

Is Common Core “Working”? And Where Does Common Core Research Go From Here?

Morgan S. Polikoff

University of Southern California

Five years into the Common Core initiative, researchers and the general public are interested to know whether and how the standards are “working.” In this introduction to the special topic, I discuss the state of the literature on these questions and offer suggestions for what I view as the most important work moving forward.

Keywords: *educational policy, instructional practices, mixed methods, policy*

FIVE years into the Common Core “experiment,” what do we know about the implementation and effects of the standards? Answering this question was the primary goal of this special topic. The answer, it turns out, is not as satisfying as we might like. In this article, I review both the articles that were published under the topic and my broader sense of where the literature is and where it needs to go. This introduction is organized under these two key questions. However, I start with the impact question because I think it is of more interest to policy makers.

What Is the Effect of Common Core Standards on Student Outcomes?

The million-dollar question for the Common Core Standards (CCS) in the public eye is probably whether or to what extent it is “working” to make students more college and career ready. This is a very hard question to answer, which is perhaps why none of the papers in this special topic address it. Descriptively, many jumped on the rare 2013–2015 National Assessment of Educational Progress (NAEP) declines to imply or argue that the standards were not working (e.g., Burris, 2015). However, these crude analyses (if they can be called that) did not even attempt to control for key design issues that might affect the validity of the conclusion that Common Core was not “working.”

Loveless (2014, 2016) has investigated this question using state NAEP. His work starts to address some of the design issues that affect the crude mean score change analyses mentioned above. In the 2014 analysis, he examined the NAEP gains for states with standards most and least similar to the CCS between 2009 and 2013, finding essentially no differences in gains. Here, he was testing the argument put forth by Schmidt and Houang (2012) that the states with standards

that were more “Common Core–like” saw greater NAEP gains prior to the adoption of the standards. He also analyzed the NAEP gains for states based on an index of state implementation of the CCS, finding slightly larger gains in higher implementation states (though these differences were not statistically significant). Here, he was testing the idea that states that were implementing the standards more fully might see greater gains. In the 2016 analysis, he updated the work using 2015 NAEP results. Focusing on the implementation index, he again found no evidence that high implementers were seeing greater achievement gains post-CCS.

Loveless’s analyses are the only of which I am aware that really attempt to get closer to an estimate of the Common Core’s impact on learning; however, the Institute of Education Sciences’ Center on Standards, Alignment, Instruction, and Learning (C-SAIL; on which I am a co–principal investigator) is addressing this question to some extent in its longitudinal impact study, the preliminary results of which should be available soon. The problems with Loveless’s analyses illustrate the general problems with this kind of work and perhaps suggest why few have attempted it.

First, of course, there are all manner of research design issues. Common Core was not randomly assigned to states, so the best that we can do is a quasi-experimental analysis. Within-state analyses are certainly out of the question because Common Core was either implemented or not in each state; thus, you should not put stock in analyses suggesting that within-state gains in test scores are indicative that the standards are working (see Nix, 2016). Loveless’s investigations are a sort of informal difference-in-differences analysis, where the pre- and post-NAEP changes are compared for CCS and non-CCS states. Of course, this could have been formalized with a regression model, which would have improved it somewhat. It would have been somewhat



more sophisticated to add multiple pre- and post-CCS time points, in which case Loveless might have had something like a comparative interrupted time series (CITS) analysis. Recent research suggests that CITS analyses can identify causal impacts when well designed and implemented (St. Clair, Hallberg, & Cook, 2016).

