Understanding Text Complexity

Article *in* The Elementary School Journal · December 2014

DOI: 10.1086/678446

CITATIONS
17

READS
1,617

2 authors:

Elfrieda Hiebert
TextProject.org
134 PUBLICATIONS **4,043** CITATIONS

SEE PROFILE

P. David Pearson
University of California, Berkeley
284 PUBLICATIONS **15,079** CITATIONS

SEE PROFILE

UNDERSTANDING TEXT COMPLEXITY

Introduction to the Special Issue

ABSTRACT

The Common Core State Standards represent the first standards document to address whether students are able to read progressively more complex texts as they progress across the grades. This article gives an overview of the three components of the model of text complexity that were identified in Appendix A of the Standards and also were the basis for the selection of manuscripts for this special issue: (*a*) qualitative, (*b*) quantitative, and (*c*) reader-task considerations. This introduction gives an overview of the three components and the contributions of the articles in this special issue to extending understanding about these three components.

Elfrieda H. Hiebert

TEXTPROJECT & UNIVERSITY OF CALIFORNIA, SANTA CRUZ

P. David Pearson

UNIVERSITY OF CALIFORNIA, BERKELEY

THE presence of Standard 10 in the Common Core State Standards (CCSS) (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010), which assesses students' ability to comprehend increasingly more complex texts as they move through the grades, is a game changer in the policy world. This goal, while underlying the many literacy activities that consume the majority of instructional time in schools, has been unarticulated in previous standards documents. Presumably, in prior state and national assessments, texts for fourth graders are easier than those for fifth graders but harder than those for third graders. But how much harder or how much easier? And how have assignments of difficulty been determined?

Over the past century, answers to these questions have been many (e.g., Klare, 1984), but no previous policy document has either targeted increased capacity with complex texts as a separate standard or been as explicit as the CCSS in identifying

THE ELEMENTARY SCHOOL JOURNAL VOLUME 115, NUMBER 2

quantitatively indexed goals for particular points in students' school careers. Moreover, an underlying assumption of the CCSS is that the complexity levels for all grades need to be accelerated to ensure that an average high school graduate can read the typical texts of college courses and the workplace. Evidence is accumulating that calls into question the assumption of the CCSS that the difficulty of texts from grades K to 12 has decelerated over the past 50 years, at least in the elementary grades (Gamson, Lu, & Eckert, 2013; Hiebert & Mesmer, 2013). Even so, accelerated text levels appear to be guiding text selections for school programs and the new generation of assessments. The implications of the policy shift represented by Standard 10 are substantial. For the next decade and beyond, policy-makers and educators can expect that they will be asked to interpret, justify, and explain students' English/language arts (ELA) performances in light of complexity standards.

This issue of *The Elementary School Journal* is the first collection of articles in an archival journal to address the issues involved in establishing text complexity. Specifically, this special issue is focused on the three components of the model of text complexity identified in Appendix A of the Standards: (*a*) qualitative (i.e., features that are best measured by an attentive human reader), (*b*) quantitative (i.e., features that are typically measured by computer software), and (*c*) reader-task considerations (i.e., features that are best assessed by teachers with knowledge of their students and subjects) (NGA & CCSSO, 2010). As editors of this special issue, we selected manuscripts that best illustrate current work on the three components of the model of text complexity. In this introduction, we provide a context for the three components and the articles in this issue.

Placing the Three Components of the CCSS Text Complexity Model within Historical and Current Contexts

Concern about whether what we write can be understood by our intended audience is inherent to human communication. From that standpoint, evaluating text complexity is as old as written communication. But it was not until the late nineteenth century that literary analysts (Sherman, 1893) and linguists (Rubakin, 1889, in Choldin, 1979) began the systematic study of text complexity. These early analyses of text complexity were exclusively qualitative, focusing on rich descriptions of text features that would likely impact the comprehensibility or readability of texts. As scientific methods were brought to bear on educational problems in the early twentieth century, quantitative systems appeared and quickly dominated efforts to describe the comprehensibility of texts.

Lively and Pressey (1923) proposed a formula based on word frequency and sentence length for predicting how well students would understand texts. Its use by practitioners is uncertain, but Gray and Leary's (1935) study directed attention of scholars to readability. From a group of 82 potential factors, they identified 44 factors (e.g., word length, sentence length, proportion of explicit sentences) as significantly related to reading difficulty. In an even more exhaustive investigation of linguistic factors using cloze tests (filling in a blank left for every fifth word) as a dependent measure, Bormuth (1966) found over 60 structural indices that predicted comprehension (filling in cloze blanks) difficulty.

