



PROJECT MUSE®

Common Core Standards and their Impact on Standardized Test
Design: A New York Case Study

Jody N Polleck, Jill V Jeffery

The High School Journal, Volume 101, Number 1, Fall 2017, pp. 1-26 (Article)

Published by The University of North Carolina Press

DOI: <https://doi.org/10.1353/hsj.2017.0013>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/676358>

Common Core Standards and their Impact on Standardized Test Design: A New York Case Study

Jody N. Polleck
Hunter College—CUNY
jpolleck@hunter.cuny.edu

Jill V. Jeffery
Leiden University Centre for Linguistics
j.v.jeffery@hum.leidenuniv.nl

With adoption of the Common Core (CCSS) in a majority of U.S. states came development of new high-stakes exams. Though researchers have investigated CCSS and related policies, less attention has been directed toward understanding how standards are translated into testing. Due to the influence that high-stakes tests exert on classroom teaching, research is needed to investigate what kinds of changes in test content are associated with CCSS, as well as the potential impact of these changes on students and teachers. Accordingly, this case study examines changes made to one high-stakes exam by comparing pre- and post-CCSS literacy tests administered to high school students in New York. The study responds to the following: (1) How did the adoption of CCSS alter the design of high school literacy exams in New York? (2) To what extent do exams represent measures of college readiness as opposed to early college equivalence? (3) What are the implications of CCSS exam adaptations for the goal of preparing students to be college and career ready? Findings suggest that the rush to implement more rigorous CCSS exams resulted in an exceedingly long and difficult exam that is more representative of early college equivalence rather than of college readiness.

Keywords: Common Core, Standardized Testing, Literacy

As of March 2017, 42 U.S. states and five U.S. territories have adopted the English language arts Common Core State Standards (CCSS) for College and Career Readiness (Common Core State Standards Initiative, 2016). This adoption comes with the caveat that little research has explained how the adoption of standards impacts students' literacy achievement or whether implementation of CCSS prepares students to meet college and career literacy demands (Beach, 2011). Internationally and nationally, there is no strong correlation between the existence of nationwide learning standards and achievement outcomes on standardized literacy tests (Tienken, 2008; Troia et al., 2016). In fact, Finland, ranked as one of the top developed countries for education, does not use standardized tests to drive academic performance in their schools, as educational policy makers believe these assessments narrow the curriculum and lead to harmful competition (Sahlberg, 2011). Tienken (2011) recently argued that neither the Common Core Standards nor

large-scale testing programs have been shown to improve students' literacy outcomes. Given this lack of empirical support as well as historical concerns regarding the potentially damaging effects of high-stakes accountability programs, Mathis (2010) has recommended that the Common Core Standards should be measured using low-stakes assessments until they have been "subjected to extensive validation, trials, and subsequent revisions before implementation" (p. 16). Despite these recommendations, the CCSS continue to drive standardized test design and literacy curricula in most U.S. states, and, in turn, are influencing students' learning opportunities in classrooms across the country.

In this study, we investigate the impact of the standards by focusing on changes made in New York's high school English language arts (ELA) exam following the state's CCSS adoption. The Common Core Standards are comprised of five core areas designed to reflect "college and career readiness" competencies: writing, reading informational text, reading literature, speaking and listening, and language. Upon their release, literacy researchers quickly moved to examine the standards' alignment with empirical research as well as with standards established by national and international professional organizations—and noted crucial gaps. For example, with respect to writing, Applebee (2013) suggested that the developmental models for CCSS writing might be based on arbitrary distinctions; Aull (2015) questioned the theoretical basis for the standards' separation of writing and language; and Troia (2014) noted a lack of alignment between the writing standards and evidence-based practices for writing instruction with respect to the roles of motivation and goal-setting in learning to write well. Troia et al. (2016) recommended that much more research be conducted, as currently there is little evidence demonstrating CCSS effectiveness in enhancing students' writing achievement. Further, they suggest revision of the standards themselves so as to keep pace with 21st century college and career demands.

In addition to the CCSS writing standards, scholars have also questioned the basis for CCSS reading. For example, Kern (2014) noted that the CCSS supporting documents (e.g., "Appendix A") do not make any reference to international reading standards and Cassidy, Ortlieb, and Grote-Garcia (2016) asserted that the CCSS do not meet the needs of our most struggling readers. Literacy scholars have expressed concern in particular regarding the standards' explicit calls to sharply increase text complexity across grade levels. Gamson, Lu, and Eckert (2013), for example, analyzed a corpus of textbooks and found that contrary to the CCSS claim that the level of reading difficulty has declined in K-12 academic materials, the complexity of texts has actually remained constant or even increased. Hiebert and Mesmer (2013) reviewed research and concluded that CCSS pressure to introduce more difficult texts in the early grades could "widen a gap that is already too large for students who, at present, are left out of many careers and higher education" (p. 49). Fisher and Frey (2014) also claimed that the standards have not sufficiently attended to characteristics of individual readers when considering levels of text complexity. Similarly, in discussing the "close reading" strategy for reading complex text that has been explicitly advocated by CCSS ELA designers, Snow and O'Connor (2013) have raised concerns that this approach could, if applied without sufficient attention to reader characteristics such as prior knowledge, experience, and motivation, actually widen historic achievement gaps. Taken together, scholarship regarding the Common Core literacy standards has highlighted substantial gaps between the standards and the existing body of literacy research. If the standards themselves are imperfectly aligned with literacy theory, research, and professional standards, then

corresponding accountability programs, including high-stakes exams, may also present gaps.

Literacy scholars have long questioned the high-stakes use of standardized tests that has been part and parcel of standards-based reform efforts such as the CCSS initiative. For decades, standards-based initiatives in the U.S. have been implemented alongside testing programs that are designed to hold students, teachers, and school leaders accountable for student achievement by linking exam results to teacher evaluations, school restructuring decisions (e.g., school closure), and, in some cases, to students' graduation prospects. For example, some states publish their test scores, which may trigger more explicit test preparation as well as harmful competition among teachers and schools (Davis & Willson, 2015; Dooley & Assaf, 2009). Further, 25 states serving over 34 million students include exit exams as a graduation requirement (Ujifusa, 2012), representing 69% of the nation's enrollment. Given the multifaceted and federally incentivized nature of the CCSS initiative, which has required states to "link student achievement and student growth data . . . to students' teachers and principals" in order to receive federal funds (U.S. Department of Education, 2009, p. 10), CCSS implementation is likely to raise the already high stakes associated with standardized tests in the U.S. In fact, Brooks and Dietz (2012/2013) have gone so far as to say that "the initiative conflates standards with standardization" (p. 65).

It is important to examine the tests meant to measure student achievement of CCSS because, in addition to carrying high stakes, mandatory testing is known to impact instruction in unintended ways. For example, researchers have questioned the consequential validity of high-stakes tests due to their association with high levels of test preparation (Davis & Willson, 2015). As a result, due to the accountability focus of standards-based reform movements, standards themselves may matter less in terms of their influence on instruction than does their representation in standardized tests. Though high-stakes assessments were already known to exert a strong influence on instruction prior to CCSS, it is especially important to examine CCSS assessments given that these standards were explicitly designed to sharply raise the bar for literacy expectations. Thus, many students and educators in the U.S. face the dual challenge of more difficult tests that have been tied to higher stakes.

Despite the potential impact of CCSS tests, there is a dearth of research regarding their content, due at least in part to the fact that their development has been largely contracted to major educational publishing corporations, which limit access to the tests. For example, a majority of CCSS states have joined one of two conglomerates to develop Common Core aligned exams. First, the Partnership for the Assessment of Readiness for College and Careers (2015), which contracted the publishing giant Pearson, developed tests that have been implemented in 11 states and in the District of Columbia. Second, the Smarter Balanced Assessment Consortium (2015), which contracted CTB/McGraw-Hill to develop assessments, provided standardized tests for 18 U.S. states. Each conglomerate was awarded \$175 million in U.S. Department of Education grants to develop common assessments (U.S. Department of Education, 2010).

Following CCSS adoption and the subsequent design of standards-aligned tests, questions have been raised regarding the intersection of private and public interests in large-scale CCSS test development, the appropriateness of test items, and the lack of transparency with respect to how items are developed and scored. However, some CCSS adopting states such as New York, the focus of the present study, have used

locally developed CCSS-aligned exams. Though the state initially contracted Pearson to develop CCSS tests for grades 3-8¹, local governmental agencies design the CCSS-aligned high school assessments known as the Regents exams. Unlike the tests developed by educational publishing corporations such as Pearson, New York Regents test items are made public quickly following each test administration, providing stakeholders immediate and comprehensive access to the tests. Thus, changes made to New York Regents exams provide a relatively transparent case study of the types of test design adaptations that have followed CCSS adoption.