Even if he had used multiple time points, however, there are other challenges with a CITS design in this case. For instance, when did “treatment” begin in the case of Common Core? Most states adopted the standards in 2010 in response to *Race to the Top* (LaVenia, Cohen-Vogel, & Lang, 2015), but does adoption by a state legislature really mean that the standards are being implemented? What if, as is the case in California, a CCS-aligned assessment was not implemented until 2013–2014 and English language arts textbooks were not adopted by the state until 2014–2015? For our C-SAIL project, we asked states for their timelines for standards implementation, and 41 states replied that their college- and career-readiness standards were not fully implemented until either 2013–2014 or 2014–2015 (C-SAIL, 2016). If this is true, it may not make sense that the implementation date in Loveless’s analysis was 2010. Also, what about the comparison states? All 50 states have content standards, and non-CCS states each have state-specific standards. Furthermore, some non-CCS states claim that they had adopted college- and career-readiness standards prior to CCS adoption; for instance, on our C-SAIL map, the earliest adopter of college- and career-readiness standards is Texas, in 2009–2010. At a minimum, the interpretation of any of these potential quasi-experimental analyses would therefore be “compared to the standards adopted in other states.”

Second, and related, there are endogeneity concerns with trying to identify the impacts of CCS. States that adopted the standards certainly differed from those that did not adopt the standards. For instance, states that did not adopt the CCS may have been more politically independent or in better financial shape at the time of *Race to the Top*, or they may have had existing standards that they thought were particularly strong. Whether these preexisting differences are also related to outcomes is not obvious, but certainly it could be the case that states’ decisions about CCS adoption could have been related to the quality of their education system at the time of the decision. In this case, efforts to disentangle the effects of the standards would be more fraught. States that slow-played the implementation of the standards, such as California, may also be systematically different from those that implemented the standards more expeditiously. Of course, a few states also adopted and then unadopted the standards (not to mention the states that either modified the standards by adding content or modified them several years later in a formal revision). Endogeneity issues are here too—at a minimum, the political climate in these states clearly differed from the states that did not unadopt. But there are also obvious analytical issues with including these states.

Finally, many data issues plague efforts to analyze the impact question. What outcome measure should be used? Given that cross-state analyses are required, the NAEP is the default—it is the only test that is administered nationwide and in state-representative samples. But there are several issues here. For one, the CCS are not perfectly aligned to NAEP (Hughes, Daro, Holtzman, & Middleton, 2013); given this, would we expect to see the CCS have an impact on NAEP scores? Even if we would expect this, recent work suggests that student-level NAEP data (which Loveless and most others utilizing NAEP data do not often use) are optimal for correct inferences (Chingos, 2015).

Test scores are also a narrow outcome measure, especially when we know that their predictive power for future success, while undeniable, is imperfect (e.g., Angrist, Cohodes, Dynarski, Pathak, & Walters, 2016; Dobbie & Fryer, 2016). The CCS are billed as college- and career-readiness standards—might we not want to evaluate their impact on outcomes of this sort? Measuring college readiness is fraught with data challenges (e.g., Porter & Polikoff, 2012). Certainly we can use the SAT or ACT, but these are only sometimes administered statewide, meaning that there is student selection into these exams. Measuring student enrollment or persistence in college is challenging due to student mobility and data availability issues; plus, these outcomes are highly affected by labor market and macroeconomic conditions. Career readiness is even more difficult to measure at scale. The consequence of these data availability issues is that test scores in general and NAEP in particular become the outcome variable in virtually all analyses of the impact of the CCS.

The upshot of these issues—and there are more—is that no analysis of which I am aware provides convincing causal evidence of the impact of the CCS on any student outcome. Furthermore, it is not obvious to me that such an analysis is even possible. Even if it is possible, the outcomes that would be used for such an analysis would almost certainly be quite narrow. This may be a political challenge for the standards moving forward—if we can never convincingly answer the “did it work” question, critics may use this lack of evidence against the standards. Perhaps this was an oversight on the part of the CCS’s authors, or perhaps it was an inevitability of any attempt to institute national or nearly national standards in the United States, but it is a reality that CCS proponents must deal with.

How—and How Well—Are the CCS Being Implemented?

While the impact question is perhaps of more interest to the general public, the implementation question is likely more interesting and relevant to educators and the research community. Here is where the articles in the special topic make an important contribution.

