Significance of the Focused Program of Research for C-SAIL

For roughly 3 decades, standards-based reforms have been a core element of state and federal efforts to improve education through policy (National Commission on Excellence in Education, 1983; Smith & O'Day, 1991). Content standards are the heart of standards-based reform and are intended to spell out what students should know and be able to do. The standards are meant to be supported with coherent high-quality assessments, curriculum materials, and professional development (PD). And to give weight to the standards and encourage their adoption, policy makers have often attached rewards or sanctions to students' test results (Clune, 1993; Smith & O'Day, 1991).

Considering the notable longevity of standards-based reforms, the success at achieving desired policy goals has been modest. In terms of teachers' instruction, standards-based reforms have led teachers to improve the alignment of their instruction with standards (e.g., Hamilton & Berends, 2006; Polikoff, 2012a), and, in some cases, attempt more intellectually ambitious instruction when called for by the standards and reinforced in the assessments (Koretz, Stecher, Klein, & McCaffrey, 1994). However, they have also led to unintended gaming behaviors such as narrowing the curriculum to the exclusion of untested subjects (Stecher & Barron, 2001), targeting instruction at students just below the proficiency cut score (Booher-Jennings, 2005), and excessive test preparation (Wong, Anagnostopoulos, Rutledge, & Edwards, 2003). The most methodologically sophisticated research suggests that standards-based reform and accountability have led to modest improvements in achievement, especially in math (Dee & Jacob, 2012; Hanushek & Raymond, 2005). National Assessment of Educational Progress (NAEP) scores at most grades, high school graduation rates, and college enrollment rates are at or near all-time highs for all student groups. However, there is little to no evidence of the achievement gap closing.

Why has the success of standards-based reforms been so modest? Research points to a number of flaws in both design and implementation that have undermined the ideals of the policy. These include poor-quality content standards with unclear language (Finn, Petrilli, & Julian, 2006; Hill, 2001), poor-quality and poorly aligned assessments (Polikoff, Porter, & Smithson, 2011), flawed school accountability metrics that identified too many successful schools as failing (Linn, 2004), and inadequate supporting materials, including textbooks and PD (e.g., Schmidt et al., 2001). In short, the vision of standards-based reform as laid out by its earliest advocates has not been realized.

These design and implementation challenges are all the more pressing with the adoption of new college- and career-ready (CCR) standards. Unlike many states' previous standards, the CCR standards call for mastery of ambitious content that will prepare students to succeed in college and career, raising the bar on expectations for student success (Achieve, 2013).

By far the most prominent set of CCR standards is the Common Core State Standards (CCSS). The CCSS were developed under the auspices of governors and chief state school officers working with advocacy and research groups. Most academic reviews of the CCSS view them as relatively high quality (Beach, 2011; Cobb & Jackson, 2011). The CCSS were adopted by approximately 46 states (depending on the subject). Recently, several states have removed themselves from the CCSS, mainly citing concerns about states' rights, federal intrusion, and the quality of the standards. However, more than 40 states remain in the Common Core at this time.

To support implementation, there has been a proliferation of developments in new assessments, curriculum materials, and other supporting policies. Two federally funded assessment consortia recently piloted their tests; full implementation of those assessments will commence in spring 2015. They are expected to bring multiple advances to student assessment, including computerized testing (and adaptive testing for one consortium) and the use of more advanced and diverse item types. Curriculum developers, both traditional and new, have also rushed to create "Common Core-aligned" curriculum materials, though early reviews of these materials have found them lacking (Chandler, 2014).

Our Center's Focus

Our *Center on Standards, Alignment, Instruction, and Learning (C-SAIL*) has an integrated research program designed to address some of the most pressing questions about standards-based reform. Our studies are grounded in a common framework—the policy attributes theory—focusing on the key role of instructional guidance for teachers. We are conducting four sets of studies: an implementation study, a longitudinal analysis, a measurement study, and an intervention study. Our two studies of natural variation—the implementation and longitudinal study—seek to understand state, district, and school efforts to implement CCR standards since the publication of the CCSS and other CCR standards. Our *implementation study* is being conducted in five partner states—Kentucky, Massachusetts, Missouri, Ohio, and Texas—that present sharp contrasts in their approaches to implementation to see how the new CCR standards have or have not brought on new and different approaches to implementation. A complementary study of natural variation is a *longitudinal analysis* that looks at all 50 states, plus urban districts in the NAEP Trial Urban District Assessment (TUDA), to assess how timing of adoption of CCR standards and approaches to their implementation have or have not resulted in different effects on NAEP. Our *measurement study* is revising and modifying instruments used to measure the alignment of standards, assessments, instruction, and curriculum materials, which we will use in our longitudinal and intervention studies.

Our studies of natural variation are important and will provide timely new information about standards-based reform. But studies of natural variation are limited in two important ways. First, they are limited to describing and investigating what state, district, and school policy makers and practitioners have tried. Second, using them to determine causal relationships is problematic because of the co-variation between (a) the natural variation in policies and practices and (b) nuisance variables. As for the first weakness, we believe that standards-based reform has typically stopped at the classroom door, and because of that, has not realized its full potential. To address this shortcoming, we plan to implement an *intervention*—Feedback on Alignment and Support for Teachers (FAST)—which will provide real time feedback to teachers on the alignment of their instruction to state standards, so that teachers can take corrective action where appropriate. In the words of the famed social psychologist Kurt Lewin, "To understand a system, you have to change it" (as quoted in Schein, 1988, p. 27). To address the second weakness of standards-based reform, we will test our intervention with a randomized control trial (RCT) in 4^{th-} and 5^{th}-grade classrooms in 10 districts across five states. Together, our *implementation, longitudinal, measurement,* and *intervention* studies afford a comprehensive examination of the new CCR standards—their current implementation and effects, their evolution over the next 5 years, and where they might go if implementation of standards-based reform were to penetrate the classroom, to directly support teachers as they enact standards-aligned curricula.

A Focus on College and Career Readiness

The Obama administration called for upgrading the rigor of existing standards and testing requirements to ensure that students are on track to be college and career ready. While the theory underlying CCR standards parallels that of past standards-based reform efforts, the new set of standards raises the bar on what students are being asked to know and do.

CCR math standards. In math, the CCR standards focus on the knowledge, skills, and dispositions necessary for students to succeed in college or the workplace (U.S. Department of Education, 2010; National Research Council, 2012). States that have adopted the CCSS math content standards have fulfilled the federal expectation for math required by the Race to the Top and ESEA flexibility policies. The CCSS math standards include specific learning standards that describe what students are expected to understand and be able to do at each grade level in K-8. Similar standards are grouped into clusters within broader domains, which span grade levels. At the high school level, domains are grouped by topic (e.g., Number and Quantity, Algebra, Functions, Modeling, Geometry, Statistics & Probability) rather than by grade level.

A key shift from earlier content standards is that the CCSS in math focus on fewer topics per grade level, but with more depth (National Governor's Association, 2010; Hiebert et al., 2003; Stigler & Hiebert, 1997). Another shift is that topics are linked with greater coherence across grade levels. Conceptual coherence provides opportunities for students to deepen and extend their understanding from one year to the next. A third shift relates to rigor. CCSS standards define rigorous math content as consisting of three equally important parts: conceptual understanding, procedural fluency, and application. Students are not only expected to carry out procedures with speed and accuracy, they must be able to explain the concepts underlying the procedures and apply both their conceptual understanding and procedural skills in applied problem solving situations (National Governor's Association, 2010). These attributes are best captured in the CCSS *Standards for Mathematical Practice*, which include: making sense of problems and solving them with persistence, reasoning abstractly and quantitatively, constructing viable arguments and critiquing the reasoning of others, and modeling with mathematics, among others. Creating learning environments that foster these characteristics is likely to require substantial instructional shifts for teachers.

CCR ELA standards. In ELA, CCR standards aim to reflect 21st-century skills that are expected of a literate individual who must critically read and hear, comprehend, and synthesize a broad array of information, as well as write and speak cogently for diverse audiences. As in math, most states have adopted CCSS as the way to fulfill the federal incentives for CCR ELA standards.

The CCSS standards for ELA provide CCR anchor standards (one set for K-5 and one set for 6-12) for each of four strands: Reading, Writing, Speaking and Listening, and Language. The anchor standards define broad, cross-disciplinary literacy expectations. The CCSS individual grade-level standards (and 2-year bands for high school) explicate specific expectations for each anchor standard[1]. Though the standards are divided into separate strands, it is intended that literacy is taught and learned in an integrated manner, with students writing about what they read or speaking about what they research and synthesize.

[1] The CCSS provides another set of literacy standards for grades 6 to 12 teachers of history/social studies, science, and technical subjects. These are designed to supplement content standards in these disciplines, with the intention that content area teachers help students learn the particular competencies needed to read, write, speak, and listen in each content area.

Educators have suggested that to fully implement the CCSS, teachers will need to shift their instruction to include: greater balance between informational and literacy text; content knowledge built directly from text; rigorous conversations about text; writing that is based on evidence and synthesis of sources; and use of increasingly complex academic vocabulary. Illustrating the overall emphasis on increased rigor, the CCSS emphasize the increased use of complex text at all grade levels. This requirement addresses current deficits that exist in the reading level of texts used in many schools as well as in the reading proficiency among the adult population (CCSSO, 2010). There is also an emphasis on the informed and strategic use of technology.

Theory Underlying CCR Standards

The theory underlying CCR standards is grounded in the original vision of standards-based reform that posits that improved teaching and learning will result from: (a) creating high-quality content standards that provide uniform and meaningful learning goals; (b) designing student assessments aligned to those standards; (c) providing a system of supports to help build teachers' capacity to successfully teach to the high-quality standards; and (d) establishing accountability mechanisms to motivate compliance with the standards (Smith & O'Day, 1991). At the most basic and broadest level, the evidence suggests that previous attempts at standards-based reform have been derailed because of failure to address the teacher and classroom implementation issues that are central to translating standards policy into productive practice.

Policy Attributes Theory

Our Center's study of standards-based reform is undergirded by the *policy attributes theory*, a simple yet powerful theoretical framework that posits the attributes that are related to successful policy implementation. Policy attributes theory (Porter 1994; Porter, Floden, Freeman, Schmidt, & Schwille, 1988) relates five components to successful policy implementation: specificity, consistency, authority, power, and stability. We apply this theoretical framework to our Center's work to guide our implementation, longitudinal, measurement, and intervention studies. See our conceptual framework in Appendix B Figure B1.