To investigate the kinds of testing adaptations associated with CCSS adoption, this study examined the New York CCSS Regents ELA test and compared it to the previously administered New York ELA Comprehensive Regents exam. We qualitatively and quantitatively analyzed both versions to determine how the CCSS Regents differed from the previously administered Regents exam, and the extent to which each exams' items corresponded to the CCSS literacy standards. To investigate the extent to which new Common Core exams evaluated college and career readiness as opposed to college equivalence, we also comparatively analyzed Regents and Advanced Placement (AP) examinations. This study asked, (1) How did the adoption of CCSS alter the design of high school literacy exams in New York? (2) To what extent do the exams represent measures of college readiness as opposed to early college equivalence? and (3) What are the implications of CCSS exam adaptations for the goal of preparing students to be college and career ready?

Review of Literature

Content of High-stakes Literacy Exams

The purpose of this study is to analyze changes in the content of one recently modified high-stakes literacy exam. Previous content analyses, specifically of writing items on high-stakes tests prior to CCSS, have suggested potential instructional consequences such as narrowing of curriculum and over-reliance on formulaic writing instruction. For example, an analysis of 41 high-stakes writing exam items administered across U.S. states found that these tended to include persuasive writing items and to focus on formal features in scoring (Jeffery, 2009). We could not locate studies that examined the content of reading comprehension items in U.S. states' exams, pre- or post-CCSS adoption. However, researchers in Australia, where a new accountability-based assessment initiative was implemented in 2008, examined the content of high-stakes assessments administered nationally. The researchers analyzed the cognitive complexity of test items and found that these offered "little opportunity for students to demonstrate higher-order thinking" (Pendergrast & Swain, 2013, p. 15). Though we were unable to locate similar studies for mandatory tests developed to meet accountability demands in the U.S., other research regarding large-scale reading assessments suggests that validity issues associated with these tests exist. For example, Shanahan (2014) noted that performance on standardized literacy tests cannot be consistently linked to discrete reading competencies (e.g., drawing inferences) they are designed measure. Rather, student performance has been consistently tied only to the qualities of the reading passages included. That is, the more challenging a text is for a particular reader, the less likely it is that reader can produce the desired response, regardless of the discrete reading skill that an item is supposed to measure. It follows that an analysis of high-stakes literacy exams must address the complexity of included reading passages in addition to the competencies targeted in the questions that follow these

¹ Following questions regarding the quality of these exams, the state announced that it would not renew its contract with Pearson, which expired in December 2015.

texts. As such, in this study we examine reading passages as well as selected- and constructed-response items.

Consequences of High-stakes Exams

Content analyses of high-stakes exams are especially important to undertake given research regarding their consequences. For example, research has suggested that high-stakes testing programs actually widen achievement gaps in learning opportunities (Christenson et al., 2007; Clarke et al., 2003). Laitsch (2006) and Nichols (2007) warned that in high-poverty schools, while high-stakes testing may enhance some students' learning, such tests more often lead to negative consequences, including narrowing of the curriculum and marginalization of struggling readers. Laitsch (2006) theorized that incentives, measured only by standardized tests, such as increased salaries may motivate administrators and teachers to provide "better services to low-achieving students," however he questions how responsibly these services are delivered (p. 6). For instance, administrators might add instructional time for reading or math but then cut art or social studies classes, or teachers might only focus on what is tested—which might mean creating a more "impoverished academic experience" for students (Laitsch, 2006, p. 7).

In line with this thinking, research has suggested unintended consequences of high-stakes exams, particularly on the most vulnerable students (Afflerbach, 2005). In his meta-synthesis of scholarship on this topic, Au (2007) found that standardized, high stakes tests fragmented the "structure of knowledge" and increased "teacher-centered pedagogy" (p. 263). For example, the high stakes associated with the tests have often made teachers feel pressured to teach to the middle at the "expense of the lowest performing students" (Kesler, 2013, p. 510). These findings were similar to those by Clarke et al. (2003) and Roderick and Engel (2001) who found that high-stakes tests undermined struggling students' motivation. Further research has suggested that correlations exist between high-stakes tests and dropout rates (Amrein & Berliner, 2003). Afflerbach (2005) also found that high stakes tests not only negatively impact teachers but are also limited in their capacity to accurately measure reading achievement.

While research on the effects of CCSS adoption is only beginning to be published, post-CCSS studies have further reinforced previous research regarding the effects of test-focused policy initiatives. For example, a recent study of New York's CCSS implementation examined teachers' instructional shifts in nine elementary and nine middle schools (Wilcox & Jeffery, 2016; Wilcox, Jeffery, & Gardner-Bixler, 2016), finding that teachers made substantive changes in their instructional focus in order to align to the standards, particularly with regard to their increased focus on complex informational text. The study also found that teachers felt they were incorporating less instruction in narrative genres than they had prior to CCSS, and that they believed the neglect of narrative and imaginative writing assignments was diminishing student engagement (Wilcox & Jeffery, 2016; Wilcox et al., 2016). Davis and Willson (2015) also found dissatisfaction regarding pressure to enact test-preparatory practices among teachers interviewed about the implementation of CCSS tests in Texas. Given the powerful influence of these tests on multiple stakeholders, research regarding how CCSS is influencing assessment design is needed. To that end, in this article we examine how the Common Core literacy standards were translated within one standardized test and then consider how CCSS adoption could impact teachers' instruction and students' opportunities to learn, particularly for those who are marginalized and/or struggle with literacy.

Case Study Context: CCSS in New York State

The stakes associated with CCSS adoption are particularly high in New York as it is one of 25 states where exam scores are tied to graduation requirements. While most states require only math and literacy tests, high school students in New York must pass a series of five Regents tests in order to graduate. These include one test in math, one in science, one in U.S. History, one in Global History, and one in English language arts (ELA). The high school history exams include assessments of both reading and writing, asking students to respond to “document-based questions” (DBQs) in which they incorporate multiple nonfiction source texts into extended written responses. In 2014, the state developed revised, CCSS-aligned Regents exams for math and ELA. Here we focus on the transition from the previously administered “Comprehensive Regents” ELA (COMP-ELA) to the new CCSS-aligned Regents ELA (CCSS-ELA) to examine the kinds of changes that accompanied CCSS adoption in New York.

In 2011, the state contracted the College Board to produce a technical report evaluating the alignment of CCSS to the ELA tests in grades 5, 8, and 11 (College Board, 2011). Following this report, substantive changes were made to the exit level (Grade 11) Regents ELA exam, the focus of this study. Consistent with the goals of the Common Core initiative, the new CCSS-ELA was designed to include “a noticeable change in rigor and an increased focus on text” and “more demanding and complex” items (New York State Education Department, 2014, p. v). Despite the increased difficulty of the exam, it was implemented in rapidly progressing stages (EngageNY, n.d., “Changes”). Students who entered high school in August 2013 must pass the new CCSS-ELA Regents in order to graduate. As is the case elsewhere in the U.S., rapid CCSS implementation has created pressures for teachers and school leaders, as results from administration of the new exams have been made public. Due to this sharp increase in exam difficulty for a large number of students, it is important to investigate how the COMP-ELA Regents has been adapted to align with CCSS and what kinds of challenges such adaptations might present for students and teachers. Further, this case study can provide crucial insight regarding the implications of CCSS adoption in other U.S. states.

Methods

Data sources for this case study included all items from 10 exams: eight Regents exams and two sample AP exams. We analyzed four COMP-ELA tests and four CCSS-ELA tests that were administered between June 2014 and June 2015 and two sample AP English exams, one Language and Composition and one Literature and Composition test (Table 1). Unlike the Regents, the College Board does not release AP tests to the public; only one sample of each test is provided on their website. We used the AP exams for two reasons: One, these exams are designed to measure college equivalence rather than readiness; and two, the College Board, which designs AP exams and also issued the 2011 report on Regents alignment to CCSS, may have influenced the revisions made to the COMP-ELA. (As an example, the College Board (2011) report criticized the New York Regents speaking and listening items, explaining, “discussion and presentation skills might be more holistically and authentically assessed through in-class performance and observation” (p. 4). The revised Common Core Regents ELA exam later removed the listening portion altogether.) We were interested in how the English AP exams might compare to the two ELA Regents exams with respect to CCSS alignment and text complexity, as we further explain below.

