed the articles to ensure that the articles were complete (i.e., were not truncated by the download process) and contained contiguous text (i.e., no box scores or poetry). After human review, the articles were processed using the Lexile Receptive Analyzer[®] and assigned a Lexile text measure (see Table 2).

Table 2. Lexile Measures for Selected On-Line International Newspapers (2008-2010)

<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>	<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>
<i>The Egyptian Gazette (Egypt)</i>	319	101,347	1440L	<i>The Australian (Australia)</i>	186	103,652	1390L
<i>Oman Daily Observer (Oman)</i>	315	105,126	1430L	<i>German Times (Germany)</i>	171	101,315	1390L
<i>Financial Times (Great Britain)</i>	195	100,996	1430L	<i>The Copenhagen Post (Denmark)</i>	451	101,299	1390L
<i>Gulf Times (Qatar)</i>	240	101,135	1420L	<i>Malaysian Star (Malaysia)</i>	316	104,180	1390L
<i>Straits Times (Singapore)</i>	290	105,189	1410L	<i>Bahrain Tribune (Bahrain)</i>	296	103,652	1380L
<i>China Dailey (China)</i>	235	101,186	1400L	<i>Irish Times (Ireland)</i>	198	101,160	1380L
<i>Manila Bulletin (Philippines)</i>	342	146,993	1400L	<i>New York Times (United States)</i>	111	100,800	1380L
<i>France Dailey (France)</i>	155	101,044	1400L	<i>Santiago Times (Chile)</i>	224	104,075	1380L
<i>The Moscow Times (Russia)</i>	229	100,708	1400L	<i>Sydney Morning Herald (Australia)</i>	221	112,842	1380L

Table 2 (continued). Lexile Measures for Selected On-Line International Newspapers (2008-2009)

<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>	<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>
<i>Arab News (Saudi Arabia)</i>	230	102,239	1370L	<i>South China Morning Post (Hong Kong)</i>	191	100,752	1330L
<i>Gulf News (United Arab Emirates)</i>	388	103,276	1370L	<i>Times of India (India)</i>	288	99,670	1320L
<i>Jerusalem Post (Israel)</i>	156	102,642	1370L	<i>Wall Street Journal (United States)</i>	134	100,239	1320L
<i>Sweden Globe (Sweden)</i>	169	105,459	1370L	<i>The Guardian (Great Britain)</i>	167	101,330	1310L
<i>The Chosun Ilbo (Korea)</i>	378	106,813	1360L	<i>The Independent (Great Britain)</i>	141	104,590	1310L
<i>Daily Mail and Guardian (South Africa)</i>	187	101,963	1350L	<i>Mexico Daily (Mexico)</i>	165	102,251	1310L
<i>Japan Times (Japan)</i>	175	103,537	1350L	<i>Philippine Daily Inquirer (Philippines)</i>	272	142,354	1310L
<i>Philippine Star (Philippines)</i>	162	101,516	1350L	<i>Belfast Telegraph (Ireland)</i>	295	104,118	1290L

Table 2 (continued). Lexile Measures for Selected On-Line International Newspapers (2008-2009)

<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>	<i>Newspaper (Country)</i>	<i>Number of Articles</i>	<i>Number of Words</i>	<i>Lexile Measure</i>
<i>The New Zealand Herald (New Zealand)</i>	227	105,558	1290L	<i>USA Today (United States)</i>	153	100,613	1260L
<i>New Strait Times (Malaysia)</i>	269	102,306	1280L	<i>The Daily Telegraph (India)</i>	259	103,509	1240L
<i>Pravda (Russia)</i>	175	105,157	1280L				

Materials and Procedures

The Lexile® Framework for Reading is a psychometrically-rigorous system that places readers and text on the same developmental scale. A Lexile measure, denoted by a numeric value followed by a trailing *L*, is a quantity that represents an individual's reading ability or alternatively the complexity of a piece of text. The scale ranges from 0L, or Beginning Reader, to above 1800L. Like a degree on a thermometer, the higher the Lexile measure, the more reading ability a student possesses or the more complex a piece of text. A Lexile measure is assigned to text by a piece of software, The Lexile Receptive Analyzer that evaluates the semantic and syntactic complexity of the text. The analyzer can be used at no charge for non-commercial purposes by educators around the world.

Results

The text complexities of each publication provide an interesting perspective as all the newspapers are within 200L of each other. The least complex newspaper, India's *The Daily Telegraph*, measured 1240L whereas the most complex publication, Egypt's *The Egyptian Gazette*, measured 1440L. The complexity of text for three domestic newspapers was 1380L for the *New York Times*, 1320L for the *Wall Street Journal*, and 1260L for *USA Today*.