Specificity refers to how extensive and detailed a policy is. *Consistency* captures the extent to which various policies are aligned. For example, a curriculum may be tied to the school's vision of reform through a guide that links particular parts of the curriculum to specific school goals. Policies gain *authority* through becoming law, through their consistency with social norms, through support from experts, or through promotion by charismatic leaders. *Power* is tied to the rewards and sanctions associated with policies, such as monetary incentives. *Stability* represents the extent to which people, circumstances, and policies remain constant over time.

Porter and colleagues found that policies vary in their specificity, consistency, authority, power, and stability, and that the higher a policy is on one or all of the attributes, the greater the chance of its successful implementation (Porter et al., 1988). The set of five policy attributes may vary at the school, district, and state level. The policy attributes framework has been used to analyze systemic reform efforts (Clune, 1998) and comprehensive school reforms (Berends, Bodilly, & Kirby, 2002; Desimone, 2002; Desimone, Smith, Hayes, & Frisvold, 2005; Polikoff, 2012b).

What Have We Learned About Standards-based Reform?

Decades of research on the various manifestations of standards-based reform have identified successes as well as major areas where the operationalization of standards-based reform has fallen short. As a field, we have learned that the weak and mixed implementation of standards-

based reform, so common to most types of reform (see McLaughlin, 1976, 1987, 1990), occurred due to weaknesses in the policies' specificity, consistency, authority, power, and stability.

Specificity: lack of specific guidance. One major barrier to successful implementation of standards-based reform has been the lack of provision of real-time feedback to teachers that would enable them to improve their instruction (e.g., Taylor, Stecher, O'Day, Naftel, & Le Floch, 2010); thus it is no surprise that low levels of teacher change have been linked to lack of teacher capacity to change (Loeb Knapp & Elfers, 2008; Knapp, Elfers, & Plecki, 2004; Minnici & Hill, 2007; Stecher et al., 2008).

The need for more precise feedback mechanisms to facilitate instruction that is aligned to the standards is evidenced in the variation in teachers' responses to standards-based reforms. Some studies show standards-based reform may lead to more emphasis on didactic instruction (Booher-Jennings, 2005; Diamond, 2007; Diamond & Spillane, 2004; Sandholtz, Ogawa, & Scribner, 2004), while others find teachers make more use of conceptual, problem-solving approaches (Firestone, Camilli, Yurecko, Monfils, & Mayrowetz, 2000; Hamilton, Stecher, Marsh, McCombs, & Robyn, 2007; Stecher, Barron, Chun, & Ross, 2000). Still others find no change in instruction (Wong et al., 2003), or alternatively, that the relative emphasis on didactic or conceptual instruction depended on the teacher's skill and experience (Achinstein, Ogawa, & Spiegelman, 2004). Understanding how teachers are changing in terms of both what and how they teach is a fundamental component of the new standards-based reform (McCann, Jones, & Aronoff, 2010).

Consistency: tensions between instructional materials and standards and assessments. Much of the criticism of standards-based reform is that it resulted in the narrowing of the curriculum to respond to tested content (teaching to the test) and the use of class time to practice test-taking strategies (Diamond & Spillane, 2004; Hilliard, 2000). "Teaching to the test" can mean a number of things, and it can be good or bad depending on the circumstances (Firestone & Schorr, 2004; Koretz, 2008). The alignment of standards with assessments and other instructional materials plays a critical role: when such alignment is in place, teachers can teach to the standards and not focus on the test (Polikoff, 2012a; Porter, 2000). In contrast, when alignment is absent, teachers may adapt instruction to the assessments rather than the standards or instructional materials, and this may undesirably narrow the curriculum and give teachers conflicting messages about what to teach (Stecher et al., 2000; Stecher & Chun, 2001).

Authority: teachers' commitment to, interpretations of, and beliefs in the standards. Teachers' understandings and interpretations of what they are being asked to do are necessary precursors to changing practice (Louis, Febey, & Schroeder, 2005). The considerable variation in teachers' interpretation of standards and how they respond in terms of changes in the classroom have been well documented (Cohen & Hill, 2000; Spillane, Reiser, & Reimer, 2002). For some teachers, standards have served as a platform to inspire them to better serve the needs of low-achieving children (Desimone, 2013); while in other circumstances, teachers have admitted that they believe standards are too difficult for certain students (Stecher et al., 2008). Further, teachers vary in their perspectives on how much they think they need to change in order to successfully implement the standards. While some teachers reported changing their instruction to match the standards (e.g., Desimone, 2013), others have reported that their teaching was already consistent with state standards (e.g., Jennings, Swidler, & Koliba, 2005). This is problematic in most cases, as it likely reflects either a misunderstanding of the reform or a noncritical view of

their own instruction (Archbald & Porter, 1994; Cohen, 1990; Floden, Porter, Schmidt, Freeman, & Schwille, 1981; Porter et al., 1988; Porter & Brophy, 1988; Spillane et al., 2002).

Power: incentivizing attention to struggling learners. One of the most pervasive debates in standards-based reform is whether the system's rewards and sanctions (i.e., power) properly and productively incentivize a focus on lower-achieving students. The rewards and sanctions of previous waves of standards-reform were mixed in terms of whether they fulfilled their potential to act as a mechanism to improve learning opportunities for traditionally underachieving students (Hassel & Hassel, 2010), or whether they instead undermined instruction for these students (Darling-Hammond, 2004). Some studies in particular contexts show how accountability policies can exacerbate inequalities by marginalizing low-performing students (Booher-Jennings, 2005; Diamond & Spillane, 2004; Sandholtz et al., 2004). Documented perversions of the previous system include a focus on students who are near proficiency cut scores (bubble kids), at the cost of attention to lower-performing students (Hamilton et al., 2007; Le Floch, Martinez, O'Day, Stecher, Taylor, & Cook, 2007; Stecher et al., 2008; Taylor et al., 2010). In contrast, other studies have documented an increased focus on low-achieving students in response to accountability mandates (e.g., Stecher et al., 2008; Taylor et al., 2010). Furthermore, there is a lack of evidence that any subgroup has been disproportionately harmed by standards-based accountability (Gaddis & Lauen, 2014; Lauen & Gaddis, 2012).

Stability: mobility and longevity. Research on standards-based reform and school reform in general demonstrates that high mobility of students, teachers, principals, and district leaders can be detrimental to sustaining and institutionalizing reform (Berends et al., 2002; Smith & O'Day, 1991). Additionally, shifts in the curriculum, textbooks, and PD focus can also be a source of frustration to teachers (Desimone, 2002). Furthermore, educators' perceptions of how long a reform will last have a direct impact on their willingness to invest time and attention to building knowledge and skills related to the reform (Ross et al., 1997).

The Promise of Standards-based Reform

Despite evidence of previous implementation failures, there is considerable evidence that standards-based reform can, in certain circumstances, improve instruction, academic rigor, and student learning (e.g., Center on Education Policy, 2007; Dee & Jacob, 2012; Hamilton et al., 2007; Le Floch et al., 2007; Stecher et al., 2008; Taylor et al., 2010). Many scholars believe that if standards-based reform is done right, it can lead teachers to change their practice in positive ways (Bishop & Mane, 1999; Borko & Elliott, 1999; Wolf & McIver, 1999) and promote student learning (Hannaway, 2003; Porter, 2000). Thus, our Center's work is designed to learn from both past failures as well as successes.

Need for a Research and Development Center on Standards in Schools

The CCSS and other CCR standards have created a new day for standards-based reform. But what will make standards-based reform fundamentally different in its reach and effects this time around? Certainly, many believe the new standards, with their focus on college and career readiness and with their greater emphasis on cognitive demand—the type of thinking and level of complexity of thought students are expected to engage in, such as memorize or conjecture—represent a better target for instruction than previous state content standards. Moreover, many states have adopted the same or essentially the same standards rather than each state developing its own unique standards.

Another possible difference is better assessments of student achievement. Under NCLB, each state built, used, and reported on assessments of student achievement in ELA and math in grades 3-8 and one grade in high school. These assessments were to be aligned to the states' content standards. Alignment is a property that exists in degrees, and some have concluded that the degree of alignment has been modest to poor in most states (Polikoff, Porter & Smithson, 2011). Many states have joined one of the two multistate assessment consortia (PARCC and Smarter Balanced), whose new assessments will be administered for the first time in the spring of 2015. These assessments may be much better aligned and of much higher quality than previous assessments. At this point, that remains to be seen.

Certainly, more attention is being given to CCR standards than was given to the NCLB state content standards. Perhaps this attention is creating greater commitment to implementation of the standards. This in turn may translate into greater support of classroom-level implementation from states, districts, and other sources, and lead to bigger and more-uniform effects on students.

With this next generation of standards-based reform comes the responsibility to measure and evaluate implementation and effects. If implementation and effects are to be evaluated with validity, the studies must take place at or near the time of the introduction of the new standards and assessments. Retrospective studies simply will not suffice. Certainly, for standards-based reform, now is a time of intervention and experimentation; important changes are being introduced.

There is potentially important variation among states in approaches. Some states are adopting the CCSS in their entirety; some are adopting them in part; and some are not adopting them at all, instead developing their own CCR standards in ELA and mathematics. The policy and scientific communities have a responsibility to determine what is being done by whom, why, and to what effect. These estimates of natural variation and their effects are addressed in our implementation and longitudinal studies.

But an IES R&D center on CCR standards can provide much more than studies of natural variation. We hypothesize that even the new CCR standards with their state and district supports and encouragement from the Department of Education may not reach their potential unless effective, scalable efforts are designed, tested, found effective, and made widely available—efforts that take standards-based reform through the schoolhouse door and into the classroom in powerful ways that directly support teacher implementation. This is what our Center is doing.

Research Plan for Focused Program of Research

C-SAIL is conducting four studies, including the development and testing of a new intervention designed to support implementation of CCR standards in the classroom. This research agenda will produce usable information that will inform practitioners and policy makers engaged in the implementation of CCR standards. Our studies pay special attention to examining experiences and effects for subgroups, for example, students with disabilities (SWDs), English Language Learners (ELLs) (e.g., Solorzano, 2008), and low-achieving students. The next sections outline each study's research questions and data collections and analyses. Table B1 in Appendix B provides a timeline for our four sets of studies.

Implementation Study

Previous implementation research chronicled successes and failures of standards implementation, but provided no organizing framework to consider what is necessary for effective implementation (Chatterji, 2002; Hamilton, Stecher, & Yuan, 2008). Our approach addresses this gap by using the policy attributes framework to provide an analytic lens that will allow for comparisons across states, districts, and schools.