We comparatively analyzed the exams for two qualities: text complexity and correspondence to CCSS. For our first measure of text complexity, each reading passage from the 10 included exams was measured quantitatively based on a Lexile score. We obtained

Table 1: *Sources and Methods*

COMP-ELA	CCSS-ELA	AP Samples	Methods of Analysis
June 2015	June 2015	Language and Composition	Lexile, complexity rubric, item coding
January 2015	January 2015	AP Literature and Composition	Lexile, complexity rubric, item coding
August 2014	August 2014		Lexile, item coding
June 2014	June 2014		Lexile, item coding

Lexile levels by running passages through a tool available on Lexile.com, which is widely used by school districts and curriculum and assessment developers. Lexile calculates an algorithm based on variables such as “words per sentence, the average number of syllables per sentence, and whether or not the words appear on a given list” such as the Dale-Chall readability formula (Fisher & Frey, 2014, p. 237).

We also qualitatively analyzed reading passages for text complexity, as the quantitative measure of Lexile does not take into account developmental concerns, reader interest, and other textual variables. Specifically, we analyzed 15 reading passages from the exams: Seven passages from the two sample AP exams and eight passages from the June 2015 COMP-ELA and June 2015 CCSS-ELA, as these were the most recently administered of each Regents exam at the time of data analysis. Using an abbreviated version of Fisher and Frey’s (2014) rubric for analyzing text complexity (see Appendix A), we looked at levels of meaning and purpose (i.e. density, complexity, and figurative language), structure (i.e. genre, organization, and narration), language conventionality and clarity (i.e. standard English, variations, and register), and knowledge demands (i.e. background, prior, cultural, and vocabulary knowledge). This analysis was critical in that readers are equal contributors to the transaction of making meaning—and their experiences with texts are impacted by these elements (as is the context in which students are reading). The rubric included three qualitative labels used for analysis: Texts that would *stretch* the reader, texts that require *grade-appropriate* skills and texts that are within a *comfortable* range for the reader. To obtain an average score for each category using this rubric, using a percentage of agreement, we analyzed each text separately (with an interrater reliability of .94) and then resolved discrepancies through discussion.

We also conducted an analysis of selected and constructed response items for all 10 exams to examine how the items corresponded to CCSS for reading and writing. We first generated code lists from the CCSS ELA Anchor Standards for Reading and Writing College and Career Readiness. We used the Anchor Standards for reading so as to encompass both informational text and literature. Table 2 demonstrates the overlap between the language of the Anchor Standards for college and career readiness and the Reading Standards for Grades 11-12. We used numerical codes to identify each standard. For example, if the question asked students to make an inference we coded the prompt as “R1”; if the prompt asked the students to write an argument, we coded the task as “W1”. We each separately applied these codes to the reading exam items, with an inter-rater reliability of .86. We also applied writing codes to the rubrics used to assess writing on the constructed-response items. We then compared results of the item analyses across exams to identify overarching patterns of variation.

Table 2: *Overlap of Language between Anchor Standards and Grade Level 11-12 Standards for Reading*

<p>ELA Anchor Standards for College and Career Readiness</p>	<p>ELA Common Core Reading Literature Standards: Grades 11-12</p>	<p>ELA Common Core Reading Informational Text Standards: Grades 11-12</p>
<p>CCSS.ELA-Literacy.CCRA.R1. Read closely to determine what the text says explicitly and to make logical inferences from it; cite specific textual evidence when writing or speaking to support conclusions drawn from the text.</p>	<p>CCSS.ELA-Literacy.RL.11-12.1. Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text, including determining where the text leaves matters uncertain.</p>	<p>CCSS.ELA-Literacy.RI.11-12.1. Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text, including determining where the text leaves matters uncertain.</p>
<p>CCSS.ELA-Literacy.CCRA.R2. Determine central ideas or themes of a text and analyze their development; summarize the key supporting details and ideas.</p>	<p>CCSS.ELA-Literacy.RL.11-12.2. Determine two or more themes or central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to produce a complex account; provide an objective summary of the text.</p>	<p>CCSS.ELA-Literacy.RI.11-12.2. Determine two or more central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to provide a complex analysis; provide an objective summary of the text.</p>
<p>CCSS.ELA-Literacy.CCRA.R3. Analyze how and why individuals, events, or ideas develop and interact over the course of a text.</p>	<p>CCSS.ELA-Literacy.RL.11-12.3. Analyze the impact of the author’s choices regarding how to develop and relate elements of a story or drama (e.g., where a story is set, how the action is ordered, how the characters are introduced and developed).</p>	<p>CCSS.ELA-Literacy.RI.11-12.3. Analyze a complex set of ideas or sequence of events and explain how specific individuals, ideas, or events interact and develop over the course of the text.</p>
<p>CCSS.ELA-Literacy.CCRA.R4. Interpret words and phrases as they are used in a text, including determining technical, connotative, and figurative meanings, and analyze how specific</p>	<p>CCSS.ELA-Literacy.RL.11-12.4. Determine the meaning of words and phrases as they are used in the text, including figurative and connotative meanings; analyze the impact of specific word</p>	<p>CCSS.ELA-Literacy.RI.11-12.4. Determine the meaning of words and phrases as they are used in a text, including figurative, connotative, and technical meanings; analyze how an author</p>

Continued on next page

Table 2: *Continued*

ELA Anchor Standards for College and Career Readiness	ELA Common Core Reading Literature Standards: Grades 11-12	ELA Common Core Reading Informational Text Standards: Grades 11-12
word choices shape meaning or tone.	choices on meaning and tone, including words with multiple meanings or language that is particularly fresh, engaging, or beautiful. (Include Shakespeare as well as other authors.)	uses and refines the meaning of a key term or terms over the course of a text (e.g., how Madison defines faction in Federalist No. 10).
CCSS.ELA-Literacy.CCRA.R5. Analyze the structure of texts, including how specific sentences, paragraphs, and larger portions of the text (e.g., a section, chapter, scene, or stanza) relate to each other and the whole.	CCSS.ELA-Literacy.RL.11-12.5. Analyze how an author’s choices concerning how to structure specific parts of a text (e.g., the choice of where to begin or end a story, the choice to provide a comedic or tragic resolution) contribute to its overall structure and meaning as well as its aesthetic impact.	CCSS.ELA-Literacy.RI.11-12.5. Analyze and evaluate the effectiveness of the structure an author uses in his or her exposition or argument, including whether the structure makes points clear, convincing, and engaging.
CCSS.ELA-Literacy.CCRA.R6. Assess how point of view or purpose shapes the content and style of a text.	CCSS.ELA-Literacy.RL.11-12.6. Analyze a case in which grasping a point of view requires distinguishing what is directly stated in a text from what is really meant (e.g., satire, sarcasm, irony, or understatement).	CCSS.ELA-Literacy.RI.11-12.6. Determine an author’s point of view or purpose in a text in which the rhetoric is particularly effective, analyzing how style and content contribute to the power, persuasiveness or beauty of the text.
CCSS.ELA-Literacy.CCRA.R7. Integrate and evaluate content presented in diverse media and formats, including visually and quantitatively, as well as in words.	CCSS.ELA-Literacy.RL.11-12.7. Analyze multiple interpretations of a story, drama, or poem (e.g., recorded or live production of a play or recorded novel or poetry), evaluating how each version interprets the source text. (Include at least one play by Shakespeare and one play	CCSS.ELA-Literacy.RI.11-12.7. Integrate and evaluate multiple sources of information presented in different media or formats (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem.

Continued on next page

Table 2: *Continued*

ELA Anchor Standards for College and Career Readiness	ELA Common Core Reading Literature Standards: Grades 11-12	ELA Common Core Reading Informational Text Standards: Grades 11-12
	by an American dramatist.)	
CCSS.ELA-Literacy.CCRA.R8. Delineate and evaluate the argument and specific claims in a text, including the validity of the reasoning as well as the relevance and sufficiency of the evidence.	(RL.11-12.8 not applicable to literature)	CCSS.ELA-Literacy.RI.11-12.8. Delineate and evaluate the reasoning in seminal U.S. texts, including the application of constitutional principles and use of legal reasoning (e.g., in U.S. Supreme Court majority opinions and dissents) and the premises, purposes, and arguments in works of public advocacy (e.g., <i>The Federalist</i> , presidential addresses).
CCSS.ELA-Literacy.CCRA.R9. Analyze how two or more texts address similar themes or topics in order to build knowledge or to compare the approaches the authors take.	CCSS.ELA-Literacy.RL.11-12.9. Demonstrate knowledge of eighteenth-, nineteenth- and early-twentieth-century foundational works of American literature, including how two or more texts from the same period treat similar themes or topics.	CCSS.ELA-Literacy.RI.11-12.9. Analyze seventeenth-, eighteenth-, and nineteenth-century foundational U.S. documents of historical and literary significance (including The Declaration of Independence, the Preamble to the Constitution, the Bill of Rights, and Lincoln’s Second Inaugural Address) for their themes, purposes, and rhetorical features.
CCSS.ELA-Literacy.CCRA.R10. Read and comprehend complex literary and informational texts independently and proficiently.	CCSS.ELA-Literacy.RL.11-12.10. By the end of grade 11, read and comprehend literature, including stories, dramas, and poems, in the grades 11-CCR text complexity band proficiently, with scaffolding as needed at the high end of the range.	CCSS.ELA-Literacy.RI.11-12.10. By the end of grade 11, read and comprehend literary nonfiction in the grades 11-CCR text complexity band proficiently, with scaffolding as needed at the high end of the range.