In the 30 years between the efforts of Gray and Leary (1935) and Bormuth (1966), nearly 200 different readability formulas were developed (Klare, 1984). In a compre-

hensive review of readability research, Klare distinguished between using readability formulas to forecast comprehension and using them to edit texts to attain readability targets. Readability scores were generally correlated with comprehension scores, Klare concluded, but lower scores were not necessarily the reason for better comprehension since readability scores had only been validated for purposes of prediction. Klare's cautionary observations turn up repeatedly in the history of measuring text complexity: (*a*) Does linguistic complexity produce barriers to comprehension, or simply reflect the complexity of the ideas the language represents? (*b*) How precisely can a readability score predict a reader-book match for a given student or class?

With the cognitive turn in psychology (Gardner, 1987), attention turned away from predicting reading difficulty and toward understanding the roles of particular text features in readers' cognitive processing of information. During this era, various analyses of text structure at both the micro- (sentence) and macrolevels (paragraphs and rhetorical structures) were invoked to explain text comprehension (e.g., Meyer, 1975). But the big ideas in the 1970s and 1980s were situated in the reader, not the text. Knowledge of ideas, represented as schemata in long-term memory, along with executive control processes (Brown, Armbruster, & Baker, 1986), dominated research on comprehension. The emphasis in comprehension studies was on synthesis (how humans integrate separate inputs into a coherent whole) rather than on analysis (how humans deconstruct units to examine their infrastructure).

This was not a comfortable context for a construct like readability, with its insistence that long words and long sentences were predictive of, or perhaps even shaped, comprehension. Linguists (e.g., Davison & Kantor, 1982) examined consequences of changes in lexical and syntactic form on reader processes and comprehension, demonstrating that readers could be burdened with extra inferential tasks when writers of instructional materials broke longer, grammatically complex sentences into shorter, grammatically simpler ones. Other scholars (e.g., Pearson, 1974–1975) confirmed the negative impact of readability formula-driven adaptations on comprehension empirically, showing that texts with higher readability scores elicited superior comprehension.

By the mid-1980s, these critiques had reached a point where, in the highly influential National Academy of Education report *Becoming a Nation of Readers*, Anderson, Hiebert, Scott, and Wilkinson (1985) took the position that readability formulas, while providing a general guide to text placement, need to be supplemented with more qualitative analyses of text. This critique was reflected in the policies of large states that limited the use of readability formulas to create and select texts for their centralized textbook adoptions (California English/Language Arts Committee, 1987; Texas Education Agency, 1990).

Partially in response to these questions about readability formulas and partially because experience taught them that readability formulas are especially inaccurate at the earliest stages of reading, other scholars (e.g., Carver, 1990; Clay, 1991) identified alternative approaches to scale books on some continuum of complexity, particularly for use in instruction. The fundamental procedure was to assemble expert teachers to examine and then, using their experience in the classroom with students and texts, to assign developmental levels to texts. In educational pedagogy, the emphasis remained on the reader, with the reasoning that text challenge could be compensated for if educators could just tap into or build reader knowledge and/or interest prior to reading.

At the same time that the role of quantitative measures was being disputed in research and policy, the digital revolution was well underway. Rapid digital processing meant that large numbers of texts could be analyzed quickly, and, with large corpora stored in databanks, rankings of the frequency of words in texts could be gotten in nanoseconds, rather through laborious hand scoring (Thorndike, 1921) or the time-consuming data entry for processing by mainframe computers (Carroll, Davies, & Richman, 1971; Kučera & Francis, 1967). The increased speed of processing was used by several organizations (e.g., Smith, Stenner, Horabin, & Smith, 1989; Zeno, Ivens, Millard, & Duvvuri, 1995) to develop readability systems drawing on word frequencies garnered from massive databanks of texts as well as information on sentence length. During the 1990s and early 2000s, these digitized readability systems became widely used by school publishers and districts and states to guide selection of texts for instruction and assessments.

Digitization also set the stage for more sophisticated analyses that built on the work of cognitive scientists as they simulated the functioning of human intelligence. Complex mathematical models could be used to analyze the presence of text features that had been identified in micro- and macrostudies of readers' processing of texts. When, in early 2010, CCSS writers finalized the model of text complexity within the Standards, it was understandable that their model would reflect the primary strands within practice and scholarship: qualitative, quantitative, and reader-task.

Qualitative Systems of Text Complexity

Qualitative systems predated quantitative and, in the day-to-day classroom activities in current American schools, qualitative judgments are ubiquitous. Our review (Pearson & Hiebert, 2014, in this issue), however, shows that the scholarly literature contains few validations of these current systems. Another aspect that becomes apparent in our review is that in many of the applications of the CCSS' text-complexity model, the purpose has been to assign texts to a level. Such use of qualitative analyses, we note in the article, fails to give the rich information that teachers need to guide students through texts, curriculum developers to craft lessons for teachers, or test developers to create appropriate assessments. We conclude that much more work is needed to create qualitative text-analysis systems that provide the kind of information needed for curriculum, instruction, and assessment functions.