Research questions: (1) How are educators at state, district, and school levels understanding and interpreting the CCR standards?; (2) What is the nature and quality of supports and guidance at the state, district, and school levels? (a) To what extent are supports and guidance specific, consistent, authoritative, powerful, and stable? (b) How do these levels of attributes compare with "ideal" levels of the attributes?; (3) How are schools and teachers changing their practices? (a) how do these practices differ for student subgroups (e.g., SWDs and ELLs, and low-achieving students?), for ELA and math, in elementary and high school? and (4) To what extent are changes in understandings, support, and practice related to improved student learning, and how does that differ for student subgroups, and for ELA and math, in elementary and high school?

The implementation study provides a detailed picture of the supports for CCR standards implementation and progress toward implementation in five purposively selected partner states. The study examines how educators across these states are understanding and interpreting the standards; assess the nature and quality of support and guidance at the district, state, and school levels; and observe whether and how teachers report they are changing their practices. In each state, we are collecting data on the specificity, consistency, authority, power, and stability of implementation at the state, district and school levels. Across all five states, the study is collecting information through (a) annual interviews with state and district administrators about supports for CCR standards implementation; (b) a review of state documents (e.g., standards, assessments, websites, curricular materials, etc.) designed to provide guidance around CCR standards implementation; and (c) a statewide representative survey of district administrators, teachers, and principals in years 1 and 4 in our five partner states. Cross-state analyses will reveal commonalties and differences across the five states, each representing different policy contexts.
In addition, within each state we are examining the relationship between variation in supports for implementation at the district and school levels and student outcomes.

State partners. A key component of the Center is its partnership with five states that are diverse both geographically and in terms of their approach to CCR standards: Kentucky, Massachusetts, Missouri, Ohio, and Texas. The state partners are committed to participating and supporting the data collection activities that support the implementation study, helping to recruit districts and schools for participation in the RCT, and participating in conversations about what the Center's findings mean for policy and practice. In return, state partners are participating in leadership opportunities through the Center, receive information and recommendations on CCR standards implementation on an annual basis, and receive tools and an intervention designed to support CCR standards implementation.

State selection. The five geographically diverse states collectively represent a range of policies and characteristics. These include both CCSS and non-CCSS states; states with PARCC, Smarter-Balanced, and other assessments; geographic variation; and some states with a high concentration of ELLs. See Table B2 in Appendix B for an overview of our partner states. Three states (Kentucky, Massachusetts, and Ohio) were early adopters of CCSS and Massachusetts and

Kentucky have developed particularly robust supports for implementation. In contrast, Texas has not adopted CCSS, instead choosing to develop state-specific CCR standards. Efforts to implement CCSS in Missouri have been met with concerted political opposition, leading to compromises by CCSS supporters.

We realize that CCSS is a volatile political issue and, as a result, states' participation may shift. Our Center will make adjustments as needed to meet our design requirements.

State and district administrator interviews. In each of the first 4 years of the implementation study, we are conducting interviews with state officials with primary responsibility for overseeing implementation of CCR standards (e.g., state curriculum directors or state directors of assessment) and with officials in three purposively selected districts in each state (chosen to vary on key policy attributes). The interviewees will include math or reading content specialists, and, in the case of smaller districts, the superintendent or assistant superintendent. These interviews will collect in-depth information on CCR adoption and implementation, asking questions about each of the policy attributes as they relate to CCR standards policies.

We are developing interview protocols that include a balance of closed-ended questions with consistent response options as well as open-ended questions that enable the interviewer to probe for unique circumstances across locales. We will develop state and district classifications that facilitate cross-case analyses and that permit survey analyses of schools nested in districts and states with different characteristics.

Document review. We will supplement the data collected in the state and district administrator interviews with a review of key CCR standards documents in each state and in the sampled districts. Like the interviews, this document review will be updated annually for the first 4 years of the Center. As a part of the document review, we will examine state and district websites focused on CCR standards implementation as well as other documents provided by state or district administrators to inform teacher practice (e.g., curricular guides, aligned curricular materials and lessons, guidance around instructional practices that support alignment). In addition, we will use the Surveys of Enacted Curriculum (SEC) framework to assess the alignment of standards and assessments in each of the five states following established content analysis procedures (Porter et al., 2008). Based on these documents and the data collected

through the interviews, the research team will assess the extent to which implementation reflects the five policy attributes, and then develop insights into how they interact to shape implementation.

District administrator, principal, and teacher surveys. We plan to administer district, principal, and teacher surveys in years 1 and 4. To maximize what is learned given the Center resources available for the survey, we plan to target elementary and high schools, which represent sharp contrasts in terms of the proximity to high school graduation, the target of CCR standards, and also in terms of departmentalization and teacher subject matter specialization; (middle school is not well defined as to grade level configuration and many may not be departmentalized).

Districts, schools, and teachers will be drawn from the five partner states. The sampling in each state will consist of three stages. The first stage of selection will be the school district or pseudo-district.[2] The next stage of selection will be the school (elementary or high school). The third and final stage of selection will be the teacher. We will select about 70 districts from each state with probability proportional to the number of students in the relevant grades in the district. From each sampled school district, we will randomly select two elementary schools and two high schools.

In the last stage of sample selection, for each elementary school in the sample, we will randomly select with equal probabilities two 4^{th}-grade teachers and two 5^{th}-grade teachers, a teacher of SWDs, and a teacher of ELLs. Similarly, for each sampled high school, we will randomly select with equal probabilities two ELA teachers, two math teachers, a teacher of SWDs, and an ELL teacher. The principal in each sampled school and the district administrator in each sample district will be asked to complete the principal survey and district administrator survey, respectively. Our approach to selecting the appropriate district administrator to complete the survey will parallel the approach used to identify district administrators for the interview portion of the study (described above). As a result, we will have approximately 70 district administrators, 280 principals, and 1,680 teachers per state in our sample. We designed the survey to have sufficient power to detect a difference of 10 percentage points between states or between waves within states, for a dichotomous item, at a significance level of 0.05 and 80% power.

Survey development. We will design questions to ask about the specificity, consistency, authority, power, and stability of CCR standards implementation. The district administrator survey will be designed to collect information about the district's policy, support, and guidance for implementation of CCR standards. The principal survey will be designed to collect information about principals' instructional and organizational leadership and their guidance and support for standards implementation. The teacher survey will be designed to collect information

[2] We will sample school districts and then sample both elementary and high schools from the same district. Some school districts have elementary schools but no high schools or possibly high schools but no elementary schools. Before sample selection, we will form pseudo-districts by grouping (only if necessary) a school district with elementary schools but no high school with a nearby (or overlapping) district with at least one high school. In doing so, we will endeavor to create pseudo-districts that are consistent with prevailing feeder patterns from elementary to high school.

about teachers' CCR standards implementation in their classrooms, as it relates to instructional practices, standards-aligned curricular materials, and student assessments. The teacher survey will also include questions about PD and the availability and sufficiency of supports that teachers receive from various sources, including such materials as detailed curriculum guides, frameworks, and/or pacing sequences.

To the extent appropriate, we will align teacher survey items with principal and district survey items as well as state and district administrator interviews. Our work will be guided in part by existing instruments such as Achieve's Common Core Feedback Tool for Educators, RAND's American Teacher Panel and American School Leader Panel surveys, and the teacher surveys for the National Longitudinal Study of No Child Left Behind (Birman et al., 2009; Taylor et al., 2010).

The process of survey development will include expert review of survey items for checking their face validity and adequate content coverage; cognitive interviews with a small number of district administrators, principals, and teachers to identify common survey problems (e.g., potential misinterpretations, or confusing wording) (see Desimone & Le Floch, 2002); revisions of survey items based on the results of cognitive interviews; and a small-scale pilot test of district administrator, principal, and teacher surveys. Based on the results of the pilot test, the surveys will be finalized and converted to web surveys. Using web surveys has several advantages, including the use of skip patterns and data validation procedures. However, we will provide a paper copy survey to any respondent who requests one.

Survey administration. We will administer the finalized district administrator, principal, and teacher web surveys in the spring of years 1 and 4 of the Center. With the help of sample district personnel, we plan to obtain rosters of schools, principals, and teachers and collect target respondents' e-mail addresses. We understand the importance of obtaining high response rates. To that end, we will work with each state to identify appropriate incentives to encourage survey completion. Before launching the surveys, we will ask district staff to send an e-mail to respondents introducing the survey and asking them to participate. The online survey interface will allow for follow up with non-respondents to ensure that response rates are as high as possible. Based on our experience with similar surveys, we expect that we will obtain a response rate of at least 80% (e.g., Wayne et al., 2014; Stecher, Garet, Holtzman & Hamilton, 2012).

Survey analysis and reporting. Administration of district, principal and teacher surveys at two time points (years 1 and 4) will yield both cross-sectional and longitudinal datasets. Teacher survey data will be linked to both principal and district survey data, and we will using the linked datasets to examine CCR standards implementation in districts, schools, and classrooms. We will start with basic descriptive data analyses, focusing on the overall patterns of CCR standards implementation in each state, and examining the extent to which implementation varies between districts within each state, between schools, and within schools.

We will conduct regression analysis to examine the relationship between district- and school-level policies and supports, and teachers' CCR standards implementation in the classrooms, taking advantage of the nested data structure (i.e., teachers nested within schools and schools nested within districts).

Using the longitudinal datasets, we will track overall trends in CCR standards implementation over time within states. At the same time, we will seek to account for such trends with a number of contextual or policy-relevant variables at the district, school, and teacher levels. Do schools

and teachers improve in implementing CCR standards? If so, which policy attributes account for the improvement? Which teachers, of which students, show improvement? The survey will be vertically equated to allow for the rigorous measurement of change over time.

Finally, we will examine the relationship between student outcomes and the district, teacher, and principal surveys. By linking data from the survey to data on student achievement at the school level, we will explore the relationship between the specificity, consistency, authority, power, and stability of CCR standards implementation at the district and school levels and student outcomes. To the extent possible, we will control for other factors that influence student performance, including past academic performance and school and district characteristics.

Implementation data from the intervention study. In addition to drawing on the interview, document, and survey data mentioned above to describe implementation in our five partner states, we will also draw on data from our RCT on the FAST intervention, (described in detail later). The Center's RCT will measure the effectiveness of the intervention, but it will also provide rich data to measure each school's policy attributes related to standards implementation. We will be able to document teacher change as teachers engage in an intervention designed to support their implementation of CCR standards, permitting us to address key questions about implementation. For example, what is the pace of teacher change when engaged with a specific and consistent intervention? How do variations in authority, power and stability interact with the intervention's effects?

Innovations in our implementation study. Our Center's work with states allows cross-state comparisons grounded in an analytic framework (policy attributes). For example, how are Texas' CCR standards similar to and different from other state standards, and how does their policy system differ from other states? How are these differences related to the trends in implementation and student achievement? Are certain types of standards more likely to be well implemented and more quickly seen in student achievement results? What are the implications of this for other state systems? Further, we can examine levels of policy attributes across states. How close do states get to "ideal" levels? Do early-adopting states like Kentucky look stronger in terms of the policy attributes and what are the implications? What actions and conditions are associated with stronger policy systems?