Continued on next page

Table 2: *Continued*

ELA Anchor Standards for College and Career Readiness	ELA Common Core Reading Literature Standards: Grades 11-12	ELA Common Core Reading Informational Text Standards: Grades 11-12
	By the end of grade 12, read and comprehend literature, including stories, dramas, and poems, at the high end of the grades 11-CCR text complexity band independently and proficiently.	By the end of grade 12, read and comprehend literary nonfiction at the high end of the grades 11-CCR text complexity band independently and proficiently.

Previous research has raised questions regarding the reliability of item analyses with regard to the use of expert raters (Herman, Webb, & Zuniga, 2005). However, such research often relies on raters with limited or no teacher training or teaching experience. For both qualitative analytic procedures (text complexity and CCSS correspondence), we based our scoring decisions on our extensive knowledge of adolescent literacy development and interactions with high school teachers and students. The lead author has taught ELA in public high schools for 20 years and currently holds a position as a reading specialist and literacy coach in a New York City high school. The second author has taught ELA, including Advanced Placement, in public high schools for seven years. In addition, for the past 12 years, both authors have conducted research in adolescent literacy and have mentored and supervised pre-service and in-service ELA teachers. Our experience working in public secondary schools includes teaching and mentoring teachers in both urban and suburban settings. Having only two evaluators is certainly a limitation of this research; however, as this was a case study, this work is meant to spark larger, more resourced studies to explore the reliability, accessibility, and validity of standardized tests.

Results

Comparison of Regents Exams

Here we provide a comparison of the COMP-ELA and the CCSS-ELA in response to our first research question regarding how the Regents exams have been adapted to align with CCSS. We found that the two Regents exams differed substantively with regard to listening, reading, and writing requirements. The COMP-ELA includes four sections: In the first section, students listen to one informational passage twice and then answer eight multiple choice questions. The second section provides students with two reading passages (one fiction and one informational); students answer 12 multiple-choice questions. The third section has two reading passages (usually one poem and one narrative); students answer five multiple-choice questions and write two text-based paragraphs. For one paragraph students are given a “controlling idea” and they must use evidence from both passages to support a theme. In the second paragraph, students select one passage and discuss how the author uses a literary element or technique. Additionally, students must write an extended (i.e. multi-paragraph) response, a “critical lens essay,” in which they read a quotation and then select two texts

they have previously read to either support or refute that quotation. An example critical lens quotation from August 2014 is “ignorance is never better than knowledge.”

In contrast, the CCSS-ELA does not include short constructed-response items and does not invite students to draw from knowledge and experience with texts they have read previously. The CCSS-ELA instead requires students to respond only to passages within the test itself. Further, the CCSS-ELA does not assess listening skills. The first section includes three reading passages with 24 multiple-choice questions. The second section features four nonfiction passages and students are to develop an argumentative essay based on a focus question. An example question from the August 2014 exam is “should the United States bid to host a future Olympic Games?”. The final section asks students to read a passage and write a “textual analysis response,” in which they must explain the author’s use of rhetorical devices or literary elements and techniques.

In comparing the two versions of the Regents, there is a clear increase in reading demands for students within the CCSS-ELA. The length of the test went on average from 10 pages for the COMP-ELA to 18 pages for the CCSS-ELA. Further, as noted above, the COMP-ELA provides four passages (two narratives, one informational, and one poem) while CCSS-ELA has eight passages (one narrative, one poem, two non-fiction, and four connected informational pieces to be read and synthesized for an argumentative essay). Word length for the exams increased from an average of 3,972 words (range 3,968 to 3,978) for the COMP-ELA to an average of 8,547 words (range 8,282 to 8,783) for the CCSS-ELA. Despite the doubling of exam content, students are allotted the same amount of time: three hours. Analyzing specific passages, the average word length and Lexile level also increased substantially. For the COMP-ELA, the average words per passage for all four exams was 515 words (range 509 to 532), and for the CCSS-ELA, 752 words (range 734 to 786). Regarding the Lexile levels for exam passages, we found that for the COMP-ELA, the average was 1138L with a range of 710L to 1380L. For the CCSS-ELA the average Lexile was 1279L with a range of 1010L to 1650. Table 3 represents averages for both Regents and AP exams, the latter of which are discussed in greater detail below.

With respect to the qualitative analysis of text complexity, we analyzed the two ELA Regents exams from June 2015. Overall, by using Fisher and Frey’s (2014) analytical tool for textual complexity, we found that both Regents exams remained within a grade-appropriate level. The first complexity feature focused on levels of meaning or purpose, or whether a text provided an explicit and unambiguous purpose. The COMP-ELA scored higher in qualitative measures when considering purpose. In all

Table 3: *Exam Length and Complexity Measures*

Exam	Average words per passage	Average Lexile and range per passage	Time allocation
Comprehensive ELA Regents	515 words	1138L (range 710L-1380L)	3 hours
CCSS-Aligned ELA Regents	752 words	1279L (range 1010L-1650L)	3 hours
AP Language and Composition	585 words	1411L (range 990L-1780L)	3 hours and 15 minutes
AP Literature and Composition	500 words	1176L (range 910L-1550L)	3 hours

four passages, the purpose was withheld from the reader and students needed interpretative skills in order to identify what that purpose was. This result is due to the fact that COMP-ELA emphasized literary texts, one essential feature of which is implicit or ambiguous meaning. Because the four passages of the eight in CCCS-ELA were informational (with a purpose clearly stated)—the exam was scored, on average, lower in the category of meaning and purpose. Second, regarding text structure (e.g., genre, organization and narration), both exams were within either the grade-appropriate or comfort range for students. Similarly, language conventionality and clarity (e.g., standard English, variations, and register) were found to be at grade-appropriate and comfort levels for both exams. As we discussed results from our independent analyses, we also noted salient structural elements of the COMP-ELA that may aid students' comprehension. For example, white space, which is a feature of text that allows easier processing by the reader, was utilized to a greater extent on the COMP-ELA than on the CCSS-ELA. Further, the passages within the COMP-ELA, when compared to CCSS-aligned Regents, had much more "movement," including action and dialogue in the fictional texts that might make them more engaging, and thus more accessible, to adolescent readers. Another observation was that both Regents exams often provided excerpted texts that occur in the middle of a larger text.

The category of knowledge demands was the most difficult to analyze as this factor depends on the readers of the text. The knowledge demands include students' background, prior, cultural, and vocabulary knowledge. Based on our extensive experience and expertise with high school students, we generated informed analyses of the passages with regard to knowledge demands. We found that CCSS-ELA passages corresponded to higher levels of complexity on the qualitative rubric than the COMP-ELA when considering demands of background, cultural, and prior content knowledge. Notably, on the CCSS-ELA exam some passages might be particularly challenging for English language learners and/or for students who do not come from middle-class or privileged environments. For example, the June 2015 CCSS-ELA included a passage from Edith Wharton's *Age of Innocence* about a group of people who are conversing in a parlor at a lush dinner party with "some of 1870s New York aristocracy." The passage references "Neopolitan love-songs" and dukes and duchesses—references to which many adolescents will not have access. The same applies for vocabulary demands. The CCSS-ELA was more demanding, as evidenced in the Wharton reading passage, which included such words as "precocious", "cordiality", "solemn", "formidable", and "amiable". The glossing of text, which in the case of the Wharton passage included eight definitions provided as endnotes, does little to fill these gaps. For instance, "ducal" is defined as "relating to a duke." In comparison, the COMP-ELA had lower levels of knowledge demands such that students might be more likely to use their prior knowledge to aid in their comprehension of the passages. The June 2015 COMP-ELA, for example, included a much more recent, shorter passage from *Lonesome Dove* for which only two words were glossed. According to Webb (2007, p. 9), this would be a "source-of-challenge" issue where the problem with the item is that the student does not have the background knowledge in order to answer items correctly as opposed to the skills being assessed.