Quantitative Systems of Text Complexity

Despite explicit guidance on the quantitative component within Appendix A (which the qualitative and reader-task components lacked), the only supplement to the Standards that has been issued to date is on the quantitative component. This supplement reports on a study (Nelson, Perfetti, Liben, & Liben, 2012) commissioned by CCSS developers and claims, in its subtitle, "New Research on Text Complexity." The text-complexity measures in the study may be new relative to those reviewed by Klare (1984), but Nelson et al.'s finding of moderate to high correlations between features of texts and assigned text levels or student outcomes is neither new nor one that imputes causality. Cunningham and Mesmer (2014, in this issue) raise questions about conclusions and interpretations made on the basis of such correlations. In particular, they describe the kind of studies that would be needed to provide

valid data for designing diets of text-complexity elements or designating text-difficulty levels of students.

Moving beyond the use of quantitative systems to assign reading levels of individual students or levels of individual texts, Williamson, Fitzgerald, and Stenner (2014, in this issue) consider how quantitative data can be the basis for evaluating and describing growth models. Williamson et al. use historical data to illustrate the manner in which growth models can lead to distinctly different policies. The historical data on which Williamson et al. build the growth models illustrate the current trajectories of students, especially in different quartile groups. The models can give insight into where resources and efforts would need to be allocated if the trajectories of students who are failing to reach college- and career-ready levels by high school graduation are to change.

The contributions by Graesser et al. (2014, in this issue) and by Sheehan, Kostin, Napolitano, and Flor (2014, in this issue) remind us that just because we assign a number to a particular feature does not mean that a feature is exclusively quantitative. In fact, quantitative features have no validity until and unless they are "realized" qualitatively. Average sentence length is a number, but it is the qualitative nature of those long sentences, most probably the relationships among ideas, that influences comprehension. Quantitative data can alert us to the need for qualitative attention to particular features of texts.

Further, as Graesser et al. (2014, in this issue) and Sheehan et al. (2014, in this issue) show, there may be text features that have not yet been analyzed or that can be analyzed in new ways and in new combinations to better scale text complexity. Both Graesser et al.'s Coh-Metrix system and Sheehan et al.'s TextEvaluator system may, in their new and more nuanced forms, focus attention on new features that permit more thorough and productive qualitative analyses of texts by educators and researchers.

Reader-Task Leg

The cognitive revolution turned attention to the central role of reader differences and attributes. The interactive model of reader, text, and context has been prominent in the conceptualization of comprehension for several decades. In the widely influential RAND model (Snow, 2002), these three factors are augmented by a task variable. This interaction is recognized in the triadic system of text complexity offered in the CCSS where reader-task variables form the base of the triangle. To date, the guidance for teachers in how to make decisions about reader-task variables has been vague—recommending that teachers consider students' motivation, knowledge and experience, and the purpose and complexity of tasks. But what is the professional knowledge that teachers need for making such assessments? Valencia, Wixson, and Pearson (2014, in this issue) draw on scholarship to describe the rich knowledge base on how features of tasks and readers influence comprehension of texts. They also propose text-task scenarios that could be used by teacher educators and professional development specialists in ensuring that teachers have foundations to make choices about texts, tasks, and contexts that support their students in increased capacity with complex text.

Issues of readers and tasks are complex and, as Goldman and Lee (2014, in this issue) emphasize, can be shuttled to the background in the current surge of activity

on text. We agree. Both of us have been long associated with a view of the central role of reader in the text-task-context interaction (Anderson et al., 1985; Anderson & Pearson, 1984). Nonetheless, in selecting articles for this special issue, we believe that it is essential to foreground issues of text complexity as we move fully into the era of the CCSS—if for no other reason than elucidating how, when, and why text features shape (or don't shape) deep understanding of text.

There are other reasons for highlighting text in the mix of variables that shape reading performance. One is that in the wake of the constructivist turn that captured the psychology and pedagogy of reading in the 1960s through the early 2000s, text features were relegated to a secondary status in both reading research and pedagogy. Particularly in the 15 years prior to the release of the CCSS in 2010, the study of text, especially at different developmental levels, had been underrepresented in research and pedagogy. More attention was given to the scaffolding of text (often interpreted as teachers doing the first reading of a text for students) than to ways in which particular text features influence reading proficiency. Further, Standard 10 calls for readers to demonstrate proficiency with texts of particular levels at specific points in their school careers. Within a unique context (students at individual computers) and unique tasks (those that can be scored digitally), the next generation's comprehension will be evaluated on texts deemed to represent particular levels. Much more knowledge is needed about what influences the comprehensibility of texts—and the articles in this special issue focus on that topic.