Expected deliverables. Annual reports will be issued for each of the five states, describing the status of implementation in the state. The research team will work closely with our state partners to share the results of this work broadly with policy makers, practitioners, and other stakeholder in each state. Annual reports will also draw lessons from across the five states and be designed for a broader national audience looking to learn from the work being done in the five states. Findings from this report will be shared widely through webinars, policy briefs, conference presentations, and peer-reviewed journals.

Longitudinal Outcomes Study

Drawing on longitudinal data from the NAEP, high school graduation rates, and college enrollment and persistence records, this study will exploit the natural variation in the timing of standards and assessment implementation, as well as variation in the cognitive demands of individual state standards prior to implementation, to examine as rigorously as possible the effect of implementing CCR standards and assessments. We will further explore whether the effect of CCR standards and assessment adoption is moderated by the specificity, consistency, authority, power, and stability of state implementation; and the extent to which state standards are aligned

with assessments in math and ELA in elementary and high school. We will examine effects overall on student performance as well as whether effects vary for subgroups of students, including SWDs and ELLs.

Research questions: (1) Does implementing CCR standards result in increases in student college and career readiness?; (2) Does the adoption of assessments aligned with CCR standards by state accountability systems result in increases in student college and career readiness?; (3) How does the effect of implementing CCR standards and aligned assessments vary by student subgroup (including ELLs and SWDs), subject (ELA vs. math), and level (elementary, middle, and high school)?; and (4) Is the effect of implementing CCR standards and assessments on student learning moderated by the specificity, consistency, authority, power, and stability of state implementation? and (5) Is the effect of implementing CCR standards and assessments on student learning moderated by the extent to which state standards are aligned with assessments in math, ELAR in elementary and high school?

Forty-six states initially signed on to the CCSS, and others have adopted their own CCR standards. Such broad adoption means that examining the effect of implementing CCR standards is one of the critical questions facing education researchers over the next decade. Educators, policy makers, parents, and other observers will want to know whether CCR standards have made a positive difference in whether students have the skills they need to be successful after high school—in other words, *was this investment of resources worth it?* But the broad adoption poses challenges to answering this question. With almost all states implementing the reform, it is difficult to assess what would have happened had the new standards not been implemented.

Our primary means for answering these questions is a comparative interrupted time series (CITS) design that exploits variation in the timing and intensity of CCR standards implementation across states. In its simplest form, an interrupted time series (ITS) design measures the same outcome for a treatment group multiple times before and after the introduction of an intervention. The effect of the intervention is then estimated by examining the difference in outcomes before and after implementation. Comparison group cases often are added to this simple version of the design to guard against potential threats to internal validity, converting an ITS design to a comparative ITS or CITS (Dee, Jacob, & Schwartz, 2013; Wong, Cook & Steiner, n.d.).

Sample. The study sample will include all 50 states and the districts that participate in the Trial Urban District Assessment (TUDA) study for the National Assessment of Study Progress. The study will exploit two sources of variation in implementation to assess the effect of implementing CCR standards: timing and strength of previous standards.

States implemented CCR standards at various times between 2010 and 2014. Further, some states that initially adopted the standards have withdrawn (e.g., Indiana and South Carolina). Likewise, the timing of the implementation of standards-aligned assessments varies across states, with some states implementing a new assessment as early as 2011 and others not planning to implement until the 2015-16 school year.

In addition, there is variation among states in how demanding their proficiency standards were prior to their adoptions of CCR standards. Following Wong, Cook, and Steiner (n.d.), we will compare student proficiency on the NAEP with student proficiency on state assessments to examine how challenging state standards were prior to CCR implementation. For each state, we will calculate the difference between (a) the percentage of students who were proficient on state

assessments in ELA and math in 2009 (prior to CCR standard implementation in all states) and (b) the portion of students proficient in that state on the NAEP.

We will calculate the difference for ELA and math for grades 4, 8 and 12 for each state. We hypothesize that implementing CCR standards will have a greater impact the less demanding the state's previous proficiency standards were. To rule out alternative explanations, the study will compare outcomes in states that were early implementers with those outcomes in states that were later implementers, with the latter serving as a nontreatment comparison group in the early years of the study.

Data sources. The longitudinal study will draw on extant data on key student outcomes collected over time. To supplement this extant data collection, we will build a 50 state database that describes (a) each state based on the five policy attributes and (b) the alignment of state standards and assessments. We will draw on existing documents and summaries of state policies. When necessary, we will supplement the documents with interviews with state administrators. The database will be created in Year 1 of the Center and updated in years 2 through 4. Our research team brings extensive experience developing and populating this kind of state policy database (see, for example, Desimone, Smith, Hayes, & Frisvold, 2005; Polikoff, 2012b).

To assess the extent to which state standards and assessments are aligned, we will examine state standards and assessments in all 50 states grades 4 and 8 in ELA and mathematics. The task is manageable because many states will share common standards and assessments. Content analysis will be conducted by trained analysts using the SEC taxonomies. Content analysis can be used to generate an index of alignment between a set of standards and an assessment. Previous research describes the content analysis methods in detail and indicates that the methods provide reliable data (Porter, Polikoff, Zeidner, & Smithson, 2008). As a result of these analyses, the state data base will include measures of the extent to which state policy is consistent with the five policy attributes and the extent to which state assessments and standards are aligned. These variables will be used in the CITS to examine whether the moderate the effect of the implementation of the effect of CCR standards and assessments.

In addition, the following outcome data will be collected:

- **NAEP assessments.** State NAEP has been administered biannually since 1990 in math and reading. We will draw on these data to track changes in student achievement in math and ELA for grades 4, 8, and 12.
- **High school graduation.** To assess the effect of implementation of CCR standards and assessments on high school graduation, we will draw on data on the average freshman graduation rate (AFGR) from the National Center for Education Statistics. While the adjusted cohort graduation rates are generally considered more accurate, this measure was introduced in the 2010-11 school year, the same year the first states adopted CCR standards. In contrast, AFGR data are available beginning in the 2007-08 school year.
- **College enrollment and persistence.** The study will draw on data from the National Student Clearinghouse and Graduation Rate Survey to estimate college enrollment persistence rates over time.

Data analysis. The study will employ a year and state fixed effects approach to estimate the effect of implementing CCR standards and assessments on these outcomes. The basic modeling approach is illustrated in Figure B2, Appendix B.

We will estimate effects for each of the outcome measures described above using two definitions of treatment. First, a state's treatment status each year will be defined based on whether a given state is implementing CCR standards or assessments in that year. Here the implicit comparison group is states that are not implementing in that year. Second, treatment status each year will be defined based on whether a given state had low standards prior to implementing CCR standards or assessments and is implementing in that year. Here the "comparison group" is states that had high standards prior to the implementation of CCR standards. We will estimate these models for all students in the states and for particular subgroups, including SWDs and ELLs. In addition, we will use the measures from the state policy database and the alignment study to include interactions in the model to explore how the effects of supports and alignment moderate the effects of the policy.

***Addressing common threats to validity*. C**ommon threats to validity in interrupted time series include history (e.g., other forces besides standards introduction may account for any measured change), instrumentation (e.g., changes in measurement), and selection (e.g., population change that coincided with the introduction of new standards). Our design is relatively robust to each, though we also will employ sensitivity tests to provide a sense for how these threats may influence our results. Table B3, Appendix B provides details about validity threats to our approach and the extent to which the design is robust to each.

Innovations in our longitudinal study. Our longitudinal study will use state-of-the-art analysis with robustness checks to provide the most rigorous results on associations between policy attributes, implementation, and student learning trajectories. In addition our innovations in this work include building a unique and powerful 50 state policy database that will allow us to do cross-state comparisons linking state policy to achievement trends. We will be able to answer question such as: How quickly are states changing, and on which attributes? How are these changes related to implementation and student learning trajectories? Further, our approach allows the Center to chart explicitly how the new standards are influencing instruction for different groups for different students (e.g., ELLs, SWDs, and low-achieving students). We will also quantitatively compare the alignment of standards to assessments for each state, to answer pressing policy questions, such as, how well aligned are the new multi-state assessment consortia assessments with CCR standards? Our procedures will allow us to identify specific areas of alignment and misalignment, which will provide actionable information for states and districts.

Expected deliverables. In Year 1, we will build and populate the state policy database and assess the alignment of states' standards and assessments. The results of these analyses will be published on the Center's website in a format that is searchable by state. We will make the database accessible to state and district leaders to allow states to conduct their own inquiries to see what other states are doing and the possible impact those actions are having. Our longitudinal analyses will be disseminated in multiple forms, including issues briefs, research papers, and webinars. In years 3 and 5, we will update the state policy database and alignment analysis as we add another year of outcome data to our analyses. Updated reports will be published every two years and shared widely.

The Measurement Study

One of the key requirements to support and monitor the implementation of CCR standards, we argue, is the availability of valid, practical measures that can be used to assess the extent to which teachers are engaging in classroom instruction consistent with the standards. Such measures are needed to provide feedback to teachers to guide improvement, and to provide feedback to districts and states, to guide the development of supports. One central element of the Center is the refinement of measures that will meet these goals.

Research questions: (1) What is the validity of teacher reports of their content and cognitive demand coverage for a single lesson?; (2) What is the reliability of content analyses of assessments based on the revised SEC? (3) What is the validity of teacher reports of content coverage over an extended period of time?; (4) What is the reliability of coding of classroom observations?; and (5) Does the validity or reliability in the above studies differ based on student type or subject?

The measures that are the focus of our work will be designed to assess the extent to which the *content* of instruction and assessments aligns with CCR standards. However, content alignment is not enough in the context of new CCR standards. In particular, many CCR standards documents also specify important skills and habits of mind that do not neatly fall into traditional conceptions of "content." For example, the Common Core Mathematics Standards for Mathematical Practice call for students to "Construct viable arguments and critique the reasoning of others," to "Look for and make use of structure," and six other skills. In general, we view these additional instructional changes called for by the standards as falling under the category of *cognitive demands.* Thus, we must measure not only topic coverage, but also coverage of various levels of cognitive demand representative of the key expectations in the standards.

The measurement studies will focus on 4^{th} grade math and 5^{th} grade ELA for feasibility and because the products of that work will be used in our intervention described later. Resources permitting, the instrument development will be extended to other grades.

