Effects of States' Adoption of College- and Career-Ready Standards on Student Achievement

(4-1-2019)

[DRAFT—PLEASE DO NOT CITE OR CIRCULATE WITHOUT AUTHOR PERMISION]

Mengli Song

Rui Yang

Michael Garet

American Institutes for Research

Introduction

Since the release of the report *A Nation at Risk* (National Commission on Excellence in Education, 1983), standards-based reforms have been a crucial part of federal and state efforts to improve education. College- and career-ready (CCR) standards--the focus of the current wave of standards-based reform—differ from states' previous standards in important ways. Most notably, the new CCR standards were explicitly designed around the goal of ensuring college and career readiness for all students upon high school graduation. This explicit focus on college and career readiness in the CCR standards stemmed from the concern that too many students do not have the knowledge and skills needed for success in college and the workplace. Nearly half of new college students, for example, had to take remedial courses,[1] and surveys of employers also showed widespread dissatisfaction with the literacy and mathematics skills of young job applicants (U.S. Department of Education, 2010).

In addition to high college-remediation rates, another impetus for the current wave of standards-based reform was the recognition that the rigor of states' standards varied widely across states and declined in many states as an unintended consequence of the accountability requirements under the No Child Left Behind Act of 2001 (Bandeira de Mello, Blankenship, & McLaughlin, 2009). To encourage states to adopt more rigorous standards, the Obama administration built into its $4.35 billion Race To the Top grant program the requirement that states applying for the grant need to demonstrate their commitment to adopting rigorous CCR standards. Specifically, states were required to participate in a consortium consisting of a significant number of states working toward jointly developing and adopting a common set of K-12 standards that "are internally benchmarked and build toward college and career readiness by the time of high school graduation" (U.S. Department of Education, 2009, p.7).

The push for common standards across states was further strengthened through the Elementary and Secondary School Act (ESEA) flexibility, which provided states with waivers of certain requirements under ESEA. As one of the conditions for receiving the waivers, states must adopt CCR standards in English language arts (ELA) and mathematics that are:

> either (1) standards that are common to a significant number of States; or (2) standards that are approved by a State network of institutions of higher education, which must certify that students who meet the standards will not need remedial course work at the postsecondary level. (U.S. Department of Education, 2012, p.5)

In total, 45 states, the District of Columbia (DC), and two U.S. territories submitted requests for ESEA flexibility, and approval has been granted to all but two states. The great majority of those states met the flexibility requirement regarding CCR standards by adopting the Common Core State Standards (CCSS).

Indeed, a strong emphasis on common standards across states—CCSS in particular—has been a distinctive feature of the current standards-based reform. Spearheaded by the National Governors' Association and the Council of Chief State School Officers, the CCSS Initiative was launched in 2009 and aimed to develop a common set of ELA and mathematics standards for all states, based on evidence of what knowledge and skills are needed to be ready for college and career upon high school graduation,

---

[1] Based on the estimates from the National Center for Education Statistics (2010), for example, 40 percent of 2003–04 high school seniors who had enrolled in college by 2006 and 51 percent of the entering students in public 2-year institutions took remedial courses.

and internationally benchmarked to the world's highest-performing countries.[2] Proponents of the CCSS argue that consistent standards across states promote equity by providing teachers, parents, and students with a set of clear expectations to ensure that all students have the skills and knowledge necessary to succeed in college and career, regardless of where they live. Common standards are also expected to benefit states by facilitating collaboration among states on a range of tools and policies, such as textbooks and assessments, and thus enabling cost-sharing and economies of scale.

Released in June 2010, the CCSS was quickly adopted by 45 states and DC by the end of 2011, and adopted by one more state (Washington) in June 2012.[3] The extraordinary initial response of states to the CCSSS, however, was followed by a steady decline in public support. The annual Education Next public opinion polls, for example, show that in 2012, 90% of all those taking a side indicated that they supported the CCSS, and this percentage declined steadily to 83% in 2013, 58% in 2015, and 50% in 2016 (Peterson, Herderson, West, & Barrows, 2016). By fall 2017, 11 states had announced a major Common Core rewrite or replacement.[4] In recent years, though, public support for the CCSS has increased slightly—52% in 2017 and 54% in 2018 according to the Education Next polls.

There are many reasons for the rising opposition to the CCSS. Some educators, for example, do not agree with certain aspects of the standards themselves (e.g., the increased emphasis on non-fictional text in ELA). The lion's share of attacks on the CCSS, however, are grounded in politics (Gewertz, 2015). Many conservative activists and policymakers see the active involvement of the federal government in promoting the CCSS as an encroachment on states' rights, even though the creation of the standards was entirely a state-led effort. Some on the liberal side have also voiced concerns about the CCSS, as they fear that common standards may undermine teachers' creativity. Additional factors contributing to the opposition to the CCSS, as revealed by a state representative poll of California voters, included disapproval of President Obama, opposition to testing, and certain misconceptions about the CCSS (Polikoff, Hardaway, Marsch, & Plank, 2016).[5]

While there is no shortage of opinions in the contentious field of CCR standards, there is clearly a lack of rigorous empirical evidence on the impact of the new standards on student learning, partly due to the relatively short history of the new standards. The study presented in this paper is intended to begin to fill in this gap. As part of a larger research agenda on the implementation and impact of the CCR standards, this study was designed to assess the effects of CCR standards on student achievement. The overarching question guiding the study is as follows:

> *Did states' adoption of CCR standards result in increases in student achievement in reading and mathematics, both overall and for key student subgroups (e.g., SWDs and ELLs)?*

Relying on 1990–2017 state-level NAEP data in reading and mathematics for grades 4 and 8, we addressed the above question using a comparative interrupted time series design. Before we describe in

---

[2] See http://www.corestandards.org/ for detailed information about the CCSS.

[3] The 45 states include Minnesota which adopted CCSS only in ELA but not in math. Four states (Alaska, Nebraska, Texas, and Virginia) did not adopt CCSS, but adopted their own CCR standards between 2008 and 2015.

[4] See the current CCSS adoption map at https://www.edweek.org/ew/section/multimedia/map-states-academic-standards-common-core-or.html.

[5] The authors found, for example, the misconception that states were not allowed to add content to the CCSS was associated with a 64 percent increase in the odds of opposition ($p < .001$).

detail the study design and present study findings, a brief review of existing research on CCR standard is in order.

Existing Research on CCR Standards

During the past few years, there has been a growing body of research on the implementation of the CCR standards, focusing primarily on the CCSS. Research on the impact of the CCR standards on teaching and learning, however, has been scant. In this section, we first review findings from existing research on the implementation of the CCSS standards, which provides useful context for understanding the potential impact of the new standards. We then describe the findings from three studies that attempted to assess early impact of CCSS standards on student achievement (Loveless, 2014, 2015; Xu, 2018).

Research on the Implementation of CCSS Standards

Most of the existing studies of CCSS implementation are descriptive studies based on surveys of various types of stakeholders (e.g., teachers, principals, and district and state officials). Key issues examined in these surveys include support for the new standards, perceptions of the standards, and implementation challenges. Below we highlight some of the key findings on each of these issues.

Support for the CCSS

Stakeholder support is critical for successful implementation of any initiatives. Findings about support for the CCSS are mixed, depending on the type of respondents and timing of the survey. As mentioned earlier, the annual Education Next polls of nationally representative samples of adults show a steady decline in public support for the CCSS (Peterson, Herderson, et al., 2016). From 2013 to 2016, for example, the percentage of respondents taking a side who supported the use of the CCSS decreased from 83% to 50%, with the decrease particularly pronounced among Republicans (from 82% to 39%) than among Democrats (from 86% to 60%). Further, the Education Next polls also show that support for the CCSS declined among teachers as well—from 87% in 2013 to 54% in 2014 and 44% in 2015, but it stabilized between 2015 and 2016.

The teacher surveys jointly conducted by the Scholastic and the Bill & Melinda Gates Foundation (2014) also revealed a decline in support for the CCSS among teachers between 2013 and 2014, but to a lesser extent. Specifically, the Scholastic and Gates surveys found that about three quarters (73%) of teachers in CCSS-implementing states said in 2013 that they were enthusiastic about the implementation of the new standards, and this figure decreased slightly to 68% in 2014. Among teachers who had experienced more than one year of full CCSS implementation by 2014, however, a much higher percentage (84%) indicated that they were enthusiastic about CCSS implementation.

Results from the 2013 poll by the National Education Association were similar to the results from the 2013 Scholastic and Gates teacher survey. Among members of the association, about three quarters supported the CCSS either “wholeheartedly” (26%) or “with some reservations” (50%) and only 11% expressed opposition to the standards (Walker, 2013). Among the reasons that the respondents cited for their support for the CCSS were the clearer guidelines and education goals of the new standards (38%), the alignment of the CCSS with what they teach (27%), and the higher level of rigor of the new standards (23%).

While less than half of the teachers nationwide supported the CCSS in 2015 based on the Education Next poll, support for the CCSS was much stronger among teachers taking a 2015 survey of educators in five states that participated in the PARCC or SBAC assessment consortia (Kane, Owens, Marinell, Thal, &

Staiger, 2016). In these states, about three quarters (73%) of the teachers surveyed reported that they have embraced the CCSS "quite a bit" or "fully," and more than two thirds (69%) of the principals agreed that the new standards will have a positive effect on student learning in the long run.

Perceptions of the CCSS

While public support for the CCSS deteriorated from the earlier years due perhaps in a large part to political turmoil, perceptions of the quality of the CCSS actually improved substantially among district leaders, who have been on the front lines of implementing the new standards. For example, based on a 2014 survey of a nationally representative sample of school districts conducted by the Center on Education Policy (CEP) (Rentner & Kober, 2014), the overwhelming majority (about 90%) of district leaders in CCSS-adopting states agreed that the new standards are more rigorous than their state's previous standards, reflecting a dramatic increase from 2011 (58% in math and 57% in ELA) (Kober & Rentner, 2011a). The percentage of district leaders who agreed that the new standards will lead to improved student achievement also increased considerably according to the CEP surveys—from less than 60% (55% in math and 58% in ELA) in 2011 to over three quarters (76% in math and 79% in ELA) in 2014. Perceptions of the CCSS among state administrators in CCSS-adopting states were even more positive—all 40 state administrators responding to a 2013 CEP survey and all but a few of the 35 state administrators responding to a 2011 CEP survey agreed that the CCSS are more rigorous than their previous standards and will improve students' skills in math and ELA (Kober & Rentner, 2012; Renter, 2013).[6]

Teachers, however, appeared to have more reservations about the impact of the CCSS on student achievement. Less than half (48%) of the teachers participating in the 2014 Scholastic and Gates survey indicated that the new standards will have a positive impact for most students, down from 57% in 2013. A small percentage of teachers (8% in 2013 and 17% in 2014) believed that the CCSS will have a negative impact for most students, and about a third of teachers in both years did not think the CCSS will make much of a difference for most students. Notably, teachers who expressed more positive views on the new standards were more likely to report that their in-school, CCSS-related experiences were "extremely" or "very" helpful, more likely to report having received information about the CCSS through professional development, and less likely to report having received information about the new standards through the media. A more recent survey of a nationally representative sample of K–8 math teachers from 43 CCSS-adopting states and DC produced similar results—a little over half of the teachers surveyed in 2015 believed that that the CCSS in math will have long-term benefit for students (Bay-Williams, Duffett, & Griffith, 2016).