Measuring CCS Implementation

At the most basic level, it is not obvious how to measure whether teachers are actually implementing the standards. It would be simple to survey representative samples of teachers in CCS states and ask them, “To what extent are you implementing the Common Core Standards?” This would not be a particularly effective research strategy, for many reasons, not the least of which is that teachers in CCS states undoubtedly know that they are supposed to be implementing the standards and are thus likely to respond favorably to such a question (Schwarz & Oyserman, 2001).

More thoughtful and specific questions could be constructed and used, and RAND and the Center on Education Policy (among others) recently used state-representative surveys of this sort to gauge implementation (e.g., Center on Education Policy, 2016b; Opfer, Kaufman, & Thompson, 2016). These surveys include questions that probe practices that are or are not aligned with standards, where the teachers would be less likely to know what the “correct” answer is unless they were familiar with the standards. This work has found that, even as of 2015, large proportions of teachers in both subjects (mathematics and English language arts/literacy) have misconceptions about what the CCS are calling for (in terms of content and practices), suggesting that their instruction is likely to be questionably aligned at best (Opfer et al., 2016).

Even more sophisticated survey methods for gauging standards alignment exist and could be used. For example, I analyzed teachers’ instructional alignment and changes over time using the Surveys of Enacted Curriculum (Polikoff, 2012), and we recently updated the surveys to better match the CCS and other college- and career-readiness standards as part of our C-SAIL work. The surveys ask teachers about their topic and cognitive demand coverage and then compare their responses to content analyses of the standards—these surveys should be even more difficult to game, but because they ask about fine-grained practices, they may be more difficult to reliably report on than broader questions. The same kinds of questions and issues could apply to interview studies (e.g., Center on Education Policy, 2016a), though the challenges of obtaining representative samples would, of course, also be an issue if interviews were used.

While there is good evidence that teachers can validly report on certain dimensions of their instruction, observations by external raters may provide the most valid inferences about teachers’ enactment of standards-aligned practices. Here too, however, are challenges. First, the “alignment” question is not one that can be satisfactorily answered by observing one or a few lessons. To know if a teacher’s instruction is aligned with standards, we must know what she or he is teaching throughout the school year—or at least a reasonable portion thereof. Of course, it is infeasible to observe a classroom every day for a full year.

Second, we may need to develop appropriate instrumentation to rate the lessons that we observe, since observational protocols that existed prior to the CCS may not match well with the standards. Here, the article in this special topic by Stein and colleagues (Stein, Correnti, Moore, Russell, & Kelly, 2017) can offer a useful template. Their work demonstrates not only a new tool that can be used to measure mathematics instruction in the vein of the CCS but also a thoughtful, theory-driven approach to the development of such measures. Future scholars working to extend their work would be well advised to take such a sophisticated approach to the development of instrumentation.

Process of Implementation

The work in the special topic also makes clear that studies of CCS implementation will be strongest when they carefully attend to the process of implementation. Hodge, Salloum, and Benko (2016) and Supovitz, Fink, and Newman (2016) both focus on the ways that networks play a major role in CCS implementation. In Supovitz and colleagues’ analysis, they looked inside schools at the networks that teachers used to access knowledge about Common Core. They found that more knowledgeable teachers were more likely to receive requests (in mathematics, not in English language arts) and more likely to go outside the school for additional knowledge. They also found that coaches and administrators were playing an important role in teachers’ social networks. These are positive findings suggesting that teachers may be targeting some of the right individuals to gain additional knowledge. They also indicate the importance of developing some internal expertise about new standards within each school if the standards are to take root. Whether the new knowledge that teachers can gain from colleagues translates into changed practice is an important next step for this line of work.