We plan to use three modes of data collection to capitalize on the strengths of each: teacher surveys, classroom observations, and content analyses of the assessments (quizzes or exams) teachers routinely administer at the end of each unit of instruction. Classroom surveys are low cost and relatively low effort and can provide high-quality data about certain aspects of instruction over an entire year (Mayer, 1999), but are limited in their ability to measure instructional quality. Observational methods can provide data on complex instructional approaches for one or a few lessons, but they cannot provide data across an extended period of time, given the costs and burden involved. Content analyses of assessments or curriculum materials are less expensive than observations and can be reliable (Polikoff, 2014); however, assessments represent only a portion of students' opportunity to learn, and content analyzing curriculum materials for a whole year is expensive (Polikoff's recent content analysis work using the SEC to analyze textbooks cost $10,000 per book).

Using multiple measures of instruction will allow us to characterize instruction along a number of dimensions and triangulate our findings. We propose to build and validate these instruments during Year 1, so that we can use them in our planned pilot of the intervention in the Year 2 and then in the controlled trial of our FAST intervention in years 3 and 4 of the Center.

Building the teacher survey and content language for analyzing assessments. The basis for our survey and content analysis of the assessments that teachers administer to their students will be

the SEC in math and ELA (Porter, 2002). The SEC surveys include subject-specific content languages that were developed over time with the input of content experts and educational practitioners (e.g., Porter et al., 1988). The content languages can be used to measure instruction, assessments, standards, and curriculum materials in terms of both topic and cognitive demand. See Appendix C Table C1 for more details on the SEC. The content languages include 133 fine-grained topics in ELA and 187 in math. These fine-grained topics (for example, Word Origins) fall underneath coarse-grained topics (in this case, Vocabulary), of which there are 14 in ELA and 16 in math. The cognitive-demand levels, defined in greater detail in Table C2 Appendix C, are based on modified Bloom's taxonomies. The five levels in math are (1) memorize, (2) perform procedures, (3) demonstrate understanding, (4) conjecture, generalize, prove and (5) solve non-routine problems, make connections. In ELA, they are (1) memorize/recall, (2) perform procedures/explain, (3) generate/create/demonstrate, (4) analyze/investigate, and (5) evaluate.

A recent review concluded that the SEC is the only existing tool that can be used to analyze alignment among instruction, curriculum materials, standards, and assessments (Martone & Sireci, 2009). (See Appendix C Table C3 for details of how alignment is calculated). Furthermore, the SEC has been widely used by researchers, states, districts, schools, and teachers nationwide. For example, the tools have been used to analyze the content of state and Common Core standards and assessments (Polikoff et al., 2011; Porter, 2002; Porter, McMaken, Hwang, & Yang, 2011), and the content of instruction (Polikoff, 2012a); as part of a randomized experiment testing the impact of teachers' receipt of alignment information on instruction (Porter, Smithson, Blank, & Zeidner, 2007); and to investigate the relationship of content coverage with student achievement gains (Gamoran, Porter, Smithson, & White, 1997; Polikoff & Porter, 2014). The SEC tools have also been used to examine the opportunities of ELLs (Smithson & Blank, 2007) and SWDs (Blank, 2013; Smithson, 2013) to learn standards-based content.

Studies on the quality of SEC data indicate that teachers understand the SEC languages and can use the languages to describe their instruction in survey form (Porter et al., 1993). Also, teacher reports of content coverage correlate well with ratings by external observers (Porter, Kirst, Osthoff, Smithson, & Schneider, 1993). Furthermore, content standards and assessments can be reliably content analyzed (Porter et al., 2008). The aforementioned experiment found that providing feedback to teachers on the content of their instruction resulted in statistically significant increases in instructional alignment in math (Porter et al., 2007). Most importantly, teacher-reported content coverage on the SEC predicted student achievement gains in high school math classes (Gamoran et al., 1997), though a more recent study found weaker relationships (Polikoff & Porter, 2014). In short, evidence shows that the SEC content languages capture meaningful distinctions in the content of instruction, standards, assessments, and curriculum materials.

Though the SEC surveys represent a strong foundation for our work, we believe they would benefit from refinements that will strengthen their utility. Potential weaknesses of the SEC identified in prior research include too broad of a grain size of the topic distinctions and lack of clarity of the definitions of the cognitive demand levels, particularly in ELA (Beach, 2011; Cobb & Jackson, 2011).

The instruments will be reviewed by a panel of math and ELA experts that we convene, and, depending on the feedback from the panel, substantially revised in the Year 1. The goals of these

revisions are (a) to ensure they comprehensively describe math and ELA content and the key content shifts called for in CCR standards in our partner states; (b) to ensure the categories and definitions of cognitive demand adequately capture instructional approaches recommended in CCR standards documents; and (c) to build a user-friendly version of the SEC in the form of a weekly log sheet that can be completed in 15-20 minutes (for our FAST intervention, described later, and also for potential use as a monitoring/continuous improvement mechanism for schools).

We will assemble two teams of content experts who will review (a) the existing surveys and frameworks and (b) the CCR standards in our partner states. The teams will have four experts each (i.e., eight experts total); the experts will be content-area experts (e.g., mathematicians), experienced teacher educators (e.g., professors of reading and writing and math education), and experts in survey design and measurement. They will meet in person for a 3-day meeting at the start of Year 1, led by Andy Porter and Morgan Polikoff, who are experts on the SEC and its properties. Their task will be broken into two portions.

The first portion of the task will be for each group to independently review the SEC and the CCR standards documents and propose revisions to the SEC. For instance, we will ask participants to ensure the SEC languages adequately cover *all* topics in the chosen sets of CCR standards, and at an appropriate grain size. We envision the participants will suggest adding some topics to the instrument and possibly revising the cognitive-demand levels or their response scales. For example, panelists may review the standard for the mathematical practice "make sense of problems and persevere in solving them" (Common Core Standard for Mathematical Practice 1) and determine that elements of this standard are not fully captured by either the topic or cognitive-demand classifications in the SEC. Similarly, the panelists may conclude that the Common Core in ELA's focus on text complexity is not adequately captured by the SEC. In these areas, panelists will propose new data-collection strategies.

The second portion of the task is for the combined group to take the modified SEC instrument and develop a weekly log that can be used to provide regular feedback to teachers on their instruction – a key component of the FAST intervention, described below. The goal for the weekly log is something that can be completed in 15-20 minutes. The log creation activity will take place on the Day 3 of the meeting.

At the end of the 3 day meeting, the goal will be to have a revised SEC (full version) and a revised SEC (weekly log), both of which measure the key content and cognitive-demand approaches called for in the CCR standards. The revised content languages will also be used for content analyses of assessments in the implementation and intervention studies following existing content analysis procedures (Porter et al., 2008).

We will examine the new tools in a series of cognitive interviews with practicing teachers (Desimone & Le Floch, 2004). The goal is to ensure consistent understanding between researcher and respondent (Fowler, 1995). We will interview five teachers, revise the instrument, and conduct five additional interviews. We will have completed the modification of the instruments by January of Year 1.

Validating survey and content analysis instruments. We will assess the quality of the survey and content analysis measures in the second half of Year 1. There will be three parts of this study. Part 1 will answer the question "*What is the validity of teacher reports of their content and cognitive demand coverage for a single lesson?*" We will identify 40 teachers in each subject

across grades and school districts (to ensure different content is being taught), who will use the weekly log tool to report on their content and cognitive demand coverage for a single lesson. Project personnel and advanced graduate students will also observe these lessons using video and categorize the content and cognitive demand of teachers' instruction using the same scales. Based on the teacher reports and the ratings by external observers, we will calculate traditional indices of agreement, such as Cohen's Kappa and correlation coefficients. Forty teachers per subject will give us 80% power to detect a correlation of .4 in each subject.

In Part 2, we will use the same 40 teachers in each subject for a study to examine the technical properties of content analyses of teachers' assessments, answering the question "*What is the reliability of content analyses of assessments based on the revised SEC?*" We will ask teachers to provide classroom assessments for a month's worth of instruction. Four content-area experts in each subject will content analyze these materials using the revised SEC framework. The content analyses will then be subjected to a generalizability theory D-study analysis to determine the reliability of the procedures. Previous analyses in math and ELA have generally found generalizability coefficients of .75 for 4 raters or higher (Porter, 2002; Porter et al., 2008); we will aim to meet or exceed these figures.

In Part 3 of the validation study, we will seek to answer the question "*What is the validity of teacher reports of content coverage over an extended period of time?*" This portion is essential to ensure that the surveys that will measure instruction across extended periods of time in the intervention study accurately represent teachers' enacted curricula. For this portion, we will ask a set of 40 teachers from across grades and districts to complete the SEC weekly logs for the spring semester. We will also ask them to complete a semester-end survey reporting on their content across the semester. We will aggregate the weekly logs to the semester level and compare the aggregated logs to the semester-end survey using correlations and descriptive analyses.

Together, the three parts of the validation study will provide extensive quality evidence about the SEC surveys and content analysis procedures.

Building and piloting observational instruments. Another goal of the first year's work is to develop an observational instrument that can be applied in classrooms to complement the data captured by teacher surveys and assessment content analyses. In particular, though the existing validity evidence cited above supports the conclusion that teachers understand and can report on their cognitive demand emphasis on the SEC surveys, we recognize that these data could be strengthened if supported by high-quality observational data.

To that end, we will build and pilot highly focused observation protocols targeting the key skills called for in the CCR standards in our study. In math, these are mainly the cognitive- demand skills indicated in the Standards for Mathematical Practice (for Common Core). In ELA, these are mainly included in the Anchor Standards and text complexity guidelines for each grade. Because we believe the standards are clearly different between the two subjects, we plan to build a protocol in math and in ELA. We will draw on existing instruments in this development work, including the Student Achievement Partners rubric, the New Teacher Project Core Teaching Rubric, the Mathematical Quality of Instruction (MQI) protocol, and the Protocol for Language Arts Teacher Observation (PLATO).

The protocols will be developed in Year 1.

We will pilot the instrument in the summer between years 1 and 2, using the video-recorded

observations mentioned above for the survey validation (40 classroom observations in each subject). The videos will be coded by four team members using the protocol, and generalizability theory analyses will be used to estimate reliability. After the reliability analyses, the coders will meet to discuss discrepant codes and use their discussions to propose any revisions to the instrument. Revisions will be tested by recoding samples of the classroom observations using different coders and repeating the process of estimating reliability and discussing discrepant codes. This work will answer the question "*What is the reliability of coding of classroom observations?"* The goal is to obtain a complete observation protocol by the end of the summer of Year 1. Both the survey and observation instruments will be used in the intervention pilot in Year 2, and they will be revised as needed based on findings from the pilot.