Implementation Challenges

Statewide transition from older standards to the new CCR standards is a massive undertaking and has created challenges at multiple levels. At the state level, finding adequate resources to support all the necessary CCSS implementation activities continued to be the most frequently cited challenge faced by states based on annual surveys conducted by the CEP. Among the 40 CCSS-adopting states that responded to the CEP 2013 survey, for example, 22 states considered inadequate funding as a major

[6] A 2010 content analysis of the CCSS and each state's standards conducted by the Thomas B. Fordham Institute gave the CCSS an overall rating of "A-" in math and "B+" in English with respect to the content, rigor, clarity and specificity of the standards (Carmichael, Martino, Porter-Magee, & Wilson, 2010). The study concluded that the CCSS are clearly superior to the state standards in 39 states in math and 37 states in English.

challenge in implementing the CCSS and 12 states considered it a minor challenge (Rentner, 2013). These figures were similar to what was found in earlier surveys (Kober & Rentner, 2011b, 2012).

In addition to inadequate funding, the majority of the states participating in the 2013 CEP state survey also reported challenges such as developing educator evaluation systems that hold educators accountable for student mastery of the CCSS (32 states) and identifying and/or developing curriculum materials needed for implementing the new standards (26 states) (Rentner, 2013). The majority of the states also reported major challenges in the area of professional development. For example, 37 states considered it a challenge to provide sufficient high-quality professional development to help teachers implement the new standards, and 33 states felt it challenging to provide all principals with state-sponsored professional services on the CCSS (Kober, McIntosh, & Rentner, 2013).

As is the case with state-level challenges, finding adequate resources to support all the implementation activities also topped the list of challenges faced by districts in the implementation of the CCSS. Based on the 2014 CEP district survey (Rentner & Kober, 2014), inadequate resources was cited as a major implementation challenge by two thirds of the districts and a minor challenge by one quarter of the districts in CCSS-adopting states. This level of challenge was similar to what was found in a 2011 CEP survey, where inadequate funding was deemed a major challenge by 76% of the districts and a minor challenge by 21% of the districts surveyed (Kober & Rentner, 2011).

A large majority of the districts taking the 2014 CEP survey also identified the following issues as either a major or a minor challenge: having adequate district staffing levels (87%) and staff expertise (86%) to implement all aspects of the CCSS, identifying and/or developing the curriculum materials necessary to implement the CCSS (90%), providing high-quality professional development and other support to ensure that teachers are able to implement the CCSS instructional activities (88%), and having enough time to implement the CCSS before consequences are tied to student performance on the CCSS-aligned assessment (89%).

The 2014 CEP survey also revealed that about three quarters of the districts in CCSS-adopting states considered overcoming resistance to the CCSS *within* the K-12 system as either a major challenge (25%) or a minor challenge (49%), and that a similar percentage of districts considered overcoming resistance to the CCSS from *outside* the K-12 system as either a major challenge (34%) or a minor challenge (39%). In contrast, a similar survey conducted three years earlier indicated that only 10% of the districts deemed overcoming resistance to the CCSS *within* the K-12 system as a major challenge and only 5% deemed overcoming resistance to the CCSS *from outside* the K-12 system as a major challenge (Kober & Rentner, 2011a). The increased resistance to the CCSS from both within and outside the K-12 system as revealed by the CEP district surveys is consistent with the decline in public support for the new standards based on national polls (Peterson, Herderson, et al., 2016).

Implementation of the CCSS at the school level also proved to be challenging. According to the Scholastic and Bill & Melinda Gates Foundation (2014) survey, almost three quarters (73%) of teachers reported in 2013 that they believed the implementation of the CCSS in their schools is or will be challenging, and the sentiment was shared even more widely in 2014 (81%). These teachers cited a variety of issues that had been problematic for their schools in implementing the CCSS, most notably factoring student results on new tests into teacher evaluation (59%) and uncertainty about which assessments their state will use (51%).

Finally, at the teacher level, one major challenge to successful implementation of the CCSS in teachers' classrooms is that many teachers were not well prepared to implement the new standards. The

Scholastic and Bill & Melinda Gates Foundation (2014) survey, for instance, revealed that less than one third (31%) of the teachers surveyed felt "very" prepared to teach the CCSS, about half felt only "somewhat" prepared," and 21% felt "somewhat" or "very" unprepared in 2014. On a positive note, however, these figures reflect an increasing level of teacher preparedness for teaching to the CCSS—in 2013, in comparison, only 20% of the teachers surveyed felt "very" prepared and 29% felt unprepared to teach the CCSS. In a more recent survey (Kane et al., 2016), a large majority (85%) of teachers in five states that participated in PARCC or SBAC reported having good or excellent knowledge of the CCSS. Despite their familiarity with the new standards, however, only a third of the teachers reported feeling "quite" prepared or "extremely" prepared to teach their students what they need to know to succeed on the new CCSS-aligned assessments.

Another obstacle to successful transition to the CCSS at the classroom level is the lack of CCSS-aligned curricular and instructional materials. In a recent survey of K–8 math teachers from 43 CCSS-adopting states and DC, over 40% of the teachers surveyed reported that the math materials available to them were not well aligned to the new standards (Bay-Williams et al., 2016). Several recent content analyses of textbooks reached a similar conclusion: claims about textbooks aligned to the CCSS are often unfounded. The first round of reviews of K-8 math instructional series released by EdReport.org, for instance, shows that, contrary to the publishers' claims, 17 of the 20 math curricula reviewed failed to meet criteria for alignment with the CCSS (Heitin, 2015). Findings from EdReport.org's initial review of seven ELA series were mixed but more positive—three were considered fully aligned to the CCSS, three partially aligned, and one fully unaligned (Heitin, 2016). These findings mirror the findings from Polikoff's (2015) analysis of the alignment of four popular textbooks to the CCSS for fourth grade mathematics, which revealed substantial areas of misalignment and challenged the publishers' claims of alignment.

Research on the Impact of CCSS Standards

While there has been a large body of research examining the implementation of the CCSS, research on the impact of the CCSS is rather limited, possibly due to the novelty of the standards and challenges in designing rigorous impact studies given the nearly universal adoption of the CCSS. In this section, we review findings from a number of survey-based studies which gathered data on the impact of the CCSS on teaching and learning as reported by teachers. These "impact" findings based on teachers' self-report are descriptive in nature and need to be interpreted with caution, as they do not have the same level of causal validity as findings from impact studies based on a more rigorous design such as a randomized experiment.

Also reviewed in this section are findings from three recent studies that explicitly examined the relationship between CCSS implementation and student achievement, two based on NAEP data (Loveless, 2014, 2015) and one based on statewide data from Kentucky (Xu & Cepa, 2018). Given design limitations, findings from these studies also need to be interpreted with caution.

Impact of the CCSS on Teaching and Learning as Perceived by Teachers

There has been evidence based on teachers' self-reports that CCSS implementation has produced positive changes in both teachers' instructional practice and student learning. The majority of the teachers participating in the survey conducted by Kane and colleagues (2016), for example, reported making major changes to their instructional practices and materials to align with the new standards. More than three quarters (76%) of teachers surveyed, for example, reported having changed at least half of their classroom instruction as a result of the CCSS, and about four out of five mathematics teachers (82%) and three out of four English teachers (72%) reported having changed more than half of their instructional materials in response to the CCSS.

The majority of the K-8 math teachers responding to the survey conducted by Bay-Williams and colleagues (2016) similarly reported making changes to many of their practices in ways consistent with the CCSS. Almost two thirds (64%) of the teachers surveyed, for example, reported that they were devoting more attention to requiring students to explain in writing how they got their answers than before the CCSS were introduced, and 55% of the teachers reported a greater focus on requiring students to use proper math vocabulary than before. The survey also found that teachers in high-poverty schools were more likely to report that their own teaching practice had changed as a result of the CCSS than teachers in low-poverty schools. In high-poverty schools, for example, 62 percent of the teachers reported an increased focus on requiring students to use proper math vocabulary, as compared with 47 percent for teachers in low-poverty schools. The percentage of teachers reporting an increased focus on teaching multiple methods to solve a problem was also substantially higher in high-poverty schools than in low-poverty schools—61% versus 41%.

Teachers also reported positive changes in student learning. Even though the CCSS was not fully implemented until the 2013–2014 school year in many states, over half (53%) of the teachers in CCSS-adopting states reported in 2014 that they had already seen a positive change in their students' ability to think critically and use reasoning skills as a results of CCSS implementation, and an even higher percentage (68%) of the teachers in schools where CCSS implementation was fully complete in 2012–13 or earlier reported the same (Scholastic and the Bill & Melinda Gates Foundation, 2014). Teachers' views on the impact of the CCSS were most positive among elementary school teachers—62% of elementary teachers reported seeing a positive impact on students' ability to think critically and use reasoning skills, as compared with 47% for middle school teachers and 37% for high school teachers. Similarly, half of the teachers in the Scholastic and Gates survey reported having seen a positive impact of the CCSS on their students' ability to read and comprehend informational text, and the percentage was again the highest among elementary school teachers (59%, as compared with 45% for middle schools teachers and 35% for high school teachers).

The survey of K-8 math teachers conducted by Bay-Williams and colleagues (2016) also found more positive views of the impact of CCSS on student learning among teachers in lower grades than teachers in higher grades. Over three quarters (77%) of the K-2 teachers surveyed, for instance, indicated that their students were developing a stronger number sense and more ability to apply math in real-world situations, as compared with about two thirds (66%) of teachers in grades 3–5 and just over half (52%) of teachers in grades 6–8. Over two thirds (68%) of the K-2 teachers also reported that their students were developing a stronger capacity to persevere in math and come up with solutions on their own as a result of the CCSS, as compared with 61% of grades 3–5 teachers and less than half (46%) of middle school teachers. The less optimistic views of middle school teachers about the impact of the CCSS on students' math abilities, according to the authors, is likely to attributable to the fact that the new middle school standards are much harder than the elementary school standards, particularly relative to the standards they replaced.

Impact of the CCSS on Student Achievement

To the best of our knowledge, there have been only a handful of studies that attempted to assess the impact of the CCSS on student achievement (Loveless, 2014, 2015, 2016; Xu & Cepa, 2018). The first two studies—one focusing on math and one on reading—were conducted by Loveless (2014, 2015) as part of the annual Brown Center reports on American education. Both studies were intended to "estimate

CCSS's early impact" by comparing changes in NAEP scores from 2009 to 2013 between states with different levels of CCSS implementation. Relying on a measure of "congruence" or similarity between each state's 2009 mathematics standards and the CCSS for mathematics created by Schmidt and Houang (2012), Loveless (2014) compared the 2009-2013 8th-grade NAEP gains across five groups of states with congruence ratings ranging from 1 (i.e., "least like CCSS") to 5 (i.e., "most like CCSS"), and found no systematic relationship between states' congruence ratings and changes in their NAEP scores.

In a second set of analyses, Loveless (2014) used a CCSS implementation index created based on a 2011 survey of state education agencies to classify states into three groups: strong implementers (n=19), medium implementers (n=26), and non-adopters (n=5),[7] and concluded that strong implementers experienced a larger gain in NAEP scores in 8th-grade math from 2009 to 2013 than non-adopters. The difference, although in the desired direction, is very small (1.27 points on the NAEP scale, or .04 standard deviations (SDs)).