When analyzing passages for knowledge demands, each of us independently noted a mismatch between the reading passages in both tests and the interests and cultural backgrounds of diverse learners served in New York schools. Despite the statement that "the State Education Department . . . ensures that the diversity of New York State students is represented in the test development process" (New York State Education Department, 2013), we noted a lack of representation culturally—with predominantly

white and male authors. This observation led us to analyze the authors of 32 texts in both Regents exams (excluding the informational text set authors for the argumentative writing in CCSS-ELA which were mostly from newspapers, magazines, and websites). We found that of these authors, 9 were women and 23 were men. Of these 32 authors, 27 were white, two were Asian, one was black, one was Native American, and one was of Middle Eastern descent.

We also analyzed test items for correspondence to the Common Core Anchor Standards. In terms of the selected response items, the tests varied greatly as to what standards were implicated (Table 4). Overall, we found that the CCSS-ELA had more equal distribution when looking at the first six College and Career Readiness Anchor (CCRA) reading standards. In contrast, the COMP-ELA placed far more emphasis on CCRA Reading Standard 1 (inferences and citation of textual evidence). Another notable difference is that the CCSS-ELA had over twice the number of questions that focused on CCRA.R5, which requires students to analyze the structure of texts. The same is true for CCRA.R6, which requires students to analyze the point of view or purpose of the passage. Finally, the COMP-ELA included only one question pertaining to CCRA.R2 (central idea/theme and summary) while CCSS-ELA had 10% of its questions focusing on this skill.

There were some notable commonalities between the tests, though, in that neither exam's items aligned to CCRA.R7 (integration and evaluation of content in diverse media and formats). Further, in all eight exams analyzed, CCRA.R8 (evaluation of arguments, claims, and reasoning) is only implicated in one COMP-ELA exam question and one CCSS-ELA exam question. Considering the knowledge and skills that are needed for 21st century literacies, these results represent considerable gaps.

In analyzing the writing tasks and the rubrics for such tasks, we again found variety in what was assessed for the writing and what prior knowledge was needed from students. The biggest difference we found was that the COMP-ELA allows for student choice, in that the critical lens essay invites students to select any texts they have read in the past in order to support their argument about the prompt's quotation. Another major difference between the COMP-ELA and CCSS-ELA is that the CCSS-ELA asks students to construct a source-based essay based on provided passages as opposed to previously read texts. As we discuss further below, writing to informational sources is already assessed on New York's two history Regents exams (Global and U.S. History). It should be noted, however, that the topics for CCSS-ELA nonfiction texts were, in our assessment, fairly accessible to students. These included tracking consumers' shopping preferences, hosting of future Olympic games, bringing extinct species back into existence, and most recently, in June 2015, paying college athletes. Overall, both the COMP-ELA and the CCSS-ELA only assessed the following three Anchor Standards for writing: writing an argument (CCRA.W.1), writing an informative/explanatory text (CCRA.W.2), and writing for coherence and organization (CCRA.W.4).

Comparison of Regents and Advanced Placement Exams

We also explored how Regents "college and career readiness" exams compared to Advanced Placement (AP) early college equivalence tests in response to our second research question. Like the Regents, students are given three hours for the AP exams, although the Language and Composition AP exam provides an additional 15 minutes. The sample Literature and Composition AP test—at a total of 6,443 words—exceeded the word count of the COMP-ELA but was below the CCSS-ELA word count. This exam also had fewer passages than the CCSS-ELA (six compared to eight) and more when

Table 4: *Alignment to Anchor Standards for Reading in Selected-response Items*

Anchor Standard for Reading	Comprehensive Regents (N = 100)	CCSS-aligned Regents (N = 96)	AP Language Composition Sample Test (N = 50)	AP Literature Composition Sample Test (N = 46)
CCSS.ELA-Literacy.CCRA.R.1 (inferences and citation of textual evidence)	49% (49) Range 9-16 questions per test	15% (14) Range 1-6 questions per test	26% (13)	46% (21)
CCSS.ELA-Literacy.CCRA.R.2 (central idea/theme and summary)	1% (1)	10% (10) Range 1-4 questions per test	0	2% (1)
CCSS.ELA-Literacy.CCRA.R.3 (analysis)	12% (12) Range 2-4 questions per test	18% (17) Range 3-5 questions per test	18% (9)	4% (2)
CCSS.ELA-Literacy.CCRA.R.4 (interpretation and analysis of words/phrases)	18% (18) Range 1-7 questions per test	16% (15) Range 2-5 questions per test	6% (3)	24% (11)
CCSS.ELA-Literacy.CCRA.R.5 (analysis of the structure of texts)	10% (10) Range 1-4 questions per test	22% (21) Range 3-8 questions per test	42% (21)	17% (8)
CCSS.ELA-Literacy.CCRA.R.6 (point of view or purpose)	8% (8) Range 0-3 questions per test	17% (16) Range 2-6 questions per test	8% (4)	7% (3)
CCSS.ELA-Literacy.CCRA.R.8 (evaluation of arguments, claims, and reasoning)	1% (1)	1% (1)	0	0
Did not fit a standard	1% (1)	0	0	0
Unresolved coding	0	2% (2)	0	0

N equals the number of testing items. For example, we analyzed 100 testing items on the Comprehensive Regents exams. We then coded each of these items with a reading standard. For instance, R1 was the code given to the testing items that measured the first Common Core Anchor Standard for reading (inferences and/or citation of textual evidence). 49 testing items were given this code out of 100, giving the percentage of 49% (49/100).

compared to the COMP-ELA (four). The Language and Composition sample test had more words and passages than both Regents exams at 9,696 words and 11 passages. As to word length and Lexile level, the average word length for the AP Language and Composition exam reading passages was 585 words and for the Literature and Composition 500 words. These numbers are similar to COMP-ELA reading passages, which averaged 515 words. The CCSS-ELA passages were longer than both AP exams with an average of 752 words per passage.

Regarding the Lexile levels, the AP Language and Composition test had the highest average at 1411L, while the AP Literature and Composition test averaged 1176L per passage (see Table 3). This Lexile level is almost 100 points *lower* than CCSS-ELA. According to the supplementary document to “Appendix A” of the Common Core State Standards, the CCSS Lexile band for 11th to 12th grade “Common Core Readiness” (CCR) is 1185L-1385L (NGACBP & CCSS, 2015). While the COMP-ELA misses this mark by nearly 50 Lexile points, so does the AP Literature and Composition, which is designed to assess early college equivalency for high school students. The CCSS-ELA seems to be more in alignment with college equivalency, not readiness, with an average Lexile that exceeds the 11-CCR Lexile band.

We also coded all passages from the sample AP exams for qualitative aspects of text complexity. With respect to structure and organization, the AP exams were similar to both of the Regents exams in remaining within appropriate complexity levels. However, with respect to levels of meaning and purpose, AP tests were scored higher than both Regents tests in terms of density, complexity, and figurative language. Similarly, the AP tests were located in the stretch range in terms of register (i.e. the register was either archaic, formal, domain-specific, or scholarly). AP tests also had higher scores when considering demands of background, cultural, and prior knowledge. The same applied for vocabulary demands.

With respect to alignment with CCSS Anchor Standards, we found that the AP Literature and Composition sample test was similar to the COMP-ELA in its focus on inferences and textual evidence, with 46% of its questions focusing on CCRA.R.1 (the COMP-ELA was at 49%). While the AP exams resembled the CCSS-ELA in an emphasis on source integration, they also resembled the COMP-ELA in inviting students to draw from prior knowledge in extended written responses. For example, like the “critical lens” essay in the COMP-ELA, the AP Literature and Composition test provides students with a prompt (i.e. “how cruelty functions in the work and what cruelty reveals about the perpetrator and/or victim”) where they are required to draw evidence from texts they have read previously in support of their argument.

What did the all three tests have in common? What they did not cover. Neither the Regents nor the AP tests assessed students’ abilities to plan, revise, and/or edit (CCRA.W.5). None asked for students to be able to use technology to write (CCRA.W.6), a critical skill for 21st century literacies. The Common Core Writing Anchor Standards also ask that students be able to conduct research from print and digital sources (CCRA.W.7-9), yet this skill is not assessed on any of the tests. Further, again, the CCRA.R.8 standard for reading requires that students be able to evaluate arguments, claims and reasoning—yet this is not evident on any of the four exams. Additionally, none of the tests asked students to write narratives (CCRA.W.2), as all prioritized argumentative and expository writing.