Innovations in Our Measurement Study. The measurement study will capitalize on the SEC expertise of Porter and Polikoff, as well as the extensive research history of the instrument, to ensure the SEC is well prepared to serve as a measurement tool in the context of CCR standards moving forward. Furthermore, the study-developed observational instrument focused on CCR standards will have potential value in both research and practice for measuring teacher implementation of CCR-aligned instruction.

Expected deliverables. The measurement study will produce a modified SEC that can be used as a yearly survey or weekly log, and a validated observation protocol that can be used to measure the quality of teachers' instruction in terms of its alignment with CCR standards. We will make these instruments available on our website and will work directly with our partner states to provide them the tools and assist them if they choose to use them as a means of supporting and monitoring instruction.

The Intervention: Feedback on Alignment and Support for Teachers (FAST)

As the final component of the Center's work, we plan to test our FAST intervention, designed to provide valid, practical, real-time feedback to teachers. Our central assertion is that for CCR standards to be implemented in a way that makes a difference for student outcomes, each of the policy attributes articulated in our conceptual framework needs to be in place and operating at a high level. In particular, instructional guidance in the form of feedback, examples, materials, and lesson plans needs to be specific and consistent. In addition, this guidance needs to be backed by authority, appropriately incentivized (power), and seen as a stable target.

Research questions: (1) Is the FAST intervention implemented with fidelity?; (2) Does the FAST intervention lead to greater alignment of teachers' content coverage with state standards?; and (3) Does the FAST intervention lead to increased student achievement as measured by state and Center-developed assessments?

Instructional alignment as the critical mechanism for the success of CCR standards. The central focus of the FAST feedback will be the *content of a teacher's instruction*. We purposefully avoid the use of the term "pedagogy" within the context of our intervention. While particular pedagogical approaches may be differentially effective in students' achievement of the standards, the CCSS and other state standards are not intended to prescribe how content is taught. This perspective is reflected by CCSS developers, who note: "Teachers know best about

what works in the classroom. That is why these standards establish what students need to learn but do not dictate how teachers should teach" (Common Core State Standards Initiative, 2014).

Our focus will be supporting teachers in aligning their instruction to the content standards—for example, supporting a teacher in providing students with opportunities for problem solving, and critical thinking in their math instruction; and providing students with opportunities to engage in evidence-based conversations about complex text, both informational and literary, in their ELA instruction.

Description of the components of our intervention. FAST includes three components: (1) a *real-time, online, personalized feedback* mechanism based on the revised SEC framework to provide each teacher with specific, accessible, and instant information on how aligned their weekly instruction is to CCR standards and the provision of accessible materials and examples related to addressing specific areas of improvements calibrated to the personalized feedback, (2) an *offsite coach* to assist the teacher in understanding and applying the materials and examples; and (3) school-level *collaborative academic study teams* (CASTs) that provide teachers the opportunity to engage in substantive discussions with each other to share strategies for addressing their alignment feedback.

Real-time online personalized feedback. The Center will create a web-based feedback system that provides teachers with feedback on alignment in real time, after they have completed their weekly logs (see the Measurement section for details on logs). In addition to weekly feedback, teachers will be provided with summative feedback as the intervention progresses so they can see how their instruction is evolving. See Appendix C Figure C1 and Table C4 for examples of alignment feedback. As described below, this feedback will be *specific* in its detail, focusing on concrete aspects of instruction, and *consistent* in its role in facilitating the alignment of instruction to the new standards. This addresses critiques that previous standards-based reform attempts provided no mechanisms for teachers to receive feedback on their implementation, to guide them toward improvement (Porter, 2002; Wang & Oddell, 2002).

We know teachers interact with students differently, and we are especially interested in how instruction may differ in the standards environment for SWDs, ELLs, and other students. Thus, we have designed our intervention to allow for such differentiation. Weekly, teachers will be asked to enter their log data for each of three target students—one ELL, one student with disabilities, and a student judged by the teacher to be near the class median in achievement (hereafter referred to as median student). We will provide guidance for teachers in how to choose the students as we have done successfully in the past (e.g., Desimone, McMaken, & Hochberg, 2014).

In addition to having teachers complete the weekly log, we will ask them to scan in copies of the assessments that they administer to their students (i.e., exams, quizzes, and tests). We anticipate that teacher-administered tests will be an especially useful mechanism for indicating the teachers' understanding of how to align assessments to the standards. The teacher-administered assessments will be coded and entered into the teacher's online FAST account, so they can see how aligned the assessments are to the content standards. We plan to provide each school with a low-cost scanner; alternatively, if teachers have a smart phone, they will have the option of using the free smart phone application that allows the creation of pdfs from the camera phone. If the school's technology is a barrier to scanning and sending over e-mail, or if a teacher prefers, we will provide self-addressed stamped envelopes to mail in the documentation.

To complement data from the weekly SEC log and teacher-administered assessments, teachers will be asked once a month to provide a videotaped classroom lesson that aligns with content from a pre-specified list of standards. Our team has experience and expertise in collecting large (hundreds) of video observations, with 80 to 90% response rates (Garet et al., 2014; Wayne et al., 2014). We will code the videos using the rubric for alignment with CCR standards described in the measurement study. This will provide the teacher with feedback on the quality dimensions of standards-based instruction enactment that are usually not possible to elicit from survey and assessment data (e.g., the complexity of students' analysis of texts). Further, we will ask the teachers to fill out the log specifically for the lesson that was videotaped. This will be an opportunity to obtain feedback on how teachers think about their own instruction.

Nature of feedback. This array of logs, assessments, and video data will allow us to obtain a comprehensive picture of each teacher's instruction. Based on these data, we will provide feedback that takes several forms. First, we will provide clear, accessible indications of the teacher's alignment, highlighting where their content coverage is aligned and exactly where it is misaligned. Second, we will provide materials and examples that will be calibrated to the areas in which that particular teacher is misaligned. In the past, translating the standards into actionable lessons has been problematic—usually because the materials teachers have access to are poor quality and not directly aligned with the standards (Kesidou & Roseman, 2002; Krajcik, McNeill, & Reiser, 2008; Schmidt, Wang, & McKnight, 2005). In the first year of the Center, the intervention team will work through the objectives and standards for the intervention schools and link aligned resources to each objective. We will take advantage of the considerable online resources already available (e.g., achievethecore.org; EngageNY.org; educore.ascd.org), but will use our team's expertise and improved SEC instrument to ensure materials are well aligned with a particular standard, which is not always the case for publicly posted material (e.g., Kober & Rentner, 2012). For math, the resources will be a combination of problems, extended tasks, and lessons. For ELA, the resources will be example activities and inquiries along with strategy guides (e.g., how to choose a complex text that is suited for close reading and text-dependent questioning).

Guiding teachers in interpreting feedback. We know from our previous work and the broader literature on implementation that interventions are more successful when teachers are provided with explicit guidance on collecting data and interpreting feedback (LaFleur, Witt, Naquin, Harwell, & Gilbertson, 1998; Sterling-Turner, Watson, & Moore, 2002). We will provide a video for teachers to view before the start of the data collection as well as one designed to be viewed after the first set of feedback is received, when we think teachers will be more engaged and have a concrete example to work through (Gilbertson, Witt, Singletary, & VanDerHeyden, 2007). The introductory videos will (a) walk teachers through how to complete the SEC log and use other data-collection tools; (b) provide examples of the FAST feedback; (c) explain how to interpret the feedback; and (d) describe how to adjust instruction based on the feedback. These videos will also provide teachers with the opportunity to practice examining, understanding, and responding to the feedback. Further, our "offsite coaches" (described below) will discuss the introductory videos in their initial interactions with each teacher.

In addition to introductory videos, we will provide a series of instructional videos that teachers can access during the year. They will be calibrated to the standards they are teaching. For example, in math, videos might include a teacher working through a problem showing multiple solution paths or having students explain why certain approaches are incorrect.

Offsite FAST coach. To aid teachers in translating the feedback into *specific* actions, we will have an offsite FAST coach assigned to each of our intervention teachers (Herll & O'Drobinak, 2004; Matsumura et al., 2010; Koh & Newman, 2006; Neuman & Cunningham, 2009; The Regents of the University of California & Trustees of the University of Pennsylvania, 2011). Each coach will be math or ELA expert, and be trained by our team. Each teacher's coach will be available by e-mail/internet chat to respond to questions or concerns within 24 hours. In addition to this ongoing, real-time availability, coaches will provide detailed feedback during a prescheduled 30-minute skype session every 10 school days. As described in the RCT design section below, we expect to have 300 teachers in our intervention study schools, with 150 assigned to the intervention, 75 in mathematics and 75 in ELA. We anticipate having three math and three ELA coaches (each assigned 25 teachers) providing feedback for an average of 12 or 13 teachers a week (2 or 3 teachers a day). This will allow the coach to allocate sufficient time to each teacher, including viewing videos and alignment data, gathering materials, talking with the teacher, and providing a chronicle of the meeting, which will become part of the teacher's online FAST information archive.

The FAST coach will assist the teacher in using the materials provided by the online feedback system. Additionally, the coach will provide individualized feedback based on the teacher's log, observation, and assessment data in order to guide the teacher in improving the quality of and alignment of their instruction.

Further, in the baseline teacher survey administered as part of the RCT (described below), we will ask teachers about the advice and resources on aligning their instruction to the standards that they are currently getting from state, district, and other providers. We will use this information in our intervention by relaying it back to the FAST coach, who can then help the teacher navigate any mixed messages and also provide higher quality guidance, in terms of consistency and specificity.

Collaborative academic study teams (CAST). Our intervention includes opportunities for teachers to work together in improving their aligned instruction. Our approach is a hybrid adaptation of learning communities (e.g., Fulton & Britton, 2011) and consultant-teacher problem-solving teams (e.g., Fuchs & Fuchs, 1990, 2006). An intervention study inclusion criterion will be that the principal ensures teachers have the time and space to hold 45-minute meetings twice a month. Recent rigorous research provides evidence of the efficacy of such working collaborative mechanisms in changing teachers' instruction in ways that improve student achievement (Biancarosa, Bryk, & Dexter, 2010; Gersten, Dimino Jayanthi, Kim, & Santoro, 2010).

Our Center leadership team has extensive experience developing and facilitating successful teacher collaborations (e.g., Fuchs & Fuchs, 2006; Fuchs, Fuchs, & Bishop, 1992; Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Fuchs, Fuchs, Phillips, & Simmons, 1993). We will structure the CAST process to reflect four principles: (a) guiding conversations in ways that support teachers' responsiveness to instructionally relevant data; (b) providing expert input, tailored to team-specified problems; (c) designating a leader (e.g., Fulton & Britton, 2011); and (d) focusing on particular students or types of learners. If the school has an instructional coach in reading or math, we will integrate that school-based coach into the CASTs, to ensure instructional guidance from the multiple coaches is *consistent*, moving the teacher in the same direction.