The same conclusion was also reached in a more recent study. In his 2015 study, Loveless replicated his 2014 analyses using NAEP data for 4th-grade reading, and conducted similar analyses using an alternative CCSS implementation index that designated states as strong implementers (n=12), medium implementers (n=34), and non-adopters (n=4) based on whether the state was expected to fully implement the new standards by the 2012–2013 school year.[8] Taken together, the two sets of analyses suggest that the 2009–2013 gain in NAEP 4th-grade reading score was only slightly higher (by 1.11 to 1.51 points or 0.03 to 0.04 SDs) in strong implementers than in non-adopting states. Similar analyses based on the two alternative CCSS implementation indices and 2009–2015 NAEP data, however, revealed that the 2009–2015 gain in NAEP 4th-grade reading score was actually slightly smaller in strong implementers than in non-adopting states--by 0.01 to 0.02 SDs (Loveless, 2016). For 8th-grade reading, the 2009–2015 NAEP gain in strong implementers was slightly smaller (by 0.003 SDs) based on one implementation index and slightly larger (by 0.02 SDs) based on the other implementation index relative the gain in non-adopting states.

Findings from the three studies described above, however, need to be interpreted with a large grain of salt, as they are based on simple descriptive comparisons of group means between non-equivalent groups of states, and thus reflect associations rather than causal effects. In particular, the "control group" used in all three studies included a small set of non-adopting states, which were quite unique given the almost nationwide adoption of the CCSS. These non-adopters therefore may not be an appropriate comparison group as selection bias may be a serious concern. In addition, given the very small number (4 or 5) of states in the comparison group, results from the analyses presented in Loveless (2014, 2015, 2016) were sensitive to substantial changes in NAEP scores in one or two states, as the author acknowledged.

While the three studies discussed above analyzed NAEP data from all 50 states, the fourth study (Xu & Cepa, 2018) focused on the early effect of the CCSS in a single state—Kentucky. In this study, the authors tracked three cohorts of students from 8th grade through 11th grade, and found that students

---

[7] Strong implementers are states that: (1) adopted the CCSS, (2) provided, guided, or funded professional development on the CCSS, (3) provided curriculum/instructional materials for the CCSS, and (4) worked with a consortium to develop assessments aligned with the CCSS (Loveless, 2014). Medium implementers are states that adopted the CCSS but did not engage in any of the three implementation activities. Non-adopters are states that did not adopt the CCSS.

[8] Minnesota, which adopted the CCSS in ELA but not math, was counted as a non-adopter in the Loveless (2014) study, but an implementer in the 2015 study.

exposed to the CCSS (i.e., students in the two more recent cohorts) scored significantly higher on the ACT taken in the 11th grade than similar students not exposed to the new standards (i.e., student in the earliest cohort) (differences= 0.03–0.04 SDs). The authors caution, however, that the observed differences between the cohorts may not be completely attributable to the CCSS implementation, as cross-cohort differences in student achievement occurred in both the year before and the year after the adoption of the new standards.

Clearly, the empirical research base on the effects of the CCSS on student achievement is still quite thin, even though the CCSS is the official state standards in the majority of the states with potentially far-reaching influence on the teaching and learning of tens of millions of school children. The study presented in this paper is intended to start to fill in this gap in the literature by providing empirical evidence on the effects of the CCR standards on student achievement.

Methods

In this section, we explain in detail the methods we used to address the research question for the study. We first present an overview of the study design, and then describe the measures, data sources, and analytic approach.

Design Overview

To assess the effects of states' implementation of CCR standards on student achievement, we analyzed state-level NAEP data using a comparative interrupted time series (CITS) design, a commonly used quasi-experimental design for assessing the effects of programs and policies that do not lend themselves to randomized experiments. In its simplest form, an interrupted time series design measures the same outcome for a treatment group multiple times before and after a treatment starts (i.e., the point of "interruption"). The effect of the treatment is then estimated by examining the deviation in the level or slope of an outcome from before to after the onset of the treatment.

The simple interrupted time series design, however, is subject to various threats to internal validity, particularly threats due to history—in this case, the possibility that forces other than the introduction of CCR standards might have influenced student achievement at the same time the CCR standards were introduced (Shadish, Cook, & Campbell, 2002). To guard against potential threats to internal validity, a comparison group is often added to this simple version of time series design, extending it to a CITS design (Dee, Jacob, & Schwartz, 2013; Wong, Cook, & Steiner, 2015). Identifying a plausible comparison group that was not affected by the intervention under study, as Dee and Jacob (2011) noted, is the central challenge for any CITS design. It is particularly challenging to identify an appropriate comparison group in a CITS study of a universally-adopted intervention such as the CCR standards, which were adopted across all states and DC by 2015.[9]

To assess the effects of CCR standards, one obvious approach is to take advantage of the natural variation between states in the timing of CCR standards implementation and compare the achievement trend between states implementing the standards (i.e., implementing states) and states not implementing the standards (i.e., non- implementing states). In other words, we would use the non-implementing states as the comparison states for the implementing, or treatment, states. This approach relies on the assumption that there is sufficient variation between states in the timing of CCR standards

[9] For simplicity, we will count DC as one of 51 states hereinafter.

implementation, which unfortunately does not hold.[10] As shown in Table 1, even though states officially adopted CCR standards in ELA over a 7-year window (from 2008 to 2014) and CCR standards in mathematics over a 9-year window (from 2007 to 2015), the overwhelming majority of states (41 for ELA and 39 for math) adopted CCR standards and started implementation in the same year (2010).[11]

Table 1. Number of States by the Year CCR Standards in ELA and Mathematics Were Adopted

Year of CCR Adoption	N of States Adopting CCR Standards in ELA	N of States Adopting CCR Standards in Math
2007	0	1
2008	1	1
2009	1	2
2010	41	39
2011	5	5
2012	2	2
2013	0	0
2014	1	0
2015	0	1
Total	**51**	**51**

Source: National Center for Education Statistics (2015); state education agency websites

Given the limited variation in the timing of CCR standards implementation across states, we used an alternative approach to assessing the effects of CCR standards on student achievement, taking advantage of the natural variation between states in the quality of their content standards prior to the adoption of the new standards. For this study, we relied on the following two existing measures of the quality of states' prior content standards:

- *Prior Rigor Index*: a measure of the rigor of each state's 2010 standards created by the Thomas Fordham Institute (Carmichael, Martino, Porter-Magee, & Wilson, 2010) (separate measures for ELA and mathematics standards), and
- *Prior CCSS-Similarity Index*: a measure of the similarity between each state's 2009 mathematics standards and the CCSS for mathematics created by researchers at Michigan State University (Schmidt & Houang, 2012).

Below we explain how we classified states into treatment states and comparison states for our CITS analyses based on these measures. (See Appendix A for how each state's prior standards are rated on each measure.)

---

[10] Another limitation of this approach is that, given its reliance on non-implementing states to serve as the comparison group, it could only assess the effects of CCR standards before the standards were adopted by all states. This is potentially a serious limitation, since it would likely take several years for the new CCR standards to take hold and produce appreciable effects on student achievement.

[11] For our CITS analyses, we consider the year of adoption as the starting point (time zero) of the timeline of standards implementation. It is possible, however, some states did not start to implement the CCR standards immediately after the standards were adopted. Thus, we could alternatively define the point of interruption for the CITS analyses as the first year the implementation of the CCR standards started, which, however, is not always clear.

State Classification Based on the Prior Rigor Index. The Prior Rigor Index rates the 2010 content standards of all 50 states and DC on a 0–7 point scale for content and rigor based on a pre-specified set of scoring rubrics (see Carmichael et al., 2010 for further details). For our CITS analyses based on this index, we define treatment states as states with a score of 0-3 and comparison states as states with a score of 5-7 on the index. We excluded states with a median score of 4 to allow for a sharper contrast. Our assumption is that the CCR standards would represent a stronger form of "treatment" for states with less rigorous prior standards than for states that already had fairly rigorous standards in place prior to CCR standards adoption. Thus, we would expect the implementation of CCR standards to lead to a larger improvement in student achievement in states with less rigorous prior standards than in states with more rigorous prior standards.

State Classification Based on the Prior CCSS-Similarity Index. The Prior CCSS-Similarity Index focuses on the congruence or similarity between a state's 2009 mathematics standards and CCSS for mathematics in terms of the focus and coherence of the topics covered in the standards (see Schmidt & Houang, 2012, for further details). Based on this measure, Schmidt and Houang grouped states into five categories, ranging from "least like CCSS" to "most like CCSS." For our CITS analyses, we define treatment states as states in the two "least like CCSS" groups, and comparison states as states in the two "most like CCSSM" groups. We excluded states in the middle group to allow for a sharper contrast. Our assumption here is that the CCR standards would induce more drastic changes in standards and thus represent a stronger form of treatment for states whose prior standards were less like CCSS than for states whose prior standards were more like CCSS. Thus, we would expect the implementation of CCR standards to lead to a larger improvement in student achievement in states whose prior standards were less like CCSS than for states whose prior standards were already quite similar to CCSS.

Given that the overwhelming majority of states adopted CCR standards in 2010, and given that the Prior Rigor Index pertains to states' 2010 standards and the Prior CCSS-Similarity Index pertains to states' 2009 standards, we restricted our CITS analyses to states that adopted CCR standards in 2010. Specifically, our CITS analyses based on the Prior Rigor Index included 17 treatment states and 12 comparison states for reading and 20 treatment states and 14 comparison states for math. Our analyses based on the Prior CCSS-Similarity Index, which is available for math only, included 14 treatment states and 12 comparison states. All states included in the two math analysis samples and all but one state (Virginia) included in the reading analysis sample are CCSS states. Because the Prior Rigor Index for math is more strongly correlated with the Prior CCSS-Similarity Index for math than with the Prior Rigor Index for ELA (0.67 vs. 0.50), there is more overlap in state classification between the two math analysis samples than between the math analysis sample and the reading analysis sample based on the Prior Rigor Index.

Based on the state classifications described above, we assessed the effects of CCR standards on student achievement by comparing treatment states and comparison states in their student achievement trajectory. Given that the NAEP assessments in both reading and mathematics were administered every other year since 2003, the available NAEP data (1990–2017) allow us to estimate the effects of the CCR standards for states in our sample 1 year, 3 years, 5 years, and 7 years after adoption. In other words, for states that adopted CCR standards in 2010, their 2011 NAEP data were used to estimate the effects of CCR standards 1 year after adoption (i.e., 1-year effect), their 2013 NAEP data were used to estimate the 3-year effects, their 2015 NAEP data were used to estimate the 5-year effects, and the latest 2017 NAEP data were used to estimate the 7-year effects for those states.[12]

In a sense, the number of years since adoption may be viewed as a proxy for degree of implementation. Given the timeline of states' implementation of the CCR standards (see Appendix B), the 1-year effects based on the CITS analysis represent effects before the CCR standards were fully implemented in any states in our analysis sample. The 3-year effects represent effects after the CCR standards were fully implemented in only a few states in our sample, and the 5-year and 7-year effects represent effects after the CCR standards were fully implemented in all states in our sample. Thus, the 1-year, 3-year, 5-year, and 5-year effects estimated based on CITS analyses capture the effects at different time points over the course of CCR standards implementation, and reflect the effects of the new standards as states were moving from initial adoption to full implementation.

Data and Measures

Measures of Student Achievement

Our primary measures of student achievement are state average NAEP scores in reading and math for grades 4 and 8. The NAEP data are well suited for our longitudinal analyses because NAEP provides a common set of measures across states and across years. Although NAEP was not designed specifically to address CCR or CCSS standards, research that examined the alignment between the NAEP item pool and the CCSS found substantial overlap between the two (Daro, Hughes, & Stancavage, 2015). For example, of the items on the 2015 NAEP grade 4 math assessment, 79 percent are covered by the CCSS for grade 4 or below (87 percent for grade 8 math). Thus, we expect that positive effects of the CCR standards would manifest in improved NAEP scores even though the NAEP assessments are not perfectly aligned with the CCSS or state-developed CCR standards.