Discussion

In response to our third research question, in this section, we examine implications for policy makers and curriculum and assessment developers. Given the impact of standardized tests on instruction, and, consequently, the academic experiences of students across the U.S., questions regarding what types of assessments are adopted, and who is invited to the table in the design process are crucial to examine. Our current educational environment is tumultuous with constant fluctuations and decisions being made regarding the standards themselves and the subsequent tests that are developed (and revised). Because of these rapid changes, it is critical to stop and reflect before making more decisions, without careful analyses of what is changing and how these changes impact students and teachers. We know these tests drive instruction and are gatekeepers for our students; therefore, we offer suggestions here for policy makers and other stakeholders who make decisions about design. We also offer these suggestions for high school teachers, parents, and administrators who can advocate for such changes.

College Readiness versus Equivalence

First, we need to consider the rigor of the standardized tests: Are they more representative of college readiness or equivalence? While we do not object to the goal of raising expectations for students, the results of our comparison of the Regents and AP exams led us to wonder if the new CCSS-ELA more closely resembled a measure of early college equivalence than of college readiness. And, if so, we also wondered why the CCSS-ELA exceeded even the AP tests in the amount of reading required without allowing additional time. If a shorter and more manageable test is sufficient to assess whether students might receive college credit for introductory English courses, we wondered why it was necessary for a test of college and career readiness to be so lengthy.

Additionally, the exceptional increase in length and complexity of the CCSS-ELA, which exceeded even the AP exam, suggests to us that, though it clearly corresponded with a wider range of CCSS standards than did the COMP-ELA, the designers placed a skewed emphasis on making a “more rigorous” exam as opposed to one that was better aligned to CCSS. That is, we conclude that it would have been possible to bring the COMP-ELA into closer alignment with the CCSS by using the same text types and formats but incorporating items that correspond to a more balanced cross-section of the standards. Specifically, we found that COMP-ELA emphasized inference and citation of evidence to a much greater extent than did CCSS-ELA, which placed greater emphasis on structural analysis and analysis of author’s purpose. The exams could be revised to achieve greater balance while maintaining appropriate levels of text complexity and exam length given the time constraints. We hope that, as policymakers again consider revisions to the exams in response to recent controversy over their implementation in New York (see State Education Commissioner’s comments in McMahon, 2015), they will take such recommendations into consideration.

Inclusivity and Prior Knowledge

We also recommend the selection of more diverse, contemporary texts that more accurately represent students’ cultural backgrounds and experiences. Despite New York State’s Education Department’s (2013) statement on ensuring diversity within the test development process, overall the Regents tests examined here demonstrate a lack of representation culturally, privileging texts by white authors over those of color. Of the passages analyzed for this study, 84% were by white authors. Yet, during the 2014-2015 academic year for New York City schools, the ethnic distribution was 40%

Hispanic, 28% African American, 15% Asian, 15% White, and 2% Other. The homepage of the CCSS Initiative makes claims that the standards are “relevant to the real world,” yet assessments that followed their implementation are not representative of students in New York City (Brooks & Dietz, 2012/2013) and elsewhere in the U.S. Thus, a shift to more cultural diversity in authorship of reading passages could help to level the playing field. We know that our students come to these tests with their own cultural lenses and experiences that affect their comprehension and analysis; thus, the relationship between texts and the experiences of readers matter (Hiebert & Sluys, 2014).

Regarding students’ prior knowledge, we were also concerned with the removal of the “critical lens” essay, which in the COMP-ELA invited students to share their own diverse reading experiences—to incorporate the texts with which they felt success. This revision is consistent with the CCSS emphasis on text and purpose at the expense of an emphasis on the reader—a shift in emphasis that prioritizes “text-based evidence over other sources of evidence that are equally justifiable” (Snow & O’Connor, 2013, p. 3). Dutro, Selland, and Bien (2013) contend that it is “both theoretically and ethically imperative to understand the social, cultural, and intellectual resources students bring to writing, the unique challenges they face, and the competencies they exhibit that an on-demand test can miss” (p. 133). Accordingly, future revisions to the CCSS-ELA should, in addition to requiring students to draw on evidence from text, also invite students to draw on reasoned judgments, social norms, prior knowledge, and personal experience. Doing so has the potential to engage students by showing them that we value their perspectives and experiences; incorporating varied forms of evidence is also what skilled argumentation requires. Further, given that studies have shown that students from the United States rank the lowest of internationally in terms of their interest in reading (Mullis et al., 2003), we question the rationale for removing an item that provides students an opportunity to write about reading with which they’ve felt engaged, as the “critical lens” task did.

Additionally, we ask high school test developers to consider more robustly the qualitative elements of the texts that are selected. While quantitative measures are important to consider when identifying appropriate text passages, qualitative analysis needs to be given equal attention (Fisher & Frey, 2014). Much research has arisen regarding the problematic nature of only using quantitative measures when selecting passages (Heibert & Sluys, 2014; Pearson & Hiebert, 2013). For example, in the June 2015 CCSS-ELA, the first passage (from *Age of Innocence*) has a Lexile level of 1590, which is over 200 points higher than the Common Core Lexile “Stretch Band” for 11th and 12th graders (1185-1385L). More so, the qualitative analysis demonstrates that the texts would stretch the reader or require instruction based on density and complexity, figurative language, purpose, organization, narration, register, and background, prior, cultural, and vocabulary knowledge—all text features that cannot be accurately captured with quantitative measures.

Narrowing the Curriculum

Finally, given research regarding the relationship between what gets tested and what gets taught (McCarthy, 2008), we recommend that policymakers and high school test designers consider a more balanced approach to assessment of the CCSS, so as to avoid, as much as possible, the narrowing of the curriculum associated with high-stakes testing. For example, we found gaps in these tests when considering correspondence to the Common Core writing standards, which include writing for diverse purposes, planning and revising, and writing for real audiences. While we agree that argumentation is an important skillset to master for academic literacy development,

we worry about a dearth of text diversity represented in the exams. In only offering expository and argumentative writing, we risk denying students opportunities to draw from their personal connections and backgrounds as well from creative and social imagination (Eppley, 2015). Destigter (2015) warns against an overemphasis on argument, which limits what “counts” as valid thought, explaining that we need to “foster diverse forms of [students’] expression and honor the countless reasons why we choose—or need—to write” (p. 31). Students should write for a variety of purposes, including for informing and arguing, but also for making meaning (e.g., imaginative narratives, memoirs), for professional communication (e.g., work memos, emails), for civic responsibility (e.g., editorials, public service announcements), and for personal and emotional growth (e.g., journals, poetry). However, one important question this issue raises is: How can we assure genre diversity without greatly expanding the number of items included on a literacy assessment? In the case of New York, we suggest eliminating the document-based argument from the CCSS-ELA since, as noted above, this type of writing is already assessed in two New York Regents exams for history. Given the multiple, discipline-specific exams that students must take in New York, we suggest that a further revised CCSS-ELA test might instead offer space for constructed response items that call for narrative, descriptive, and/or professional writing. Another possibility is the design of a test that varies task types (e.g., students may encounter a narrative, argumentative, or expository writing task depending on the test administration).

Another gap in all of the exams, CCSS-ELA, COMP-ELA, and the AP tests, is that of critical evaluation, despite the fact that the Common Core Anchor Standards call for students to evaluate diverse media (CCRA.R7) and arguments, claims, and reasoning (CCRA.R8). Students are required to make an argument—but not to evaluate the arguments of others. This is an essential skill, since on a daily basis students are confronted by media images, potentially false information, and advertisements. If college and career ready means that students should be able to evaluate texts, the media, and internet content, then exams should include items that allow for this kind of critical engagement with text. Similarly, the exams also fail to incorporate “diverse media and formats” (CCRA.R7), which means they do not address 21st century literacies (Brooks & Dietz, 2012/2013).

Additionally, despite the fact that speaking and listening are included in the Common Core standards, the new CCSS-ELA eliminates the listening section that was included on the COMP-ELA. We would encourage that CCSS-ELA designers to reconsider this decision, given that speaking and listening are important skills for academic success (e.g., participation in discussion, taking notes in lectures) as well as for civic participation.

Limitations

For this case study, we would like to address limitations and possibilities for future research that would expand our research. First, only two raters, the authors, analyzed test items, as opposed to a team of raters. Thus, we recommend replication of this study for both the Regents and other state standardized tests. We also used just one measure for text complexity, the Lexile score, while using other quantitative measurements that are available may have provided varied results. Accordingly, we do hope that this work will spark larger, more resourced studies to explore the reliability, accessibility, and validity of standardized tests. Such research is needed given what we know about the impact of high-stakes assessments on students’—particularly underserved and marginalized students’—opportunities to learn.