Involving the principal. We have designed our intervention to integrate the principal as a partner, able to support and back the intervention through *authority and power*, in several ways. First, we will meet virtually with each of the principals in our intervention study before the start of the intervention. This meeting will be an opportunity to share our conceptual framework, introduce the intervention, and receive principal feedback on the framework and intervention, as well as other issues of concern to the principal.

Second, we will include the principal in viewing the initial introductory video, to build his or her understanding of the intervention (Kam, Greenberg, & Walls, 2003). Third, we will include the principal in one early teacher coaching session with the FAST coach, in order to continue to develop understanding and buy-in as well as provide a model for the principal for giving productive feedback, which principals often lack (Frase & Streshly, 1994; Stiggins & Bridgeford, 1985). Because the principals are responsible for evaluating teachers, we do not want principals participating regularly in the coaching or team meetings. Fourth, we will ask that the principal hold teachers responsible for teaching to CCR standards and for participating in all aspects of the intervention, which will hopefully create a positive incentive for participation. Finally, the FAST coach will meet with the principal virtually during the first 4-6 weeks of the study to address any issues related to the specificity, consistency, authority, power, and stability attributes that may be problematic, based on early teacher feedback.

FAST pilot process. During Year 2 (prior to the intervention study), we will pilot the intervention in two schools, working through the entire input, feedback, guidance, and CAST cycle for at least 6 months. The piloting will include weekly interviews with participating teachers to identify technological, logistical, and substantive issues as they arise. We will refine the FAST intervention based on this pilot process.

The Innovations in Our Intervention. Our intervention has the potential to be of great interest to policymakers and practitioners, in that it provides concrete steps districts and schools can take to aid teachers in recognizing how well aligned their instruction is to standards, and providing concrete guidance to improve that alignment. Further, the intervention is designed to improve alignment in real-time; rather than providing end-of-year results, this intervention has a built-in improvement cycle. This is an innovative mechanism for addressing the policy-practice gap given that our intervention adapts to meet the needs of a teacher. Also, this emphasizes the partnership aspect that is often missing from interventions: FAST works with teachers by design.

Further, while many studies conduct post-hoc analyses that identify variation in effects by student characteristics (e.g., Hamilton, Stecher, & Yuan, 2008), few anticipate such variation in their study design. Our intervention is explicitly designed to provide teachers with guidance calibrated to the different needs of their students, and likewise our studies of implementation of

both the intervention and more broadly in our study states, is set up to characterize how policy attributes, implementation and achievement may differ for the very sub-groups of students for whom adequate positive effects from earlier waves of reform were not realized (e.g., ELLs and SWDs).

Expected Deliverables. Products from the intervention study will include be a detailed, targeted replicable intervention. We will provide a step-by-step handbook on how to implement the intervention. We will post this on our website and make it available to our partner states and participating districts and schools.

The Intervention Study

To provide a strong test of the impact of FAST on teachers' instruction and student achievement, we plan to conduct an RCT during years 3 and 4 of the Center. To maximize our capacity to deliver an intervention of the required intensity, we will focus on the implementation of CCR standards in a single grade for each subject (as we explain above, 4th grade in math and 5th grade in ELA). We will deliver the FAST intervention in both years 3 and 4, gathering outcome data in both years, which will permit us to assess the benefits of a second year of support. The FAST intervention will be compared with "business as usual"—the support typically received by 4th and 5th grade teachers in the participating districts.

To implement the RCT, we plan to recruit 60 elementary schools (approximately 240-480 teachers), across 10 districts (2 each in the 5 participating states). Within each district, schools will be randomly assigned to treatment and control groups. We chose the school rather than the teacher as the unit of assignment primarily because FAST includes a component to encourage teacher collaboration within schools. In addition, the videoing of classroom instruction will be more efficient with multiple teachers in the same school. The choice of the school as the unit, however, reduces the power of the design, in comparison to a teacher-level design with the same number of teachers. To improve efficiency, schools within districts will be blocked on prior achievement or other characteristics that might be associated with outcomes or moderate the impact of the treatment.

Measures. Data will be collected from teachers and students in all 60 study schools. See Appendix C Table C4 for a table reflecting the data collection schedule. The study will focus on three outcome measures:

- **Alignment of teachers' content coverage with state standards.** To measure alignment, teachers in both the treatment and control condition will complete a revised version of the content section of the modified SEC, which has a web-based interface. The modified SEC will be administered in the 2017-18 and 2018-19 school years (years 3, and 4 of the Center), at the end of each of the two years of implementation. It will also be administered in the spring of 2016-17 (Year 2 of the Center), just prior to random assignment, to serve as a baseline measure. The baseline will allow us to test the equivalence of the treatment and control groups prior to the intervention, and it will be used as a covariate in outcome analyses. One potential concern is that post-intervention, treatment teachers will be more familiar with the SEC than control teachers, and as a result may have a different mindset in responding, which might affect the SEC's capacity to elicit real differences in instruction. To assess the extent to which this is an issue, we will ask a randomly chosen subset of treatment and control teachers to watch and code a selected instructional video, which should reveal any differences. If differences are found, they will inform interpretation of later results.

- **Student achievement on state assessment.** Each of our partner states is expected to have adopted an assessment designed to be aligned with CCR standards for 4th and 5th grade by the time the study is underway. We will request student-level math scores for all students enrolled in 4th grade in the 60 study schools in the 2017-18 and 2018-19, the two implementation years, along with the students' 3rd grade scores for 2015-16 and 2016-17 to use as covariates. Similarly, we will request ELA scores for students enrolled in 5th grade. To enable pooling data across the five states, we will standardize the scores based on the state mean and SD for each grade and year.

- **Student achievement on Center-created assessment.** Because the assessments used in the five states will differ, we plan to administer a short study-created test. This will permit us to examine the effects of the FAST intervention on a common test aligned to state standards. In spring of years 3 and 4 of our center, the study administered mathematics test will be administered to all students enrolled in 4th grade in the 60 study schools, and the study administered ELA test to all 5th grade students. The test construction approach will follow recent work by Porter, Polikoff, Barghaus and Yang (2013). The work begins by determining an appropriate "target" (i.e., domain) for assessment, in this case a combination of CCSS and Texas standards. Next, we will assemble an item bank, in this case drawing from items made public by the testing consortia and other CCR test vendors, designed to be as comprehensive as possible with respect to the domain. Finally, items will be selected for inclusion on the assessment, one at a time, using an algorithm designed specifically to build highly aligned assessments to the target. This technique has been shown to yield more tightly aligned assessments to a given target that are more sensitive to instruction on that target than typical assessments (Porter et al., 2013).

We will use universal test accommodations, such as extended time and reading items aloud (Vang et al., 2012) when administering the Center-constructed achievement test at the end of each year of the intervention (e.g., Smarter Balanced Assessment Consortium, 2014). For SWD students we will be further guided by the IEP and for ELLs we will follow guidelines from the Selection Taxonomy for ELLs Accommodation (STELLA).

In addition to the outcome measures described above, we also plan to collect measures of the supports available for teachers in implementing state standards in both treatment and control schools. One of the noted weaknesses of prior impact studies is the lack of information on what happens in control schools. While we refer to the controls are "business as usual," in K-12 schools, especially urban districts, it is rarely the case that there is no PD, curriculum or leadership intervention operating, and usually there are many occurring simultaneously. To permit a rigorous comparison of the supports available in treatment and control schools, we will include items in the SEC survey asking both treatment and control teachers about professional development and the specificity, consistency, power, authority and stability of the standards policies as they affect teachers and classroom.

***Analysis strategy*.** The basic analytic strategy is to compare outcomes for schools that were randomly assigned within each district to treatment and control. We will conduct separate analyses for the impact at the end of the first and second year of implementation, making it possible to compare the effects of one and two years of implementation of the intervention. Three-level models (with students nested within teachers' classrooms within schools) will be used to estimate the impact of the PD program on student achievement, and two-level models (with teachers nested within schools) will be used to estimate the impact on teacher alignment. We will conduct all analyses separately for mathematics and ELA.

We propose to focus on "intent-to-treat" estimates, reflecting the impact on the targeted (or "intended") sample, whether or not all eligible teachers in the treatment schools take full advantage of the supports provided. To improve the precision of the impact estimates, the analysis models will include school, teacher, and student-level covariates

Because random assignment will be conducted separately within each of the 10 school districts participating in the study, the study will effectively comprise 10 separate random assignment experiments. To obtain the impact estimates, we will pool the data for all 10 districts in a single analysis, treating the districts as fixed effects, and schools and teachers as random effects. Separate program impact estimates will be obtained for each district, and the estimates will be averaged across the 10 districts.

We will examine whether the effect of FAST differs for schools with lower and higher prior achievement.[3] It is possible, for example, that the implementation of CCR standards might be more difficult for 4th and 5th grade teachers in lower-achieving schools, if students lack prerequisite skills that should have been developed in prior grades. We will also examine which of the effects differ across SWDs, ELLs and median students.

If we find an impact of FAST on student achievement, we will conduct exploratory analyses of the potential role of alignment of instruction as a mediator. In addition, we plan to conduct exploratory analyses of the relationship between alignment and achievement. We will conduct these analyses using the full sample of schools, relying on the end of year SEC as the measure of alignment. In addition, in the treatment schools only, we will repeat the analysis, using stronger measures of alignment derived from the weekly logs, videos, and analysis of teacher-administered assessments. As a final exploratory analysis for treatment schools only, we will capitalize on the fact that treatment teachers will complete both the detailed measures and the end of year SEC, allowing us to assess the validity of the end-of-year measure.

For the main impact analyses based on all 60 schools, we estimate a minimum detectable effect size (MDES) of 0.41 for the teacher alignment measure.[4] This is comparable to effects found in other recent randomized trials that have examined teacher-level outcomes (e.g., Garet et al, 2011.) We estimate an MDES of 0.16 for student achievement, assuming an alpha=.05 and power of 0.80, under a two-tailed test.[5] Our analytic model for assessing student impact is in Appendix C, Figure C2. We will test the sensitivity of our analysis to multiple ways of handling missing data, given controversies about the proper procedures for dealing with missing data.

***Sample recruitment and retention*.** The districts will be selected in collaboration with representatives of the five states. We will seek districts that are interested in the intervention and meet the following criteria: (a) the district contains at least six elementary schools enrolling students in both grades 4 and 5 and at least two teachers in each grade; and (b) the district does not provide an intensive, systematic coaching to support implementation of state standards.

---

[3] This can be tested by comparing the impact of the treatment across random assignment blocks. The power for the test will be weak, since we will have at most 30 blocks (3 blocks per district, each containing one treatment and one control school).