Table 2 summarizes the total number of states (including DC) participating in the NAEP assessments by year, subject, and grade. For states included in our CITS analyses, most have 6–8 waves of NAEP scores prior to the 2010 adoption of CCR standards and 4 waves of NAEP scores after the adoption.

Since one goal of this study is to examine whether the effects of CCR standards varied by student subgroup, we analyzed state-average NAEP scores for all students as well as scores for key student subgroups such as students with disabilities (SWDs), English language learners (ELLs), racial/ethnic groups, and students eligible for free- or reduced-price lunch. NAEP data for subgroups defined by race/ethnicity date back to 1990, whereas NAEP data for SWDs, ELLs, and students eligible for free- or reduced-priced lunch are not available until after 1996.

[12] For states that adopted the CCR standards in odd-numbered years (2007, 2009, 2011, and 2015), the available NAEP data could only be used to estimate 2-year, 4-year, 6-year, and/or 8-year effects for these states and the number of such states is too small to allow for reliable estimation (see Table 1).

Table 2. Number of States Participating in NAEP Assessments, by Year, Subject, and Grade

Year	Reading		Mathematics	
	Grade 4	Grade8	Grade 4	Grade 8
1990				38
1992	42		42	42
1994	40			
1996			44	41
1998	40	37		
2000			41	40
2002	44	42		
2003	51	51	51	51
2005	51	51	51	51
2007	51	51	51	51
2009	51	51	51	51
2011	51	51	51	51
2013	51	51	51	51
2015	51	51	51	51
2017	51	51	51	51

Note: Empty cells indicate the NAEP assessment for the given subject and grade was not administered in that year.

In addition to NAEP composite scores for all students and key subgroups, we also analyzed NAEP scores for the two reading subscales (i.e., *gaining information* and *literary experience*) and the five math subscales (i.e., *algebra*, *data analysis*, *geometry*, *measurement*, and *number properties*), as it is possible that the effects of CCR standards on student achievement as measured by NAEP might differ for different subscales due to uneven alignment between the NAEP assessments and the CCR standards in different domains. As the recent NAEP alignment study (Daro et al., 2015) revealed, 79 percent of items on the 2015 NAEP grade 4 math assessment are covered by the CCSS for grade 4 or below, and the percentage varies substantially across the five math subscales—ranging from 47 percent for the subscale on data analysis, statistics, and probability to 96 percent for the subscale on measurement. It is reasonable to assume that the effects of CCR standards on NAEP scores would be stronger for subscales where the NAEP assessment and CCR standards are more closely aligned.

Measures of Time-Varying Covariates

As will be explained in the next section, our CITS analyses include a set of time-varying covariates to control for their influence on student achievement and to improve the precision of the estimated treatment effect. A key covariate for our CITS analyses is the NAEP exclusion rate.[13] Prior to the 1998 reading administration and 2000 NAEP math administration, NAEP did not allow accommodations for SWDs or ELLs, which resulted in the exclusion of some students who could not meaningfully participate in the assessment without accommodations. To ensure that the NAEP sample be as representative as possible, beginning with the 2002 assessments, NAEP has offered accommodations to all students who

[13] The 1990–2013 NAEP exclusion rates by state for math and reading can be found at: http://www.nationsreportcard.gov/reading_math_2013/files/Tech_Appendix_Math.pdf , and http://www.nationsreportcard.gov/reading_math_2013/files/Tech_Appendix_Reading.pdf.

need them to demonstrate their knowledge. In the transition year (1998 for reading and 2000 for math), a split sample design was used, with one sample taking the assessment with accommodations and one sample without. Provision of accommodations was found to result in higher levels of inclusion, but with little effect on NAEP scale scores at the national level. [14]

Another time-varying covariate included into our CITS analyses is per pupil expenditure, which has been shown to contribute to achievement effects (Dee, Jacob, & Schwartz, 2013). To account for inflation over time, we used per pupil expenditure measured in 2016 constant dollars in the CITS analyses. Three additional time-varying covariates included in the CITS analyses are the percentage of students eligible for free/reduced-price lunch), the percentage of non-White students, and the pupil-to-teacher ratio. As indicated by Wong and colleagues (Wong, Cook, & Steiner, 2009), these indicators correlate highly with family income and other demographic variables that are related to student achievement. All these time-varying covariates are state-level measures available from the Common Core of Data.

Analytic Methods

CITS Model With Year Fixed Effects and State Fixed Effects

To assess the effects of states' adoption of CCR standards on student achievement, we conducted CITS analyses of state-level NAEP data to compare the change in student achievement trend from before to after the adoption of the CCR standards in the treatment states with the corresponding change in the comparison states. As specified below, our CITS model allows the baseline achievement trend to differ between the treatment and comparison states, and controls for state and year fixed effects as well as a set of time-varying covariates to improve the precision of the treatment effect estimates.

$$Y_{ts} = \sum_{k=1}^{K} \beta_{0k} S_{ks} + \beta_1 TIME_t + \beta_2 (T_s * TIME_t) + \sum_{m=3}^{M} \beta_{3m} YR_{mt}$$

$$+ \sum_{n=1}^{7} \beta_{4n} (T_s * POST\_YR_{nt}) + \sum_{g=1}^{4} \beta_{5g} X_{gts} + r_{ts}$$

where

- Y_{ts} is the average NAEP score in year *t* in state *s*;
- S_{ks}, *k* = 1, 2, …, and *K*, is a set of dummy indicators for the *K* states included in the analysis;
- $TIME_t$ is a continuous measure of time measured as the number of years since the first year state NAEP test for a given subject and grade was administered (for grade 4 reading, $Time_t$= 0 for year 1992, 2 for 1994, 6 for 1998, … and 25 for 2017);
- $T_s * TIME_t$ is an interaction between a school's treatment status and time;
- YR_{mt}, *m* = 3, …, and *M*, is (*M-2)* dummy indicators for NAEP testing years, where *M* is the total number of NAEP testing years included in the analysis;[15]

---

[14] Studies of the impact of No Child Left Behind on student achievement conducted by Dee and Jacob (2011) and Wong and colleagues (2011) found that their results were not sensitive to whether the analysis was based on NAEP data with accommodations or data without accommodations from the transition years. Therefore, we used NAEP data with accommodations from these years in our analyses.

[15] Indicators for the first two NAEP testing years are omitted from the model to avoid multicollinearity. For grade 4 math, YR_{1t} =1 for year 1992, YR_{2t}= 1 for 1994, …, and YR_{12t}=1 for 2017. (See Table 2 for NAEP testing years by subject and grade.)

- $T_s * POST\_YR_{nt}$, *n* = 1, 3, 5, and *7*, is a set of interactions between treatment status and dummy indicators for each of the 4 post-CCR NAEP testing years examined ($POST\_YR_{nt}$= 1 for the *n*th post-CCR year and 0 otherwise);
- X_{gts}, *g* = 1 ~ 4, is a vector of four time-varying covariates for year *t* and state *s;* and
- r_{ts} is a random error associated with year *t* and state *s*.

The estimate of primary interest from the above model is β_{4n}, which captures the treatment effect on state average NAEP score in each of the 4 post-CCR NAEP testing years included in the analysis (i.e., 1 year, 3 years, 5 years, and 7 years after CCR standards adoption). We estimated the model separately by subject (reading and math) and grade (4 and 8), for NAEP composite scores and subscale scores, and for all students and key student subgroups. For all analyses, the standard errors of the treatment effects were estimated using the block bootstrap method to account for the serial autocorrelation in the time series data (Bertrand, Duflo, & Mullainathan, 2002).

Sensitivity Analyses

The CITS analyses described above are non-experimental in nature and thus subject to threats to internal validity. The validity of the treatment effect estimates from these analyses relies on the assumption that the post-CCR deviation from the pre-CCR achievement trend in the comparison states provides a valid counterfactual for what would have happened in the treatment states had the states not adopted the CCR standards. This assumption, however, might not hold if unobserved forces (e.g., certain events or changing demographics) occurred during the post-CCR time period and affected student achievement differently in treatment and comparison states. If, for instance, the economic conditions experienced a larger improvement during the post-CCR period in the treatment states relative to the comparison states, then the differential change in the economic conditions of the two types of states may pose a "history" threat to the internal validity of our CITS analyses and may potentially result in an overestimated treatment effect.

While it is not possible to rule out all possible threats to internal validity, we will check the robustness of our CITS estimates to some potential internal validity threats following the method used by Dee and Jacob (2011). Specifically, we will estimate a model that is similar to our main CITS model but uses a time-varying measure of a state characteristic (e.g., per pupil expenditure or percentage of students eligible for free- or reduced-price lunch) as the dependent variable. A lack of treatment effect on such a measure would rule it out as a potential confounder of the treatment effect on student achievement based on the CITS analyses.

As another type of sensitivity analysis, we will check the robustness of our results to alternative measures of treatment status. In our main analyses, states were classified into a treatment group and a comparison group based on certain cut points on the Prior Rigor Index and the Prior CCSS-Similarity Index of the quality of states' prior content standards. As sensitivity analyses, we could check the robustness of our results to alterative cut points on the Prior Rigor Index and the Prior CCSS-Similarity Index that define the treatment and comparison groups for the CITS analyses.

In addition to categorical measures of treatment status, we also will use the Prior Rigor Index and the Prior CCSS-Similarity Index as continuous measures of treatment strength or dosage in the CITS model. The idea is that the lower the rigor of a state's prior content standards as indicated by the Prior Rigor Index, and the more different a state's prior standards were from the CCSS as indicated by the Prior CCSS-Similarity Index, the stronger the treatment induced by the new CCR standards, and the larger the expected effects of CCR standards.

Findings

In this section, we present findings from CITS analyses based on NAEP composite scores and subscale scores in reading and mathematics for all students as well as for select key student subgroups in the NAEP sample. Additional findings from sensitivity analyses will be available at a later time.

Effects of CCR Standards on All Students Based on NAEP Composite Scores

Table 3 presents the estimated effects of CCR standards based on CITS analyses in which the treatment and comparison states were defined based on the Prior Rigor Index and the Prior CCSS-Similarity Index of the quality of each state's prior content standards, respectively. The table presents separate estimates for effects 1 year, 3 years, 5 years, and 7 years after adoption (i.e., 1-year, 3-year, 5-year, and 7-year effects), in both the original 0-500 NAEP scale and the SD unit. Contrary to our expectation, the table reveals significant negative effects for grade 4 reading, with effect sizes ranging from -0.10 to -0.06, which are either significant at the .05 level or marginally significant at the .10 level. Negative effects were also observed for grade 8 reading, grade 4 math, and grade 8 math, although none of those effects were statistically significant except for the 7-year effect for grade 8 math with state classification based on the Prior Rigor Index (effect = -0.10 SD, p < .05).