Conclusion

While this case study is representative of only one state, New York City does have the largest school district in the nation, serving over one million students. Regardless, this analysis provides an example of how high school standardized tests are imperfectly aligned with the espoused goals of the Common Core Standards. As Rothman (2014) notes, “When there is discrepancy between tests and standards, teachers tend to place a greater emphasis on what is tested” (p. 20). Authorities in New York have recently signaled, largely in response to pressure from parents, teachers, and school leaders, that they will revise the tests, so the Regents ELA is likely to morph yet again in the near future (McMahon, 2015). Nonetheless, it is important to document such cases since enormous resources have been poured into the CCSS shift in New York and across the United States. Let us learn from the mistakes of the past, not continue to replicate them.

Thusly, we ask that test designers and school districts consider approaches to construction of standardized tests that are focused not only on increasing “rigor” but also on being responsive to students’ varied experiences and needs. This would include consideration of the cultural and personal experiences students bring to literacy tasks. We also ask that educational researchers, teacher educators, parents, students, and teachers advocate for such changes. What can this look like? It means conducting more analyses such as these with other state standardized tests. If such tests are not available to the public, it means writing our senators, our school board members, and our Congress members about being more explicit and transparent in the ways we are testing young people and how that is impacting our curriculum. It means conducting more research on the direct impact of such tests on instruction, curriculum, and our youth. It also means writing letters to our editors and informing the general public about the impact of standardized testing. In fact, we emailed this manuscript twice to the New York State Department of Education and are currently awaiting a response for further conversations and collaboration.

We want our tests to reflect the types of learning that are needed for students’ success in college and career, and tests that support the kinds of instruction that students need to achieve that success. Tests clearly should be developed and/or chosen that address the Common Core Standards; however, they should also allow for greater depth and diversity. We must begin to demand evidence that these assessments actually “measure college and career readiness” that is the goal of the Common Core Standards (Chingos, 2013, p. 15). Currently, the CCSS-ELA has not produced any evidence that it is an accurate predictor of college readiness and in fact the results of this study demonstrate it is a more accurate representation of college placement—which is not the objective of the Common Core Standards. If this is the goal for the New York State Regents, then we suggest that more educational stakeholders begin to collect data on students with passing scores and compare that to their success in college. Otherwise, these tests are not measuring what they purport to measure. If in fact the purpose of the Regents test is to guide teachers’ instruction, then the feedback should be delivered in a clearer, more explicit and specific way. Currently, teachers and students receive one numerical score (on a scale of 0-100) without any information regarding the kinds of skills high school students need to improve. A sound assessment is one that is informative to teachers, students, and parents.

In sum, the reality is that our classroom instruction will in fact change based on the new iterations of standardized tests. As Davis and Willson (2015) found in their study about the impact of standardized testing on teaching, “Instead of instructional

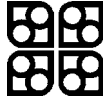
practices bending to align with the test, we see the test being allowed to enlarge and encircle all aspects of instructional practice” (p. 374). We know the tests will substantially influence our curriculum and instruction (Sundeen, 2015). But at what cost? Educational practitioners and activists Burris and Murphy (2014) predict that for the Class of 2022 in New York State, the graduation rates could drop to as low as 25 percent under the current implementation plan. Based on the subsequent analysis, overall, we ask test designers to be more thoughtful and reflective when creating such high-stakes tests. Similarly, we ask that district and school leaders, parents, and teachers be more critical about what is being asked of students and begin to advocate for who matters most: our students.

References

- Afflerbach, P. (2005). National Reading Conference policy brief: High stakes testing and reading assessment. *Journal of Literacy Research, 37*(2), 151–162.
- Amrein, A. L. & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32–38.
- Applebee, A. N. (2013). Common Core State Standards: The promise and the peril in a national palimpsest. *English Journal, 103*(1), 25–33.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267.
- Aull, L. (2015). Connecting writing and language in assessment: Examining style, tone, and argument in the U.S. Common Core standards and in exemplary student writing. *Assessing Writing, 24*, 59–73.
- Beach, R. W. (2011). Issues in analyzing alignment of language arts Common Core standard with state standards. *Educational Researcher, 40*(4), 179–182.
- Brooks, J. G. & Dietz, M. E. (2012/2013). The dangers and opportunities of the Common Core. *Educational Leadership, 70*(4), 64–67.
- Burris, C. & Murphy, J. (2014). A disturbing look at Common Core tests in New York. *Washington Post*, June 27, 2014. Retrieved September 15, 2015, from <https://www.washingtonpost.com/blogs/answer-sheet/wp/2014/06/27/a-disturbing-look-at-common-core-tests-in-new-york/>.
- Cassidy, J., Ortlieb, E. & Grote-Garcia, S. (2016) Beyond the Common Core: Examining 20 years of literacy priorities and their impact on struggling readers. *Literacy Research and Instruction, 55*(2), 91–104.
- Chingos, M. M. (2013). *Standardized testing and the Common Core Standards: You get what you pay for?* Washington, D.C.: Brown Center on Education Policy.
- Christenson, S. L., Decker, D. M., Triezenber, H. L., Ysseldyke, J. E., & Reschly, A. (2007). Consequences of high-stakes assessment for students with and without disabilities. *Educational Policy, 21*(4), 662–690.
- Clarke, M., Haney, W., Roades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut, Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy at Boston College.
- College Board. (2011, July). How the New York State testing program aligns to the Common Core State Standards in English language arts. Retrieved March 9, 2017, from https://www.engageny.org/file/341/download/ela_nys_assessment_alignment_to_the_common_core.pdf?token=Xz5LjPg3ZBTFxoLhEuxH7JVVWOTDmaua0FRsUDaUJnLA.
- Common Core State Standards Initiative. (2016, April 30). Standards in your state. Retrieved March 9, 2017, from <http://www.corestandards.org/standards-in-your-state/>.
- Davis, D. S. & Willson, A. (2015). Practices and commitments of test-centric literacy instruction: Lessons from a testing transition. *Reading Research Quarterly, 50*(3), 357–379.
- DeStigter, T. (2015). On the ascendance of argument: A critique of the assumptions of academe’s dominant form. *Research in the Teaching of English, 50*(1), 11–34.
- Dooley, C. M. & Assaf, L. C. (2009). Contexts matter: Two teachers’ language arts instruction in the high-stakes era. *Journal of Literacy Research, 41*(3), 354–391.
- Dutro, E., Selland, M. K., & Bien, A. C. (2013). Revealing writing, concealing writers: High-stakes assessment in an urban elementary classroom. *Journal of Literacy Research, 45*(2), 99–141.
- EngageNY. (n.d.) Changes to New York State Standards, curricula, and assessments: ELA and mathematics. Retrieved November 10, 2015, from <https://www.engageny.org/sites/default/files/resource/attachments/ccstimeline.pdf>.
- Eppley, K. (2015). Seven traps of the Common Core State Standards. *Journal of Adolescent and Adult Literacy, 59*(2), 207–216.
- Fisher, D. & Frey, N. (2014). Addressing CCSS Anchor Standard 10: Text complexity. *Language Arts, 91*(4), 236–250.