Hodge and colleagues investigate some of the same questions but at the state level, looking at the resources that states are providing to teachers to implement the standards and the role of networks in spreading the standards broadly. They find suggestive evidence that the CCS are leading states to share external resources, as CCS states are more likely than non-CCS states to link to other states, share instructional resources, and offer professional development opportunities through their websites. This lends some support to the theory of action for national standards in general and CCS in particular. Again, this is a promising result, but the question of take-up and impact on instruction is an important next step.

The remaining two papers in the special topic, by Reynolds and Goodwin (2016) and Herman, Epstein, and Leon (2016), begin to address the within-classroom processes of CCS implementation. Reynolds and Goodwin look inside tutoring sessions to investigate what makes literacy tutoring effective. Such individual-guided instruction may

be especially important in the CCS era, as the standards call for close reading of grade-level texts with “scaffolding as needed.” Their analysis finds that certain kinds of scaffolds—in particular motivational scaffolds—are associated with greater knowledge gains than other scaffolds. Their work is not the last word on the topic of how to effectively scaffold students’ reading of complex texts in the CCS era, but it points the way toward future research on the best approaches to support students in what is undoubtedly one of the greatest areas of need in CCS English language arts instruction.

Herman, Epstein, and Leon take a within-classroom perspective but for the purposes of an intervention designed to improve instruction and student learning. Their intervention work in two states highlighted several findings that are germane to CCS implementation more broadly. First, their artifacts, logs, and surveys indicated that teachers were able to implement the intervention with fidelity, suggesting that the kind of collaborative approaches to standards implementation holds promise for encouraging teachers to change their practice. Second, however, these initiatives were associated with achievement gains in only one of the settings, and these gains were somewhat modest in magnitude given the intensive nature of the intervention. This suggests that even instructional changes in line with the CCS might not directly result in improved student learning as measured by state tests—a challenge for maintaining support for the standards and improving implementation over time. These findings, if they were to be replicated in other settings where implementation is improving, may also raise concerns about the instructional sensitivity of state tests (Polikoff, 2010).

Together, these four studies demonstrate the kind of theory-driven, multiple-methods work that can drive the field forward toward more effective implementation research. While none of the studies, on its own, answers the question “Is Common Core working?” the research collectively points toward areas of relative strength and weakness in CCS implementation and its impact on teachers and students. With the important measurement work done by Stein and colleagues, the five articles show that the best CCS research will be complex and nuanced and will take a highly process-oriented approach to understanding how and where the standards are being effectively implemented.

Where Do We Go From Here?

Common Core remains the law of the land in more than 40 states, and while opposition to the standards remains strong—particularly among Republicans who see it as connect to President Obama (Polikoff, Hardaway, Marsh, & Plank, 2016)—the standards are likely to remain in place for the foreseeable future in many states. Given this reality, there will continue to be a need for evidence about the standards and their effects. I believe the following priorities might guide such research.

First, there will continue to be a desire to answer the impact question, and researchers should do the best that they can to meet this demand. This likely means quasi-experimental methods such as the CITS applied to NAEP data. But it also means creatively identifying and using other potential outcome measures related to college and career readiness. Regardless of the methods and the data used, authors must be humble about the challenges of the work and up-front regarding the limitations. There is a high likelihood that these findings will be politicized, so it is incumbent upon researchers to be as careful as possible in describing their research.

Second, we need ongoing implementation research. This research would be strongest if it utilized common instrumentation, and the work of the special topic (Stein et al., 2017) and C-SAIL may provide direction here. Common survey and observational instruments would allow for the kinds of comparisons, both across sites and over time, in which we are most interested. Common instruments would also likely be able to amass greater validity evidence than homegrown or one-off instruments created by individual researchers. A foundation or an ambitious researcher might also consider pulling together a database of such resources, along with corresponding validity and reliability evidence, so that researchers can have a common place to go to identify tools to conduct their work. These tools could then be used as part of the kind of microlevel research illustrated in the special topic (Herman et al., 2016; Reynolds & Goodwin, 2016).