Table 3. Estimated Effects of CCR Standards on Student Achievement as Measured by NAEP Composite Scores, by Subject, Grade, and Timing of Effect

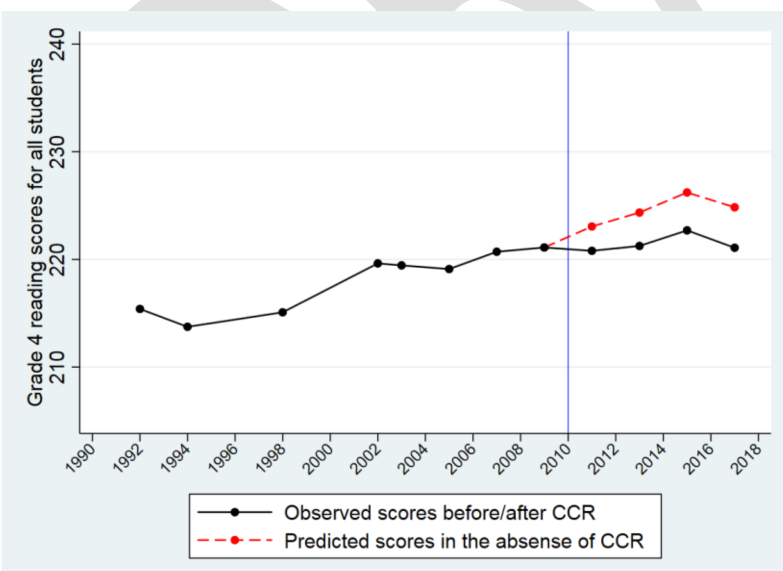
Grade/subject	Timing of effect (year after adoption)	N of states	N of observations	Estimate	Standard error	Effect in SD*	P-value
State classification based on the Prior Rigor Index							
Grade 4 reading	1 year	29	336	-2.26	0.90	-0.06	0.012
Grade 4 reading	3 years	29	336	-3.10	1.30	-0.08	0.017
Grade 4 reading	5 years	29	336	-3.52	1.89	-0.10	0.063
Grade 4 reading	7 years	29	336	-3.76	1.90	-0.10	0.047
Grade 8 reading	1 year	29	281	0.33	0.89	0.01	0.709
Grade 8 reading	3 years	29	281	-1.26	1.09	-0.04	0.248
Grade 8 reading	5 years	29	281	-1.66	1.62	-0.05	0.306
Grade 8 reading	7 years	29	281	-2.08	1.68	-0.06	0.216
Grade 4 math	1 year	34	357	0.42	1.00	0.01	0.674
Grade 4 math	3 years	34	357	-0.79	1.24	-0.03	0.524
Grade 4 math	5 years	34	357	-1.55	1.70	-0.05	0.362
Grade 4 math	7 years	34	357	-2.24	1.81	-0.07	0.217
Grade 8 math	1 year	34	381	0.16	1.02	0.00	0.875
Grade 8 math	3 years	34	381	-1.15	1.47	-0.03	0.436
Grade 8 math	5 years	34	381	-2.48	1.90	-0.07	0.190
Grade 8 math	7 years	34	381	-3.94	1.99	-0.10	0.048
State classification based on the Prior CCSS-Similarity Index							
Grade 4 math	1 year	26	271	1.50	1.03	0.05	0.147

Grade 4 math	3 years	26	271	0.13	1.51	0.00	0.931
Grade 4 math	5 years	26	271	-0.63	1.65	-0.02	0.703
Grade 4 math	7 years	26	271	-1.49	1.55	-0.05	0.338
Grade 8 math	1 year	26	287	0.64	1.03	0.02	0.533
Grade 8 math	3 years	26	287	0.12	1.31	0.00	0.926
Grade 8 math	5 years	26	287	-0.48	1.96	-0.01	0.806
Grade 8 math	7 years	26	287	-2.47	2.03	-0.06	0.223

Note: * Effect in SD was computed by dividing the estimate in the original NAEP scale by the SD of the NAEP test scores for public school students for the relevant grade, subject, and year.

To graphically illustrate the CITS analysis results, we plotted the average observed (unadjusted) NAEP grade 4 reading scores for 14 treatment states identified based on the Prior Rigor Index both before and after the 2010 adoption of CCR standards (shown by the black line in Figure 1), and their predicted scores 1 year, 3 years, 5 years, and 7 years after adoption had they not adopted the new standards (shown by the red line in the figure). The predicted score for the treatment states 1 year after adoption (i.e., 2011) was calculated by subtracting the estimated 1-year effect (i.e., -2.26 points or -0.06 SD, see Table 3) from the observed 2011 score for the treatment states; the predicted scores for the treatment states in later years were calculated in the same way. The red line in Figure 1 thus shows what the NAEP scores would have been for the treatment states in the post-CCR years had these states not adopted the new standards—i.e., the counterfactual. As is clear from the figure, the grade 4 reading achievement in the treatment states would have improved significantly more after the adoption the new standards had the states continued with their old standards, thus reflecting negative effects of the new CCR standards. Graphic illustrations of the estimated effects for other subject and grade combinations are provided in Appendix C.

Figure 1. Observed Average NAEP Grade 4 Reading Scores for Treatment States Identified Based on the Prior Rigor Index and Their Predicted Scores in the Absence of CCR Standards



Note: Results presented in the figure are based on analysis of data from 17 treatment states and 12 comparison states identified based on the Prior Rigor Index.

Effects of CCR Standards on All Students Based on NAEP Subscale Scores

In addition to NAEP composite scores, we also conducted CITS analyses using the two NAEP reading subscales and five NAEP math subscales to explore whether the effects of CCR standards might differ for different subscales due perhaps to uneven alignment between the NAEP assessments and the CCR standards in different domains (Daro et al., 2015). Table 4 presents results for the two NAEP reading subscales (i.e., *gaining information* and *literary experience*) based on state classifications using the Prior Rigor Index for states' prior ELA standards. It shows that the effect estimates for the two reading subscales for grade 4 were similar in magnitude, and all estimates were either significant (p < .05) or marginally significant (p < .10). For grade 8, however, there appear to be more notable differences in the results for the two reading subscales. Specifically, while the effects of the CCR standards on 8th graders' reading achievement as measured by the *gaining information* subscale during the 7 years after the adoption of the standards were all small and non-significant (effects = -0.04~0.02 SD, p > .10), the effects on 8th graders' reading achievement as measured by the *literary experience* subscale were larger, particularly during the period of 3–7 years after the adoption of the new standards (effects = -0.09 ~ -0.08 SD), with the 3-year effect reaching statistical significance (p < .05) and the 5-year and 7-year effects both reaching marginal significance (p < .10).

Table 4. Estimated Effects of CCR Standards on Student Achievement As Measured by NAEP Reading Subscales, by Grade and Timing of Effect (With State Classification Based on the Prior Rigor Index)

Reading Subscale	Timing of effect (year after adoption)	N of states	N of observations	Estimate	Standard error	Effect in SD*	P-value
Grade 4							
Gaining information	1 year	29	336	-2.18	1.01	-0.06	0.031
Literary experience	1 year	29	336	-2.38	0.89	-0.07	0.007
Gaining information	3 years	29	336	-2.67	1.32	-0.07	0.043
Literary experience	3 years	29	336	-3.57	1.34	-0.10	0.008
Gaining information	5 years	29	336	-3.08	1.78	-0.08	0.083
Literary experience	5 years	29	336	-3.99	2.12	-0.11	0.059
Gaining information	7 years	29	336	-3.55	1.96	-0.09	0.069
Literary experience	7 years	29	336	-4.00	1.97	-0.11	0.042
Grade 8							
Gaining information	1 year	29	281	0.77	0.93	0.02	0.409
Literary experience	1 year	29	281	-0.20	0.96	-0.01	0.834
Gaining information	3 years	29	281	-0.29	1.11	-0.01	0.792
Literary experience	3 years	29	281	-2.55	1.22	-0.08	0.036
Gaining information	5 years	29	281	-0.81	1.69	-0.02	0.633
Literary experience	5 years	29	281	-2.95	1.65	-0.08	0.075

Gaining information	7 years	29	281	-1.48	1.76	-0.04	0.402
Literary experience	7 years	29	281	-3.30	1.75	-0.09	0.058

Note: * Effect in SD was computed by dividing the estimate in the original NAEP scale by the SD of the NAEP test scores for public school students for the relevant grade, subject, and year.

Table 5 presents CITS analysis results for the five NAEP math subscales (i.e., *algebra, data analysis, geometry, measurement*, and *number properties*) based on state classifications using the Prior Rigor Index for states' prior math standards. It shows that for grade 4, the effects of the CCR standards on the five NAEP math subscales in each post-CCR year examined were of similar size and were all non-significant (p> .10). For grade 8, the results for the five NAEP math subscales were also similar 1 year and 3 years after the adoption of the CCR standards, but varied more in later years. While the effects of the CCR standards differed by 0.05 SD or less and were non-significant across the five NAEP math subscales in earlier years, the effects varied more widely from -0.13 to -0.02 SDs 5 years after the adoption of the new standards and from -0.17 to -0.05 SDs 7 years after adoption. Relatedly, there were also differences in the statistical significance of the 5-year and 7-year effects across the five math subscales. The effects of the CCR standards on *measurement* and *number properties*, for example, were marginally significant (p < .10) 5 years after adoption and significant 7 years after adoption (p < .05). In contrast, the effects of the CCR standards on *algebra* was not significant in either years (p > .10).

Table 5. Estimated Effects of CCR Standards on Student Achievement As Measured by NAEP Math Subscales, by Grade and Timing of Effect (With State Classification Based on the Prior Rigor Index)

Math Subscale	Timing of effect (year after adoption)	N of states	N of observations	Estimate	Standard error	Effect in SD*	P-value
Grade 4							
Algebra	1 year	34	357	0.57	0.91	0.02	0.533
Data analysis	1 year	34	357	0.99	1.25	0.03	0.429
Geometry	1 year	34	357	0.71	1.09	0.02	0.517
Measurement	1 year	34	357	1.06	1.14	0.04	0.351
Number properties	1 year	34	357	-0.21	1.11	-0.01	0.850
Algebra	3 years	34	357	-0.72	0.99	-0.02	0.471
Data analysis	3 years	34	357	1.02	1.60	0.03	0.525
Geometry	3 years	34	357	0.13	1.55	0.00	0.935
Measurement	3 years	34	357	-1.11	1.65	-0.04	0.501
Number properties	3 years	34	357	-1.47	1.39	-0.05	0.291
Algebra	5 years	34	357	-1.20	1.41	-0.04	0.392
Data analysis	5 years	34	357	-1.84	1.98	-0.06	0.352
Geometry	5 years	34	357	-1.61	2.12	-0.05	0.447
Measurement	5 years	34	357	-1.87	2.17	-0.06	0.387
Number properties	5 years	34	357	-1.44	1.85	-0.05	0.439
Algebra	7 years	34	357	-2.09	1.74	-0.07	0.232
Data analysis	7 years	34	357	-3.26	2.16	-0.11	0.133
Geometry	7 years	34	357	-3.20	2.30	-0.10	0.164

Measurement	7 years	34	357	-3.06	2.28	-0.10	0.180
Number properties	7 years	34	357	-1.30	1.83	-0.04	0.480
Grade 8							
Algebra	1 year	34	381	0.44	0.99	0.01	0.655
Data analysis	1 year	34	381	1.10	1.18	0.03	0.348
Geometry	1 year	34	381	0.22	1.07	0.01	0.840
Measurement	1 year	34	381	-0.55	1.48	-0.02	0.712
Number properties	1 year	34	381	-0.55	0.97	-0.02	0.568
Algebra	3 years	34	381	-0.54	1.39	-0.01	0.697
Data analysis	3 years	34	381	-0.32	1.83	-0.01	0.862
Geometry	3 years	34	381	-1.22	1.66	-0.03	0.463
Measurement	3 years	34	381	-2.27	2.02	-0.06	0.260
Number properties	3 years	34	381	-1.80	1.33	-0.05	0.176
Algebra	5 years	34	381	-0.85	1.84	-0.02	0.643
Data analysis	5 years	34	381	-2.45	2.20	-0.07	0.266
Geometry	5 years	34	381	-2.70	2.23	-0.07	0.226
Measurement	5 years	34	381	-4.97	2.59	-0.13	0.055
Number properties	5 years	34	381	-2.97	1.56	-0.08	0.057
Algebra	7 years	34	381	-2.06	1.79	-0.05	0.250
Data analysis	7 years	34	381	-4.59	2.51	-0.12	0.067
Geometry	7 years	34	381	-4.64	2.45	-0.12	0.058
Measurement	7 years	34	381	-6.59	2.90	-0.17	0.023
Number properties	7 years	34	381	-3.67	1.56	-0.09	0.019

Note: * Effect in SD was computed by dividing the estimate in the original NAEP scale by the SD of the NAEP test scores for public school students for the relevant grade, subject, and year.