Not only do the newspapers cluster themselves into a tight 200L band, but where the band is located on the Lexile scale is also intriguing. Stenner, Koons, and Swartz (2009) showed that, in order to meet the text demands of college and career, readers need to be reading at approximately 1200L-1400L. The international newspapers are distributed within this college and career readiness standard with a small amount of additional complexity (40L). This suggests that, regardless of country of origin, to successfully engage English text commensurate with college and career, readers must be reading at 1200L-1400L. While understanding the expectation in terms of Lexile measure (i.e., 1300L) is useful, it is equally important to know how readers typically grow to meet this standard (see Stenner, Koons, & Swartz, 2009).

Discussion

The growing ubiquity of technology, increased access to the Internet, and increased use of English as a common language in business, technology, and education has truly made the world smaller. People in India can read the latest *New Your Times* article with a few mouse clicks. Similarly, an individual in Mexico can read the latest op-ed published in English by a Chinese newspaper. Parents, educators, and policy-makers can work to ensure that their citizens are prepared to become members of a global community by understanding the reading demands of college and career text. Various technologies may be used to promote developing readers' ability to successfully encounter more complex text. This will, in part, insure greater understanding of events across borders, languages, cultures.

References

- Carvajal, D. (2007, April 11). English as language of global education. *New York Times*. Retrieved from <http://www.newyorktimes.com>
- Chall, J. S., Conrad, S. S., & Harris, S. H. (1977). An analysis of textbooks in relation to declining SAT scores. New York, NY: College Entrance Examination Board.
- Central Intelligence Agency (2009). *Central intelligence agency: Factbook*. Washington DC: Central Intelligence Agency.
- Crystal, D. (2003). *English as a global language*. Cambridge, U.K.: Cambridge University Press.
- Duke, N. K. (2010). The real-world reading and writing U.S. children need. *Phi Delta Kappan*, 68-71.
- Graddol, D. (2000). *The future of English: A guide to forecasting the popularity of English in the 21st Century*. London, United Kingdom: The English Company.
- Hayes, D. P., Wolfer, L. T., & Wolfe, M. F. (1996). Schoolbook simplification and its relation to the decline in SAT-verbal scores. *American Educational Research Journal*, 33(2), 489-508.
- MetaMetrics, Inc. (2008a). Texas Higher Education Coordinating Board: Text Measurement and Analysis. MetaMetrics Technical Report Update.
- MetaMetrics, Inc. (2008b). The text continuum in 2008. Presented at the Lexile National Conference, San Antonio, TX.
- Mydans, S. (2007, April 9). Across cultures, English is the word. *New York Times*. Retrieved from <http://www.newyorktimes.com>
- Phillips, V, & Wong, C. (2010). Tying together the common core of standards, instruction, and assessments. *Phi Delta Kappan*, 37-42.
- Stenner, A. J., Koons, H. H., & Swartz, C.W. (2009). Text Complexity, the text complexity continuum, and developing expertise in reading. Durham, NC: MetaMetrics.
- Williamson, G. L. (2008). A text complexity continuum for postsecondary readiness. *Journal of Advanced Academics*, 19(4), 602-632.
- Wyne, A. (2009, June 8). The language of empire: Why English will keep America's influence from waning. *Foreign Policy*. Retrieved from <http://experts.foreignpolicy.com>

Author Notes:

Correspondence concerning this article should be addressed to Carl W. Swartz, MetaMetrics, 1000 Park Forty Plaza, Suite 120, Durham, NC 27713. E-mail: cswartz@lexile.com.

Special thanks to Lauren Whitehead, Emily Weller and Carrie Jackson for selecting, editing and measuring the articles used in this research study.

MetaMetrics, an educational measurement and research organization, develops scientific measures of academic achievement that link assessment with targeted instruction to improve learning. The organization's renowned psychometric team created The Lexile Framework for Reading; El Sistema Lexile para Leer, the Spanish-language version of the reading framework; The Quantile Framework for Mathematics; and The Lexile Framework for Writing. In addition to licensing Lexile and Quantile measures to state departments of education, testing and instructional companies, and publishers, MetaMetrics offers professional development, resource measurement and customized consulting services.

METAMETRICS®, the METAMETRICS® logo and tagline, LEXILE®, LEXILE FRAMEWORK®, LEXILE ANALYZER®, the LEXILE® logo, QUANTILE®, QUANTILE FRAMEWORK® and the QUANTILE® 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 © MetaMetrics, Inc. All rights reserved.