Accomplishments and Next Steps

The articles in this special issue show that various perspectives can be used—often simultaneously—to shed light on how text features can influence comprehension. At the same time, the authors of each article caution that many aspects of text complexity and their influence on readers, tasks, and contexts have yet to be addressed. Most of the articles in this collection end with the proverbial call for "much more research" that is so typical of educational research studies. On the topic of text complexity, this call needs to be taken seriously. The educational enterprise is rife with policies and practices based on incomplete information or misinterpretations of research. There is a danger in the educational enterprise of returning to some of the practices against which scholars cautioned some 30 years ago, where texts are created to hit certain numerical readability scores (Anderson et al., 1985; Davison & Kantor, 1982; Klare, 1984). Making widespread mandates in policy and changes in practice without stronger theory and research is likely to have serious repercussions on the reading experiences of students, especially the ones who are most dependent on school instruction for their literacy experiences. Understanding what makes texts complex for readers with particular profiles in particular contexts and tasks is a central one for the goal of a fully literate society. There is a substantial amount of work yet to be done. We believe that this special issue provides direction for the work of the next decade and beyond.

References

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Champaign, IL: Center for the Study of Reading.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. **1**, pp. 255–291). New York: Longman.

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, **1**(3), 79–132.

Brown, A. L., Armbruster, B., & Baker, L. (1986). The role of metacognition in reading and studying. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 49–75). Hillsdale, NJ: Erlbaum.

California English/Language Arts Committee. (1987). *English-language arts framework for California public schools (kindergarten through grade twelve)*. Sacramento: California Department of Education.

Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston: Houghton Mifflin.

Carver, R. P. (1990). Predicting accuracy of comprehension from the relative difficulty of the materials. *Learning and Individual Differences*, **2**, 405–422.

Choldin, M. T. (1979). Rubakin, Nikolai Aleksandrovic. In A. Kent, H. Lancour, & W. Z. Nasri (Eds.), *Encyclopedia of library and information science* (pp. 178–179). Boca Raton, FL: CRC Press.

Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.

Cunningham, J. W., & Mesmer, H. A. (2014). Quantitative measurement of text difficulty: What's the use? *Elementary School Journal*, **115**(2).

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, **17**(2), 187–208.

Gamson, D. A., Lu, X., & Eckert, S. A. (2013). Challenging the research base of the Common Core State Standards: A historical reanalysis of text complexity. *Educational Researcher*, **42**(7), 381–391.

Gardner, H. (1987). *The mind's new science*. New York: Basic Books.

Goldman, S. R., & Lee, C. D. (2014). Text complexity: State of the art and the conundrums it raises. *Elementary School Journal*, **115**(2).

Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, **115**(2).

Gray, W. S., & Leary, B. W. (1935). *What makes a book readable*. Chicago: University of Chicago Press.

Hiebert, E. H., & Mesmer, H. A. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, **42**(1), 44–51.

Klare, G. R. (1984). Readability. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. **1**, pp. 681–744). New York: Longman.

Kučera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Lively, B., & Pressey, S. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, **99**, 389–398.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Atlanta: Elsevier.

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 with Appendices A–C*. Washington, DC: Authors.

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012, February 1). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Report submitted to the Gates Foundation. Retrieved from http://achievethecore.org/downloads/E0201_Measures_of_Text_Difficulty.pdf

Pearson, P. D. (1974–1975). The effects of grammatical complexity on children's comprehension, recall and conception of semantic relations. *Reading Research Quarterly*, **10**, 155–192.

Pearson, P. D., & Hiebert, E. H. (2014). The state of the field: Qualitative analyses of text complexity. *Elementary School Journal*, **115**(2).

Sheehan, K. M., Kostin I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool: Helping teachers and test developers select texts for use in instruction and assessment. *Elementary School Journal*, **115**(2).

Sherman, L. A. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry*. Boston: Ginn & Co.

Smith, D., Stenner, A. J., Horabin, I., & Smith, M. (1989). *The Lexile scale in theory and practice* (Final report). Washington, DC: MetaMetrics. (ERIC Document Reproduction Service No. ED 307 577.)

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

Texas Education Agency. (1990). *Proclamation of the State Board of Education advertising for bids on textbooks* (Proclamation 68). Austin: Author.

Thorndike, E. L. (1921). *The teacher's word book*. New York: Teachers College, Columbia University.

Valencia, S., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *Elementary School Journal*, **115**(2).

Williamson, G. L., Fitzgerald, J., & Stenner, J. (2014). Student reading growth illuminates the Common Core text-complexity standard: Raising both bars. *Elementary School Journal*, **115**(2).

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.