Table 6 presents parallel results based on state classifications using the Prior CCSS-Similarity Index. Similar to the results presented in Table 5, most of the effects of the CCR standards on the five NAEP math subscales shown in Table 6 were small and non-significant. There were some non-trivial differences in the results across different math subscales in certain post-CCR years, but generally the results did not exhibit clear patterns. Note that the results presented in Table 6 were based on a substantially smaller sample than the results presented in Table 5, and thus may contain more noise than the results in Table 5.

Table 6. Estimated Effects of CCR Standards on Student Achievement As Measured by NAEP Math Subscales, by Grade and Timing of Effect (With State Classification Based on the Prior CCSS-Similarity Index)

Math Subscale	Timing of effect (year after adoption)	N of states	N of observations	Estimate	Standard error	Effect in SD*	P-value
Grade 4							
Algebra	1 year	26	271	1.49	0.90	0.05	0.096
Data analysis	1 year	26	271	2.48	1.39	0.09	0.075
Geometry	1 year	26	271	0.96	1.22	0.03	0.434
Measurement	1 year	26	271	2.10	0.99	0.07	0.035
Number properties	1 year	26	271	1.14	1.22	0.04	0.351
Algebra	3 years	26	271	0.84	1.27	0.03	0.506
Data analysis	3 years	26	271	2.79	1.93	0.09	0.150
Geometry	3 years	26	271	-0.37	1.73	-0.01	0.829
Measurement	3 years	26	271	0.18	1.59	0.01	0.909
Number properties	3 years	26	271	-0.65	1.83	-0.02	0.721
Algebra	5 years	26	271	1.14	1.44	0.04	0.428
Data analysis	5 years	26	271	-0.05	2.22	0.00	0.984
Geometry	5 years	26	271	-2.64	2.10	-0.09	0.209
Measurement	5 years	26	271	-1.22	1.95	-0.04	0.533
Number properties	5 years	26	271	-0.41	1.91	-0.01	0.829
Algebra	7 years	26	271	-0.88	1.69	-0.03	0.603
Data analysis	7 years	26	271	-1.88	2.38	-0.06	0.429
Geometry	7 years	26	271	-3.76	2.04	-0.12	0.065
Measurement	7 years	26	271	-2.36	1.68	-0.08	0.161
Number properties	7 years	26	271	-0.35	1.76	-0.01	0.843
Grade 8							
Algebra	1 year	26	287	0.28	1.12	0.01	0.804
Data analysis	1 year	26	287	1.50	1.29	0.04	0.245
Geometry	1 year	26	287	-0.22	1.17	-0.01	0.851
Measurement	1 year	26	287	0.88	1.37	0.02	0.521
Number properties	1 year	26	287	1.28	1.03	0.04	0.211
Algebra	3 years	26	287	0.17	1.43	0.00	0.906
Data analysis	3 years	26	287	0.69	1.76	0.02	0.693
Geometry	3 years	26	287	-0.65	1.57	-0.02	0.682
Measurement	3 years	26	287	-0.33	1.68	-0.01	0.846
Number properties	3 years	26	287	0.85	1.38	0.02	0.540

Algebra	5 years	26	287	1.04	2.04	0.03	0.612
Data analysis	5 years	26	287	-1.42	2.51	-0.04	0.571
Geometry	5 years	26	287	-1.40	2.18	-0.04	0.520
Measurement	5 years	26	287	-1.48	2.57	-0.04	0.565
Number properties	5 years	26	287	-0.28	1.73	-0.01	0.872
Algebra	7 years	26	287	-0.21	1.98	-0.01	0.914
Data analysis	7 years	26	287	-4.25	2.84	-0.11	0.134
Geometry	7 years	26	287	-4.24	2.14	-0.11	0.048
Measurement	7 years	26	287	-3.67	2.76	-0.09	0.184
Number properties	7 years	26	287	-1.66	1.77	-0.04	0.348

Note: * Effect in SD was computed by dividing the estimate in the original NAEP scale by the SD of the NAEP test scores for public school students for the relevant grade, subject, and year.

Effects of CCR Standards on Student Subgroups

In this section, we present the effects of CCR standards on the achievement of a few key student subgroups, including SWDs, ELLs, Blacks, Hispanics, and students eligible for free- or reduced-price lunch (FRPL). To put these subgroup results in context, we also include in Table 7 the results for all students in the last column. One finding that is clear from the table is that the estimated effects of the CCR standards on the achievement of student subgroups varied more widely than the effects on the overall sample. While the effects for the overall sample ranged from -0.10 to 0.05 SDs across subjects, grades, and years, they tended to vary more widely for student subgroups, particularly for SWDs, ELLs, and Hispanics. The effects for ELLs, for example, ranged from -0.51 to 0.12 SDs, with 7 of the 24 effect estimates having an absolute value exceeding 0.20 SDs and 6 of the effects reaching statistical significance ($p < .05$) or marginal significance ($p<.10$). The range of effects for SWDs was narrower, but still substantial—from -0.24 to 0.14 SDs across subjects, grades, and years. In contrast, the effects for students eligible for FRPL had a much narrower range (-0.08 to 0.07 SDs), with only one effect estimate reaching marginal significance.

Table 7. Estimated Effects of CCR Standards on the Achievement of Key Student Subgroups as Measured by NAEP Composite Scores, by Subject, Grade, and Timing of Effect

Grade/subject	Timing of effect (year after adoption)	Effect in SD*					
		SWDs	ELLs	Blacks	Hispanics	FRPL	All
State classification based on the Prior Rigor Index							
Grade 4 reading	1 year	-0.04	-0.06	-0.09*	-0.11†	-0.03	-0.06*
Grade 4 reading	3 years	0.00	-0.05	-0.05	-0.06	0.00	-0.08*
Grade 4 reading	5 years	-0.05	-0.09	-0.10†	-0.18†	-0.02	-0.10†
Grade 4 reading	7 years	-0.03	-0.09	-0.11	-0.11	-0.01	-0.10*
Grade 8 reading	1 year	-0.06	0.12	0.06	0.09†	0.03	0.01
Grade 8 reading	3 years	-0.13	-0.02	0.03	0.01	-0.02	-0.04
Grade 8 reading	5 years	-0.16	-0.15	-0.02	0.00	-0.02	-0.05
Grade 8 reading	7 years	-0.24†	-0.15	-0.02	0.01	-0.01	-0.06

Grade 4 math	1 year	0.13*	0.04	0.06	0.06	0.07†	0.01
Grade 4 math	3 years	0.05	-0.07	0.02	0.03	0.01	-0.03
Grade 4 math	5 years	0.03	-0.13	0.01	0.03	-0.01	-0.05
Grade 4 math	7 years	0.04	-0.26	-0.04	0.01	-0.02	-0.07
Grade 8 math	1 year	0.02	-0.23*	0.01	-0.07	0.02	0.00
Grade 8 math	3 years	-0.09	-0.21	-0.04	-0.02	0.00	-0.03
Grade 8 math	5 years	-0.13	-0.45*	-0.08	-0.06	-0.04	-0.07
Grade 8 math	7 years	-0.18	-0.51*	-0.12	-0.09	-0.07	-0.10*
State classification based on the Prior CCSS-Similarity Index							
Grade 4 math	1 year	0.14*	-0.04	0.06	0.05	0.06	0.05
Grade 4 math	3 years	0.03	-0.23*	0.01	0.02	0.00	0.00
Grade 4 math	5 years	-0.01	-0.29†	-0.03	-0.02	-0.05	-0.02
Grade 4 math	7 years	-0.01	-0.36*	-0.08	-0.06	-0.08	-0.05
Grade 8 math	1 year	-0.02	-0.05	0.01	-0.12*	0.04	0.02
Grade 8 math	3 years	-0.08	0.02	0.03	-0.08	0.04	0.00
Grade 8 math	5 years	-0.12	-0.02	-0.04	-0.11	0.02	-0.01
Grade 8 math	7 years	-0.24†	-0.09	-0.09	-0.18*	-0.03	-0.06

Note: See Appendix D for the sample size for each subgroup analysis.
† p < .10; * p < .05.

Discussions

This paper presents findings about the effects of states' implementation of the new CCR standards on student achievement as measured by NAEP. Contrary to our expectation, we found that the CCR standards had significant negative effects on 4th graders' reading achievement during the 7 years after the adoption of the new standards, and had a significant negative effect on 8th graders' math achievement 7 years after adoption based on analyses of NAEP composite scores. The size of these negative effects, however, was generally small, ranging from -0.10 to -0.06 SDs.

Analyses of NAEP subscale scores show that the effects of the CCR standards on the two NAEP reading subscales were similar for grade 4, but differed for grade 8, with significant negative effects on 8th graders' performance on the *literary experience* subscale and smaller non-significant effects on their performance on the *gaining information* subscale over the time period examined. Similarly, we found that the effects of the CCR standards on the five NAEP math subscales were similar for grade 4 but differed for grade 8, particularly in later years (i.e., 5 years and 7 years after adoption), according to CITS analyses with state classifications based on the Prior Rigor Index. CITS analyses with state classifications based on the Prior CCSS-Similarity Index also revealed non-trivial differences in the effects of the CCR standards on different math subscales in certain post-CCR years, but generally the results did not exhibit clear patterns. Finally, our subgroup analyses show that the effects of the CCR standards on the achievement of certain student subgroups--SWDs, ELLs, and Hispanics in particular—varied more widely across subjects, grades, and years, and had a much larger size (in the negative direction) than the effects for the overall sample.

There are a number of potential reasons for the lack of positive effects that we had hoped to find in this study. One potential reason has to do with the limitations of this study. Given the timing of CCR standards implementation across states, a true "no-treatment" comparison group is not available for the study. Instead, we constructed the treatment and comparison groups based on the natural variation in the quality of states' prior content standards among states that adopted the CCR standards in 2010, which allows for an indirect test of the effects of CCR standards through CITS analyses. The CITS estimates of the treatment effects thus would represent unbiased estimates of the effects of the new CCR standards (vs. the old pre-CCR standards) for the treatment states only under certain conditions—i.e., if the CCR standards had no effect on student achievement in the comparison states, and if other factors associated with student achievement affected achievement in treatment and comparison states in similar ways. If the CCR standards had a positive effect on student achievement in the comparison states, then the CITS estimate would provide a lower bound of the true effect of CCR standards for the treatment states. Conversely, if the CCR standards had a negative effect on student achievement in the comparison states, then the CITS estimate would provide an upper bound of the true effect of CCR standards for the treatment states.