Third, we need to better understand how district and school leaders can support effective standards implementation. Early research suggests that the district office can play an important role in districts that are implementing the standards well (Durand, Lawson, Wilcox, & Schiller, 2016), in particular by supporting coherence and building professional learning opportunities for teachers. Curriculum materials may be an especially important support for teachers (Kane, Owens, Marinell, Thal, & Staiger, 2016), and there is clearly a greater need for work on this support (Chingos & Whitehurst, 2012). The work of the special topic shows the important role that networks can play, at both the teacher level and the state level (Hodge et al., 2016; Supovitz et al., 2016). More of this kind of research, which shines light on implementation variation across sites and explains such variation through policies and practices, would be of great use to practitioners. Such work could also facilitate interventions to be used in more controlled settings to improve implementation and outcomes.

Fourth, we have relatively little work that addresses the equity implications of the Common Core. The standards are widely perceived as being more rigorous than previous standards (e.g., Rentner & Kober, 2014), and one strand of opposition to them is that they are too ambitious and will cause more students to fail (e.g., Ravitch, 2016). Though continued upward trends in graduation rates suggests that these worries may be overwrought (Kamenetz & Turner, 2016), in

terms of implementation and impacts, we need to know whether these predictions are coming true. This means systematically investigating the variation in implementation and effects across classrooms and schools serving different student groups, based on poverty, English learner status, disability status, and race/ethnicity. In the No Child Left Behind era, research found systematic differences in the implementation of state standards over time (e.g., Au, 2007; Polikoff & Struthers, 2013); whether these trends will continue is an important equity question.

Finally, in this rapidly changing policy environment, we need research that is (a) timely and (b) made available and digestible to policy makers and practitioners. Faster-turnaround and open-access journals such as *AERA Open* can certainly play a role here, but it is all but certain that most policy makers and practitioners will struggle to take findings straight from peer-reviewed research to the classroom or legislative chamber. This means that scholars must strive to publish their work in blogs, briefs, and other settings that are accessible to a broader audience, if the work is to have impact. Tenure and promotion guidelines at most universities do not incentivize this kind of engagement, but it is my hope that, through training and modeling, we can develop a new generation of educational researchers who work to bring their important research to bear on policy and practice problems of the day.