- 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.
- Herman, J. L., Webb, N. M., & Zuniga, S. A. (2005). Measurement issues in the alignment of standards and assessments: A case study. (CSE Report 653.) Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hiebert, E. H., & Van Sluys, K. (2014). Examining three assumptions about text complexity: Standard 10 of the Common Core State Standards. In K.S. Goodman, R.C. Calfee, & Y.M. Goodman (Eds.), *Whose knowledge counts in government literacy policies? Why expertise matters* (pp. 144–160). New York, NY: Routledge.
- Hiebert, E. H., & Mesmer, A. E. (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.
- Jeffery, J. V. (2009). Constructs of writing proficiency in U.S. state and national writing assessments: Exploring variability. *Assessing Writing*, 14(1), 3–24.
- Kern, D. (2014). A critical review of the Common Core standards research. *New England Reading Association Journal*, 50, 75–83.
- Kesler, T. (2013). Unstandardized measures: A cross-case analysis of test prep in two urban high-needs fourth-grade classes. *The Elementary School Journal*, 113(4), 488–516.
- Laitsch, D. A. (2006). *Assessment, high stakes, and alternative visions: Appropriate use of the right tools to leverage improvement*. Boulder, CO: National Education Policy Center.
- Mathis, W. J. (2010). *The “Common Core” Standards Initiative: An effective reform tool?* Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved February 18, 2017, from <http://epicpolicy.org/publication/common-core-standards>.
- McCarthy, S. J. (2008). The impact of No Child Left Behind on teachers’ writing instruction. *Written Communication*, 25(4), 462–505.
- McMahon, J. (2015, October 27). New York, US officials retreat from push toward Common Core testing. Retrieved November 30, 2015, from http://www.syracuse.com/schools/index.ssf/2015/10/public_officials_in_central_new_york_country_shift_stance_on_common_core.html.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA’s study on reading literacy achievement in primary school in 35 countries*. Chestnut Hill, MA: International Study Center.
- National Governors Association Center for Best Practices (NGACBP) & Council of Chief State School (CCSS), 2015 National Governors Association Center for Best Practices (NGACBP) & Council of Chief State School (CCSS). (2015). *Supplemental information for Appendix A of the Common Core State Standards for English language arts and literacy: New research on text complexity*. Washington, D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.
- New York State Education Department, 2013 New York State Education Department. (2013, June 3) *Regents exams: ELA*. Retrieved November 30, 2015, from <https://www.engageny.org/resource/regents-exams-ela>.
- New York State Education Department, 2014 New York State Education Department. (2014). *Educator guide to the 2014 Regents examination in English language arts (Common Core)*. Retrieved November 10, 2015, from <https://www.engageny.org/resource/regents-exams-ela-english-language-arts-test-guide>.
- Nichols, S. L. (2007). High-stakes testing. *Journal of Applied School Psychology*, 23(2), 47–64.
- Partnership for the Assessment of Readiness for College and Careers. (2015). *PARCC states*. Retrieved November 30, 2015, from <http://www.parcconline.org/about/states>.
- Pearson, P. D. & Hiebert, E. H. (2013). *The state of the field: Qualitative analyses of text complexity* (Reading Research Report No. 13.01). Santa Cruz, CA: TextProject.
- Pendergrast, D., & Swain, K. (2013). Competing interests? NAPLAN and middle schooling assessment practices. *Middle Years of Schooling Association*, 13(1), 5–17.
- Roderick, M. & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 24(3), 97–227.
- Sahlberg, P. (2011). Lessons from Finland. *American Educator*, 35(2), 32–36.
- Rothman, R. (2014). The Common Core takes hold. *Education Next*, 14(3), 16–22.
- Shanahan, T. (2014). How and how not to prepare students for the new tests. *The Reading Teacher*, 68(3), 184–188.
- Smarter Balanced Assessment Consortium. (2015). Member states. Retrieved November 30, 2015, from <http://www.smarterbalanced.org/about/member-states/>.
- Snow, C. & O’Connor, C. (2013). *Close reading and far-reaching classroom discussion: Fostering a vital connection*. Newark, DE: A Policy Brief from the Literacy Research Panel of the International Reading Association.
- Sundeen, T. H. (2015). Writing instruction for adolescents in the shadow of the Common Core State Standards. *Journal of Adolescent and Adult Literacy*, 59(2), 197–206.
- Tienken, C. H. (2008). Rankings of international achievement test performance and economic strength: Correlation or conjecture? *International Journal of Educational Policy*, 3(4), 1–15.

- Tienken, C. H. (2011). Common Core Standards: The emperor has no clothes, or evidence. *Kappa Delta Pi Record*, 48(4), 152–155.
- Troia, G. (2014). *Evidence-based practices for writing instruction* (Document No. IC-5). Retrieved November 25, 2015, from <http://cedar.education.ufl.edu/tools/innovation-configuration/>.
- Troia, G., Olinghouse, N. G., Wilson, J., Steward, K. A., Mo, Y., Hawkins, L., & Kopke, R. A. (2016). Common Core Writing Standards: A descriptive study of content and alignment with a sample of former state standards. *Reading Horizons*, 55(3), 98–141.
- Ujifusa, A. (2012). Exit exams face pinch in Common-Core push. *Education Week*, 32(6), (1, 19).
- U.S. Department of Education. (2009). *Race to the Top Program executive summary*. Retrieved November 20, 2015, from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- U.S. Department of Education. (2010). *U.S. Secretary of Education Duncan announces winners of competition to improve student assessments*. Retrieved November 13, 2015, from <http://www.ed.gov/news/press-releases/us-secretary-education-duncan-announces-winners-competition-improve-student-assessments>.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7–25.
- Wilcox, K. C., & Jeffery, J. V. (2016, July). Writing standards reform in the US: A multiple case study of elementary and secondary teachers' practices implementing the Common Core. Paper presented at the biennial meeting of the *European Association for Research on Learning and Instruction Conference of Writing Research*: Liverpool, England.
- Wilcox, K. C., Jeffery, J. V., & Gardner-Bixler, A. (2016). Writing to the Common Core: Teachers' responses to changes in standards for writing in elementary schools. *Reading and Writing: An Interdisciplinary Journal*, 29(6), 903–928.



Appendix: *Text Complexity Rubric* (adapted from Fisher & Frey, 2014)

Complexity Feature	Texts Would Stretch Reader (3)	Texts Require Grade-appropriate Skills (2)	Texts Are Comfortable (1)
Levels of Meaning and Purpose			
Density and complexity	Significant density, complexity, with multiple levels of meaning; meanings may be somewhat ambiguous	Single, but more complex/abstract level; some meanings stated, while others are left to the reader	Single and literal levels of meaning; meaning is explicitly stated
Figurative language	Figurative language plays a significant role in identifying meaning of the text; more sophisticated figurative language is used (irony and satire, allusions, archaic or less familiar symbolism); the reader is left to interpret these meanings	Figurative language (imagery, metaphors, symbolism, personification) used to make connections within text to more explicit information, readers supported in understanding these devices through examples and explanations	Limited use of symbolism, metaphors, and poetic language that allude to other unstated concepts; language is explicit and relies on literal interpretations
Purpose	Purpose is deliberately withheld from the reader, who must use other interpretative skills to identify it	Purpose is implied but is easily identified based on title or context	Purpose or main idea is directly and explicitly stated at the beginning of the reading
Structure			
Genre	Genre is unfamiliar or bends and expands the rules for the genre	Genre unfamiliar but sample represents characteristic. OR familiar genre that bends rules for genre	Genre is familiar and the text is consistent with the elements of that genre
Organization	Organization distorts time/sequence in a deliberate effort to delay reader's full understanding of plot, process, or	Organization adheres to most conventions, but digresses on occasion to temporarily shift the reader's focus to	Organization is conventional, sequential, or chronological, with clear signals and transitions to lead the reader

Continued on next page

Appendix: *Continued*

Complexity Feature	Texts Would Stretch Reader (3)	Texts Require Grade-appropriate Skills (2)	Texts Are Comfortable (1)
	set of concepts; may include significant flashbacks, foreshadowing, or shifting perspectives	another point of view, event, time, or place, before returning to the main idea or topic	through a story, process, or set of concepts
Narration	Unreliable narrator provides distorted/limited view; must use clues to find truth; multiple narrators provide conflicting information; shifting POV	Third-person limited or first person narration provides accurate, but limited perspectives or viewpoints	Third-person omniscient narration or authoritative, credible voice provides appropriate level of detail and keeps little hidden from the view of the reader
Language Conventionality and Clarity			
Standard English and variations	The text includes significant and multiple styles of English and its variations, and these are unfamiliar to the reader	Some distance exists between the reader's linguistic base and the language conventions used in the text; the vernacular used is unfamiliar to the reader	Language closely adheres to the reader's linguistic base
Register	Archaic, formal, domain-specific, or scholarly register	Register is consultative/formal, may be academic, but acknowledges developmental level of the reader	Register is casual and familiar
Knowledge Demands			
Background knowledge	Text places demands on the reader that extend far beyond one's experiences, and provides little in the way of explanation of	Distance between reader's experiences and those in text, but acknowledges these divergent experiences, and sufficient explanation to bridge gaps	The text contains content that closely matches the reader's life experiences

Continued on next page

Appendix: *Continued*

Complexity Feature	Texts Would Stretch Reader (3)	Texts Require Grade-appropriate Skills (2)	Texts Are Comfortable (1)
	these divergent experiences		
Prior knowledge	Specialized or technical content knowledge is presumed and little in the way of review or explanation of these concepts is present in the text	Subject-specific knowledge is required, but the text augments this with review or summary of this information	Prior knowledge needed to understand the text is familiar and draws on a solid foundation of practical, general, and academic learning
Cultural knowledge	Text relies on extensive or unfamiliar intertextuality and uses artifacts and symbols that reference archaic or historical cultures	Text primarily references contemporary pop culture to anchor explanations for new knowledge; intertextuality used extensively but mostly familiar	The reader uses familiar cultural templates to understand the text; limited or familiar intertextuality
Vocabulary knowledge	Vocabulary demand is extensive, domain-specific, and representative of complex ideas; the text offers little in the way of context clues to support the reader	Vocabulary draws on domain-specific, general academic, meaning words, with text supports to guide reader's interpretations of meanings; vocabulary used represents familiar concepts/ideas	Vocabulary is controlled and uses the most commonly held meanings; multiple-meaning words are used in a limited fashion