Another limitation of the study is that our measures of student achievement—NAEP scores—are not perfectly aligned with the CCR standards. Over 20 percent of the items on the 2015 NAEP grade 4 math assessment and 13 percent of the items on the 2015 NAEP grade 8 math assessment, for example, are not covered by the CCSS for the relevant grade or below (Daro, Hughes, & Stancavage, 2015). Therefore, the NAEP assessments may be less sensitive to changes in student achievement induced by the implementation of CCR standards than assessments more closely aligned with the new standards.

Given these limitations, findings from this study need to be interpreted with caution. When interpreting the findings from the study, it is also important to bear in mind that the study was designed to estimate the effects of CCR standards 1 year, 3 years, 5 years, and 7 years after the adoption, which are not the same as the effects of 1 year of full implementation, 3 years of full implementation, 5 years of full implementation, and 7 years of full implementation. In fact, it took most states 3 to 5 years to fully transition from the old standards and the new standards (see Appendix B). For states that adopted the CCR standards in 2010, almost half of the states were still in their first year of full implementation when they participated in the 2015 NAEP assessments—5 years after adopting the new standards. Thus, findings from this study largely reflect early effects of states' implementation of the CCR Standards during the transition period, which may be different from the effects after the new standards became fully implemented.

Another potential explanation for the lack of positive effects during the transition period lies in the multitude of challenges that states, districts, schools, and teachers experienced during the transition from the old standards to the new standards. As mentioned in the literature review section of the paper, for example, lack of adequate resources was a major challenge faced by the majority of states and districts in CCSS implementation based on recent surveys. Most CCSS states and districts also reported major challenges in providing sufficient high-quality professional development to help teachers implement the new standards, which was much needed given that the majority of teachers did not feel very prepared to teach the CCSS. In addition, most states and districts implementing the CCSS indicated that the lack of CCSS-aligned curricular and instructional materials posed a major challenge for implementing the new standards, which is not surprising given that recent textbook analyses revealed substantial misalignment between popular textbooks claimed to be CCSS-aligned and the CCSS.

All these challenges may have hindered successful statewide transition from old standards to the new standards, which is a massive undertaking and a highly challenging endeavor that requires concerted efforts and support at multiple levels. Well-designed standards are essential, but not sufficient, for the success of standards-based reform—if the standards are not well implemented and not accompanied by aligned assessments tied to solid accountability systems, they would not automatically translate to improved student test scores. A great deal needs to happen between well-crafted standards and improved student achievement. When and to what extent the potential of the new standards will be realized will depend on how quickly and how adequately all the needed supports will be put in place.

References

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D.H. (2009). *Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007* (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Bay-Williams, J., Duffett, A., & Griffith, D. (2016). *Common Core math in the K-8 classroom: Results from a national teacher survey.* Washington, DC: Thomas B. Fordham Institute. Retrieved January 9, 2017, from https://edexcellence.net/publications/common-core-math-in-the-k-8-classroom-results-from-a-national-teacher-survey

Bertrand, M., Duflo, E., & Mullainathan, S. (2002). *How much should we trust differences-in-differences estimates?* (No. w8841). National Bureau of Economic Research.

Carmichael, S. B., Martino, G., Porter-Magee, K., & Wilson, W. S. (2010). *The state of the state standards—and the Common Core—in 2010.* Washington, DC: Thomas B. Fordham Institute.

Daro, P., Hughes, G. B., & Stancavage, F. (2015). *Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics*. Washington, D.C.: American Institutes for Research.

Dee, T. S., & Jacob, B. A. (2011). The impact of no Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3) 418-446.

Dee, T.S., Jacob, B., & Schwartz, N.L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis,* 35(2), 252-279.

Gemertz, C., (September 28, 2015). The Common Core explained. *Education Week*. Retrieved May 24, 2016, from http://www.edweek.org/ew/issues/common-core-state-standards/.

Heitin, L. (March 17, 2015). Review of math programs comes under fire. *Education Week*. Retrieved January 14, 2017, from http://www.edweek.org/ew/articles/2015/03/18/review-of-math-programs-comes-under-fire.html.

Heitin, L. (August 30, 2016). Common-Core reading materials get mixed results in first major review. *Education Week*. Retrieved January 14, 2017, from http://blogs.edweek.org/edweek/curriculum/2016/08/common-core_materials_for_reading_get_mixed_reviews.html.

Kane, T. J., Owens, A. M., Marinell, W. H., Thal, D. R. C., & Staiger, D. O. (2016). *Teaching Higher: Educators' Perspectives on Common Core Implementation*. Retrieved November 26, 2016, from http://cepr.harvard.edu/publications/teaching-higher-educators-perspectives-common-core-implementation

Kober, N., McIntosh, S., & Rentner, D.S. (2013) *Year 3 of implementing the common core state standards: Professional development for teachers and principals.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://cep-dc.org/displayDocument.cfm?DocumentID=422

Kober, N. & Rentner, D.S. (2011a). *Common Core State Standards: Progress and challenges in school districts' implementation.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=374

Kober, N. & Rentner, D.S. (2011b). *States' progress and challenges in implementing common core state standards.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://cep-dc.org/displayDocument.cfm?DocumentID=343

Kober, N. & Rentner, D.S. (2012). *Year 2 of implementing the common core state standards: States' progress and challenges.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=391

Loveless, T. (2014). 2014 *Brown Center report on American education: How well are American students learning? Part III: A progress report on the Common Core*. Washington, DC: The Brookings Institution. Retrieved January 15, 2017, from https://www.brookings.edu/research/a-progress-report-on-the-common-core/

Loveless, T. (2015). 2015 *Brown Center report on American education: How well are American students learning? Part II: Measuring effects of the Common Core*. Washington, DC: The Brookings Institution. Retrieved January 15, 2017, from https://www.brookings.edu/research/measuring-effects-of-the-common-core/

National Center for Education Statistics (February 2010). *2002 Educational Longitudinal Study: Second Followup, 2006*, previously unpublished tabulations. [Note: cited in https://www2.ed.gov/policy/elsec/leg/blueprint/college-career-ready.pdf]

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for education reform*. Washington, DC: U.S. Government Printing Office.

Polikoff, M. S. (2015). How well aligned are textbooks to the Common Core Standards in mathematics? *American Educational Research Journal, 52*(6), 1185-1211.

Polikoff, M. S., Hardaway, T., Marsh, J. A., & Plank, D. N. (2016). Who is opposed to the Common Core and Why? In *Education Researcher*, *45*(3), 263-266.

Scholastic & the Bill & Melinda Gates Foundation. (2014). *Teachers' views on the Common Core State Standards one year later.* Retrieved December 2, 2016, from http://www.scholastic.com/primarysources/PrimarySources-2014update.pdf

Shadish, W. R., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Peterson, P. E., Herderson, M. B., West, M. R., & Barrows, S. (2016). Ten-year trends in public opinion from the Ed Next Poll. *Education Next, 17*(1), 8–28.

Rentner, D.S. (2013). *Year 3 of implementing the common core state standards: An overview of states' progress and challenges.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=421

Rentner, D.S., & Kober, N. (2014). *Common core state standards in 2014: Districts' perceptions, progress, and challenges.* Washington, DC: Center on Education Policy. Retrieved January 8, 2017, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=440

Schmidt, W. H., & Houang, R. T. (2012). Curricular Coherence and the Common Core State Standards for Mathematics. *Educational Researcher*, *41*(8), 294-308.

U.S. Department of Education. (2009). *Race to the Top Program executive summary.* Washington, DC: Author. Retrieved November 6, 2016, from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2010). *College-and-career-ready students.* Washington, DC: Author. Retrieved November 6, 2016, from www2.ed.gov/policy/elsec/leg/blueprint/college-career-ready.pdf

U.S. Department of Education. (2012). *ESEA flexibility (updated June 7, 2012).* Washington, DC: Author. Retrieved November 6, 2016, from https://www.ed.gov/sites/default/files/esea-flexibility-request-w3.doc

Walker, T. (2013, September 12). NEA poll: majority of educators support the common core state standards. *NEA Today.* Retrieved November 26, 2016, from http://neatoday.org/2013/09/12/nea-poll-majority-of-educators-support-the-common-core-state-standards-2/

Wong, M., Cook, T. D., & Steiner, P. M. (2015). Adding design elements to improve time series designs: No Child Left Behind as an example of causal pattern-matching. *Journal of Research on Educational Effectiveness, 8*(2), 245-279.

Xu, Z., & Cepa, K. (2018). *Getting college-ready during state transition toward the Common Core State Standards*. *Teachers College Record, 120*(6), 1–36.

Appendix A. Year of CCR Standards Adoption and the Quality of Prior Content Standards for Each State

State	Year CCR standards in ELA adopted	Year CCR standards in math adopted	Prior Rigor Index for prior ELA standards	Prior Rigor Index for prior math standards	Prior CCSS-Similarity Index for prior math standards
Alabama	2010	2010	6	5	5
Alaska*	2012	2012	1	3	3
Arizona	2010	2010	5	4	1
Arkansas	2010	2010	3	3	3
California	2010	2010	7	7	5
Colorado	2009	2009	6	3	3
Connecticut	2010	2010	2	3	2
Delaware	2010	2010	2	5	3
DC	2010	2010	7	7	NA
Florida	2010	2010	5	7	5
Georgia	2010	2010	6	6	5
Hawaii	2010	2010	4	3	3
Idaho	2011	2011	4	5	4
Illinois	2010	2010	3	1	2
Indiana	2010	2010	7	7	5
Iowa	2010	2010	1	3	1
Kansas	2010	2010	4	1	1
Kentucky	2010	2010	3	2	1
Louisiana	2010	2010	6	3	1
Maine	2011	2011	4	3	2
Maryland	2010	2010	4	3	2
Massachusetts	2010	2010	7	6	3
Michigan	2010	2010	2	6	5
Minnesota*	2010	2007	4	5	5
Mississippi	2010	2010	3	4	5
Missouri	2010	2010	3	2	2
Montana	2011	2011	2	0	2
Nebraska*	2014	2015	1	3	2
Nevada	2010	2010	4	4	1
New Hampshire	2010	2010	4	3	2
New Jersey	2010	2010	4	4	1
New Mexico	2010	2010	4	4	3
New York	2010	2010	3	5	3
North Carolina	2010	2010	3	3	3
North Dakota	2011	2011	2	4	4
Ohio	2010	2010	4	3	3

(continued)

State	Year CCR standards in ELA adopted	Year CCR standards in math adopted	Prior Rigor Index for prior ELA standards	Prior Rigor Index for prior math standards	Prior CCSS-Similarity Index for prior math standards
Oklahoma	2010	2010	5	5	5
Oregon	2010	2010	4	5	4
Pennsylvania	2010	2010	3	1	3
Rhode Island	2010	2010	3	3	1
South Carolina	2010	2010	3	3	3
South Dakota	2010	2010	4	3	4
Tennessee	2010	2010	6	3	4
Texas*	2008	2008	6	4	3
Utah	2010	2010	4	6	4
Vermont	2010	2010	2	1	3
Virginia*	2010	2009	6	4	2
Washington	2011	2011	4	7	5
West Virginia	2010	2010	3	5	3
Wisconsin	2010	2010	3	1	1
Wyoming	2012	2012	3	1	2

Notes: *Alaska, Nebraska, Texas, and Virginia adopted their own CCR standards in both ELA and mathematics. Minnesota adopted CCSS in ELA but not math. All other states adopted CCSS standards in both subjects.
The Prior Rigor Index for a state's 2010 content standards is on a 0–7 point scale, with 7 presenting the highest rigor (Carmichael et al., 2010).
The original measure of the similarity between a state's 2009 content standards and CCSS in mathematics is on a 0–1000 point scale (Schmidt & Houang, 2012). For this study, we used a 1-5 version of the measure, with 1 representing "least like CCSS" and 5 "most like CCSS", based on the Schmidt and Houang's categorization.