References

- Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston’s charter high schools. *Journal of Labor Economics*, 34(2), 275–318.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Burris, C. (2015). What the national drop in 2015 NAEP test scores really means. *Washington Post Answer Sheet*. Retrieved from <https://www.washingtonpost.com/news/answer-sheet/wp/2015/10/28/what-the-national-drop-in-2015-naep-test-scores-really-means/>
- Center on Education Policy. (2016a). *Learning from and listening to teachers: A summary of focus groups on the Common Core and assessments*. Washington, DC: Author.
- Center on Education Policy. (2016b). *Listen to us: Teacher views and voices*. Washington, DC: Author.
- Center on Standards, Alignment, Instruction, and Learning. (2016). *First year college- and career-readiness standards were expected to be fully implemented across each state*. Retrieved from <http://c-sail.org/maps/first-year-implemented>
- Chingos, M. M. (2015). *Breaking the curve: Promises and pitfalls of using NAEP data to assess the state role in student achievement*. Washington, DC: Urban Institute.
- Chingos, M. M., & Whitehurst, G. J. (2012). *Choosing blindly: Instructional materials, teacher effectiveness, and the Common Core*. Washington, DC: Brookings.
- Dobbie, W., & Fryer, R. (2016). *Charter schools and labor market outcomes* (Working Paper No. 22502). Cambridge, MA: National Bureau of Economic Research.
- Durand, F. T., Lawson, H. A., Wilcox, K. C., & Schiller, K. S. (2016). The role of district office leaders in the adoption and implementation of the Common Core State Standards in elementary schools. *Educational Administration Quarterly*, 52(1), 45–74.
- Herman, J., Epstein, S., & Leon, S. (2016). Supporting Common Core instruction with literacy design collaborative: A tale of two studies. *AERA Open*, 2(3), 1–15.
- Hodge, E. M., Salloum, S. J., & Benko, S. L. (2016). (Un)Commonly connected: A social network analysis of state standards resources for English/language arts. *AERA Open*, 2(4), 1–19.
- Hughes, G. B., Daro, P., Holtzman, D., & Middleton, K. (2013). *A study of the alignment between the NAEP mathematics framework and the Common Core State Standards for Mathematics*. Washington, DC: American Institutes for Research.
- Kamenetz, A., & Turner, C. (2016, October 17). The high school graduation rate reaches a record high—Again. *nprEd*. Retrieved from <http://www.npr.org/sections/ed/2016/10/17/498246451/the-high-school-graduation-reaches-a-record-high-again>
- 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*. Cambridge, MA: Center for Education Policy Research.
- LaVenía, M., Cohen-Vogel, L., & Lang, L. B. (2015). The Common Core State Standards Initiative: An event history analysis of state adoption. *American Journal of Education*, 121(2), 145–182.
- Loveless, T. (2014). *The 2014 Brown Center report on American education*. Washington, DC: Brookings Institution.
- Loveless, T. (2016). *The 2016 Brown Center report on American education*. Washington, DC: Brookings Institution.
- Nix, N. (2016, December 18). Student test scores are on the rise. Does that mean Common Core is working? *The 74*. Retrieved from <https://www.the74million.org/article/student-test-scores-are-on-the-rise-does-that-mean-common-core-is-working>
- Opfer, V. D., Kaufman, J. H., & Thompson, L. E. (2016). *Implementation of K–12 state standards for mathematics and English language arts and literacy: Findings from the American Teacher Panel*. Santa Monica, CA: RAND.
- Polikoff, M. S. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice*, 29(4), 3–14.
- Polikoff, M. S. (2012). Instructional alignment under No Child Left Behind. *American Journal of Education*, 118(3), 341–368.
- Polikoff, M. S., Hardaway, T., Marsh, J., & Plank, D. (2016). Who is opposed to Common Core and why? *Educational Researcher*, 45(3), 263–266.
- Polikoff, M. S., & Struthers, K. S. (2013). Changes in the cognitive complexity of English instruction: The moderating effects of school and classroom characteristics. *Teachers College Record*, 115(8), 1–26.
- Porter, A. C., & Polikoff, M. S. (2012). Measuring academic readiness for college. *Educational Policy*, 26(3), 394–417.

- Ravitch, D. (2016, July 24). Renouncing the Common Core. *New York Times*, p. SR8.
- Rentner, D. S., & Kober, N. (2014). *Common Core State Standards in 2014: Districts' perceptions, progress, & challenges*. Washington, DC: Center on Education Policy.
- Reynolds, D., & Goodwin, A. (2016). Supporting students reading complex texts: Evidence for motivational scaffolding. *AERA Open*, 2(4), 1–16.
- Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the Common Core State Standards for mathematics. *Educational Researcher*, 41, 294–308.
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22(2), 127–160.
- St. Clair, T., Hallberg, K., & Cook, T. D. (2016). The validity and precision of the continuous interrupted time series design: Three within-study comparisons. *Journal of Educational and Behavioral Statistics*, 41(3), 269–299.
- Stein, M. K., Correnti, R., Moore, D., Russell, J. L., & Kelly, D. (2017). Using theory and measurement to sharpen conceptualizations of mathematics teaching in the Common Core era. *AERA Open*. Advance online publication.
- Supovitz, J., Fink, R., & Newman, B. (2016). From the inside in: Common Core knowledge and communication within schools. *AERA Open*, 2(3), 1–18.

Author

MORGAN S. POLIKOFF is an associate professor of education at the Rossier School of Education, University of Southern California, Waite Phillips Hall 904D, Los Angeles, CA 90089; polikoff@usc.edu. He researches the design, implementation, and effects of standards, assessment, and accountability policies.