Appendix B. Timeline of CCR Standards Adoption and Implementation Across States

Table B.1. Number of States by the Year CCR Standards in ELA Were Adopted and the First Year CCR Standards in ELA Were Expected to be Fully Implemented

Year CCR standards adopted	First year CCR ELA standards expected to be fully implemented								Total
	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	
2007									0
2008	1								1
2009					1				1
2010			1	5	16	18	1		41
2011					4	1			5
2012						2			2
2013									0
2014								1	1
2015									0
Total	**1**		**1**	**5**	**21**	**21**	**1**	**1**	**51**

Sources: State websites, state documents, communications with state officials, and Elementary and Secondary Education Act (ESEA) flexibility requests.
Note: The "first year CCR ELA standards expected to be fully implemented" is the first year in which a state expected all teachers in all districts in the state to be integrating CCR standards in ELA into classroom instruction.

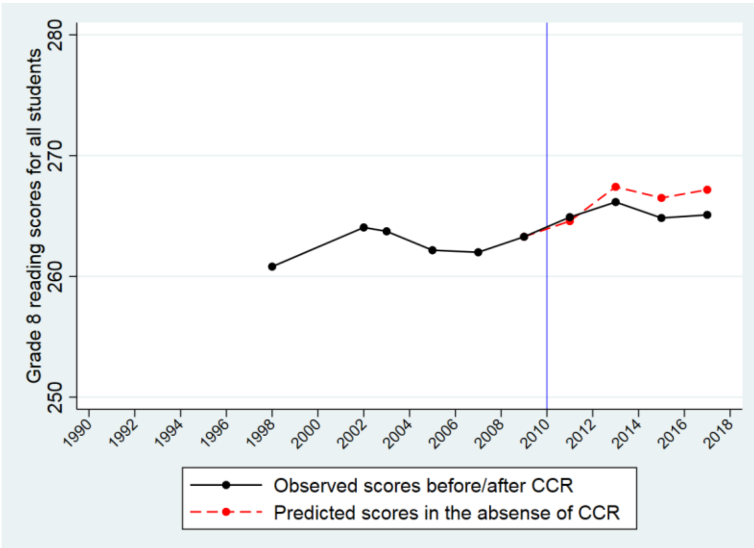
Table B.2. Number of States by the Year CCR Standards in Mathematics Were Adopted and the First Year CCR Standards in Mathematics Were Expected to be Fully Implemented

Year CCR standards adopted	First year CCR mathematics standards expected to be fully implemented								Total
	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	
2007		1							1
2008	1								1
2009				1	1				2
2010			1	4	15	18	1		39
2011					4	1			5
2012						2			2
2013									
2014									
2015								1	1
Total	**1**	**1**	**1**	**5**	**20**	**21**	**1**	**1**	**51**

Sources: State websites, state documents, communications with state officials, and Elementary and Secondary Education Act (ESEA) flexibility requests.
Note: The "first year CCR mathematics standards expected to be fully implemented" is the first year in which a state expected all teachers in all districts in the state to be integrating CCR standards in mathematics into classroom instruction.

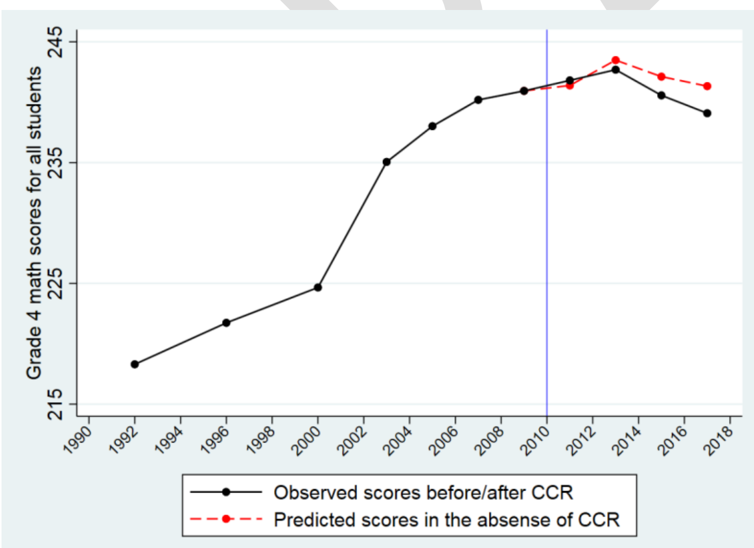
Appendix C. Observed and Predicted NAEP Scores Before and After the Adoption of CCR Standards in Treatment States

Figure C.1. Observed Average NAEP Grade 8 Reading Scores for Treatment States Identified Based on the Prior Rigor Index and Their Predicted Scores in the Absence of CCR Standards



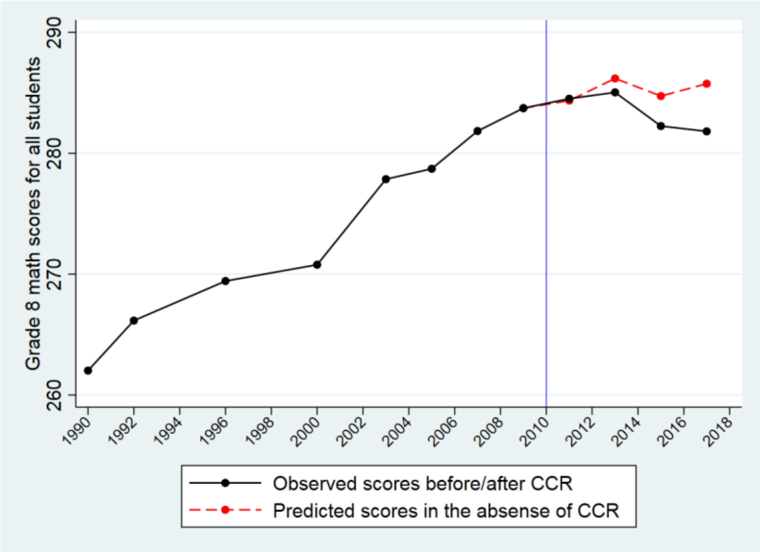
Note: Results presented in the figure are based on analysis of data from 17 treatment states and 12 comparison states identified based on the Prior Rigor Index.

Figure C.2. Observed Average NAEP Grade 4 Mathematics Scores for Treatment States Identified Based on the Prior Rigor Index and Their Predicted Scores in the Absence of CCR Standards



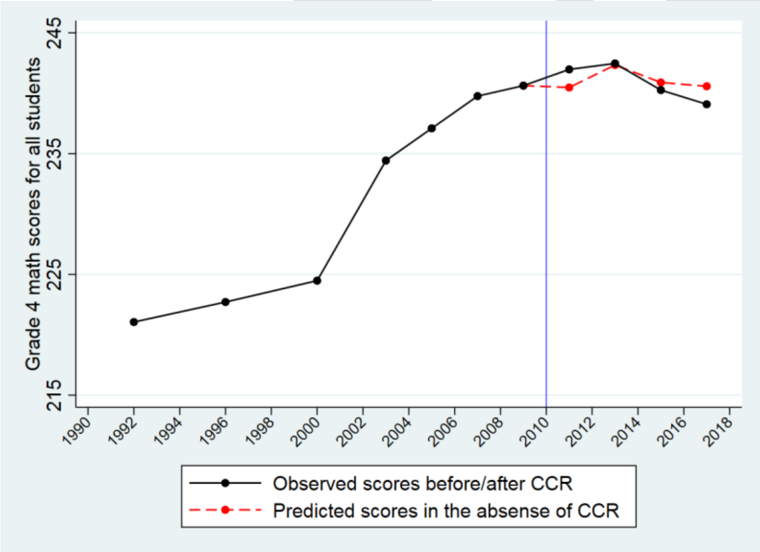
Note: Results presented in the figure are based on analysis of data from 20 treatment states and 14 comparison states identified based on the Prior Rigor Index.

Figure C.3. Observed Average NAEP Grade 8 Mathematics Scores for Treatment States Identified Based on the Prior Rigor Index and Their Predicted Scores in the Absence of CCR Standards



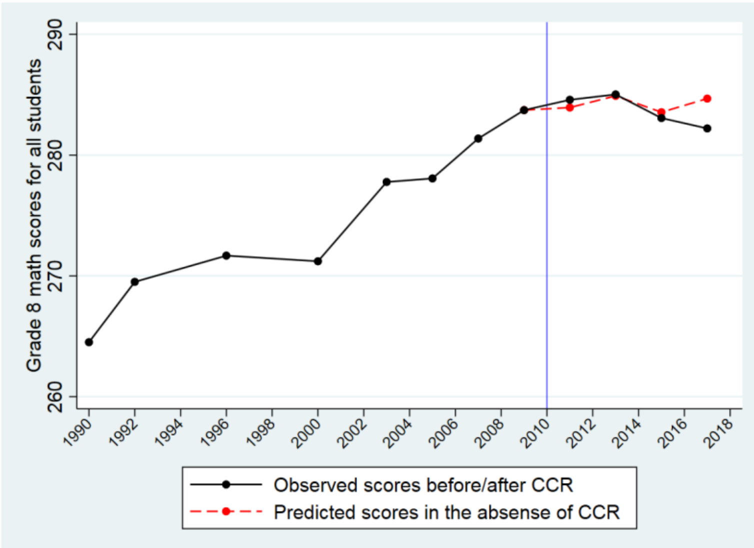
Note: Results presented in the figure are based on analysis of data from 20 treatment states and 14 comparison states identified based on the Prior Rigor Index.

Figure C.4. Observed Average NAEP Grade 4 Mathematics Scores for Treatment States Identified Based on the Prior CCSS-Similarity Index and Their Predicted Scores in the Absence of CCR Standards



Note: Results presented in the figure are based on analysis of data from 14 treatment states and 12 comparison states identified based on the Prior CCSS-Similarity Index.

Figure C.5. Observed Average NAEP Grade 8 Mathematics Scores for Treatment States Identified Based on the Prior CCSS-Similarity Index and Their Predicted Scores in the Absence of CCR Standards



Note: Results presented in the figure are based on analysis of data from 14 treatment states and 12 comparison states identified based on the Prior CCSS-Similarity Index.

Appendix D. Sample Size for Subgroup Analyses of the Effects of CCR Standards on Student Achievement

Grade/subject	Sample size unit	Subgroup Analyses				
		SWDs	ELLs	Blacks	Hispanics	FRPL
State classification based on the Prior Rigor Index						
Grade 4 reading	N of states	29	27	29	27	29
	N of observations	277	195	328	259	284
Grade 8 reading	N of states	29	23	29	27	29
	N of observations	276	135	273	222	281
Grade 4 math	N of states	34	33	34	32	34
	N of observations	298	233	324	292	328
Grade 8 math	N of states	34	27	32	32	34
	N of observations	297	173	328	277	326
State classification based on the Prior CCSS-Similarity Index						
Grade 4 math	N of states	26	25	26	26	26
	N of observations	227	182	249	232	250
Grade 8 math	N of states	26	21	24	26	26
	N of observations	227	130	253	222	247