[4] The MDES estimate for teacher alignment is based a covariate R-squared = 0.3 at the school and teacher levels, and school ICC=0.10.

[5] The MDES for student achievement is based on a covariate R-square of 0.7 at the school, teacher, and student levels; school ICC=0.10; and teacher ICC=0.10.

The districts will be identified and recruited through a multistage process that has been successfully employed in recent RCTs conducted by members of the Center's research team. (See for example, Garet et al, 2008; 2010; 2011). In the first stage, we will use information from each state to identify districts meeting the study criteria, as well as information from the Center district survey conducted in spring of year 1, which, although a sample survey, is likely to include many districts of sufficient size for the RCT. In the second stage, we will work with each state to identify about five candidate districts, and we will hold informational conference calls with officials in the identified districts. We will then visit districts that express interest, discussing both the intervention and the rationale and features of random assignment, to ensure that district staff understand how the process would work. In districts that indicate interest, we will meet with principals at eligible schools. As the final stage of securing the participation of each of the 10 districts, we will asked the superintendent and the principal of each participating school to sign a roles and responsibilities document, summarizing the information already communicated. We view this as a crucial stage in the process, facilitating a common understanding—and also important in sustaining participation over time, given superintendent and principal turnover.

Innovations in our Intervention Study. The intervention is directly targeted at the key flaws in current implementation of standards-based reform, and uses the most cutting-edge alignment tools available. The RCT we propose is implemented in the context of a larger study of what is going on in the state. Further, we can locate the schools in our RCT in the state implementation and longitudinal studies, to show how their districts and states fare on the policy attributes.

Expected deliverables. We will issue research briefs, conference papers and policy briefs chronicling the implementation and outcomes of the intervention. We will provide each district with a district-specific description of the intervention's implementation and outcomes, highlighting any district-specific contextual factors that warrant attention.

National Leadership and Outreach Activities

Previous research on standards implementation chronicled successes and failures, but not in real-time, and not in ways that policymakers and practitioners could respond to in order to improve their systems (e.g., Fuhrman, 1994; Massell, Kirst, & Hoppe, 1997). Our panels, webinars, and other interactive activities starting in Year 1 will be designed to foster conversations where states and districts can share ideas and strategies—a forum where failures as well as successes are highlighted in ways that allow participants to learn from each other, and devise concrete action steps.

The Center's stakeholders include educators at all levels of the system, education researchers, universities, policymakers and the public. We propose a multi-tier, multi-modal dissemination strategy to ensure we reach all of these audiences. Each tier will include appropriate collaborative arrangements with IES as well as quality control via an internal C-SAIL Publications and Dissemination Committee.

C-SAIL will create a portfolio of research-related publications, position papers, "working papers," monographs, and journal articles aimed directly at the research community. These will be distributed through conventional scholarly and research channels, as well as readily available on our website. C-SAIL will issue progress and end-of-year technical reports, and as a capstone publication, will produce a book synthesizing the findings from our focused program of research.

We will make available all of our products on our website. We will publicize our work using existing social media accounts (e.g., Twitter, Facebook) at our respective institutions. We will

also establish a C-SAIL presence on social media for rapid dissemination and interaction with other researchers, policymakers, and practitioners involved in CCR standards implementation."

The use of C-SAIL-produced CDs and DVDs will enable us to communicate successful practices and case vignettes in a powerful fashion, facilitating replication and social marketing. We will regularly publish a *C-SAILnotes* newsletter of activities, findings, and products and prominently display them on our website. With the assistance of our Center's editors and media staff, we will translate C-SAIL research into news releases and briefs in publications read by practitioners, policy officials and the public (e.g., newsletters, media outlets).

Our outreach activities will include active information sharing with national organizations, maintaining a presence in the media, and hosting seminars and panels timed at the beginning or end of important national conferences.

Management and Institutional Resources

C-SAIL's structure is highly interactive and collegial, but one that maintains clear lines of authority and responsibility necessary to ensure quality, accountability, direction, and leadership. Our Center's approach to studying standards in schools recognizes the importance of combining expertise in K-12 subject-area content and practice, with experts in measurement, research design, statistics and policy, in collaboration with policymakers and practitioners. It is this integration of practitioners and scholars, substance and methods, that promises to yield the most rigorous, grounded insights, in a form useful for policymakers and practitioners to act on.

Management team and goals. The Center is directed by Andy Porter, Professor at the Graduate School of Education at the University of Pennsylvania (Penn GSE). Porter will take lead responsibility for managing the Center's day-to- day activities. He will be the key point of contact for IES. Porter is joined by four co- directors to form the Center leadership team: Mike Garet and Mengli Song from AIR, Laura Desimone from Penn GSE, and Morgan Polikoff from the University of Southern California (USC).

The work of the Center will be tightly focused and guided by its research questions—the lines of work we propose are interactive and will be accomplished through a Center-wide collaborative effort, not by independent directors and managers pursuing separate lines of work. Five teams will pursue five lines of investigation: (a) the implementation study led by Laura Desimone; (b) the longitudinal study led by Mengli Song; (c) the measurement study led by Morgan Polikoff; (d) the development of the intervention led by Toni Smith; (e) and the RCT test of the intervention led by Mike Garet.

Advisory board. We have recruited to our advisory board two of the state superintendents in our five-state sample: **Mitchell Chester,** superintendent of public instruction in Massachusetts, and **Michael Williams,** superintendent of public instruction in Texas. At the district level, we have recruited **William Hite,** superintendent of the School District of Philadelphia. Helping us integrate practical wisdom with research knowledge and sophistication is **Milbrey McLaughlin,** Professor Emeritus, Stanford School of Education. A celebrated researcher of education policy, she has contributed more to understanding the importance of research on implementation than any other scholar. Bringing state-of-the-art knowledge of CCR standards and classroom implementation are **David Pearson** (for ELA), Professor Emeritus at UC Berkeley, and **Uri**

Treisman (for math), Professor of Mathematics and Education at the University of Texas and head of the Dana Center, which works with thousands of teachers to improve math instruction. **Bob Slavin,** Professor at Johns Hopkins University, will bring to our Center the wisdom of one of the most influential and well-published school reform scholars in the U.S. today. **Dan Goldhaber** from the University of Washington will bring cutting-edge knowledge and experience with using administrative data and longitudinal analysis to address problems of education policy and their effects. To advise us on our challenging issues of measurement is **Mark Reckase,** Professor of measurement and quantitative methods at Michigan State University. For expertise on our focus on ELLs, we have **Kenji Hakuta,** Professor in the School of Education at Stanford University, who is the nation's foremost authority on these issues.

We will convene our advisory board annually, and keep members involved and informed of our progress and issues on which we need interim advice through phone and e- mail correspondence. Collectively, our advisory board membership represents both practical and scholarly expertise and knowledge on the full array of issues embedded in our scope of work. They represent diversity in ethnicity and gender.

The University of Pennsylvania's excellence in research, undergraduate and graduate education, and community involvement has earned it a spot among the top 10 premier research universities in the United States. Research is an essential and esteemed undertaking at Penn, as reflected in its 165 research entities. As of fiscal year 2014, Penn's research community included more than 4,300 faculty and over 1,100 postdoctoral fellows, nearly 11,025 graduate students and 5,400 academic support staff and graduate assistants. Its research budget was $899 million, nearly 82% of which comes from the federal government. The University maintains an extensive system of 13 research libraries containing more than 4 million volumes, 34,000 journal subscriptions, and a comprehensive digital reference collection. With its 12 graduate and professional schools, Penn provides a broad-based system of support dedicated to promoting practical local, state, and national research.

The Graduate School of Education (GSE) is ranked 5th nationally and in the top 10 for educational policy research. Over the past 5 years, the school has secured $40 million in research funding from the federal government (including the Department of Education, the National Institutes of Health, and the National Science Foundation) and from a variety of private foundations, state governments, and corporate sponsors. The school and university have abundant resources to support the Center, including secure data-storage facilities, and full-time IT and administrative staff. The Center will secure for researchers needed hardware and the most recent versions of project-specific statistical and database software. Data will be stored on secure servers using state of the art encrypting. The School is equipped with high-quality teleconferencing capacity, digital cameras, and LCD projectors.

American Institutes for Research (AIR; www.air.org) is a not-for-profit corporation engaged in independent research, development, evaluation, and analysis in the behavioral and social sciences. AIR has earned a national and international reputation for efficiently and effectively conducting high-quality R&D. It has more than 1,500 research, technical, administrative, and clerical personnel. AIR is a leader in the use of rigorous research to evaluate policies and practices in education. As a firm, AIR has gained expertise in statistical analysis, frequently performing complex, multiyear studies as well as short-term data collection and analysis tasks. Researchers at AIR also have developed in-depth knowledge about standards-based reform and

college and career readiness. AIR regularly builds project teams that draw upon the expertise of staff from multiple program areas, academic disciplines, and research sites. One distinguishing characteristic of AIR is our multidisciplinary problem-solving capability. AIR's expertise spans pre-K through college and beyond, and includes finance, organizational development, research methodology, statistical and qualitative methods, and evaluation, among many other areas.

The University of Southern California (USC) is one of the nation's premier research institutions and has garnered international prestige and respect for its academic programming, research, community engagement, and the high caliber of its faculty and students. The university has raised over $508 million annually for research. For the past century, USC's Rossier School of Education has developed and prepared professional leaders in the field of education and research, including teachers and superintendents, administrative professionals, policy leaders, and scholars. The research centers at Rossier, including the Center on Educational Governance, are leading the field in studying effective and non-effective strategies for high-need student populations with special focus on urban education.

What We Will Have Accomplished At The End of Five Years

We designed the Center to provide usable information to the field that will inform practice. At the end of 5 years, we expect the following outcomes.

Our five-state implementation study will document in rigorous and replicable ways how the implementation of ELA and math CCR standards are alike and different across subjects and across states using different standards and different approaches to implementation. Our studies of

implementation will capture differences in approaches between states, districts, schools, and classrooms, and between students within a classroom.

Our longitudinal study will provide estimates of state adoption of CCR standards on student achievement as measured by NAEP in both ELA and math, high school graduation rates, and college-going and employment rates. The longitudinal study will use state-of-the-art methodologies for assessing state adoption effects.

New tools will be available for teachers to monitor in real time the content of their enacted curriculum and how that content is aligned or not to their states' CCR standards in ELA and math. Though the focus will be on grades 4 and 5, the tools will be generalizable to other grades. The web-based tools will create teacher-level data on content defined at the intersection of topics and cognitive demands and provide a mechanism for measuring how that enacted curriculum may vary or not across three types of students: SWDs, ELLs, and median students.

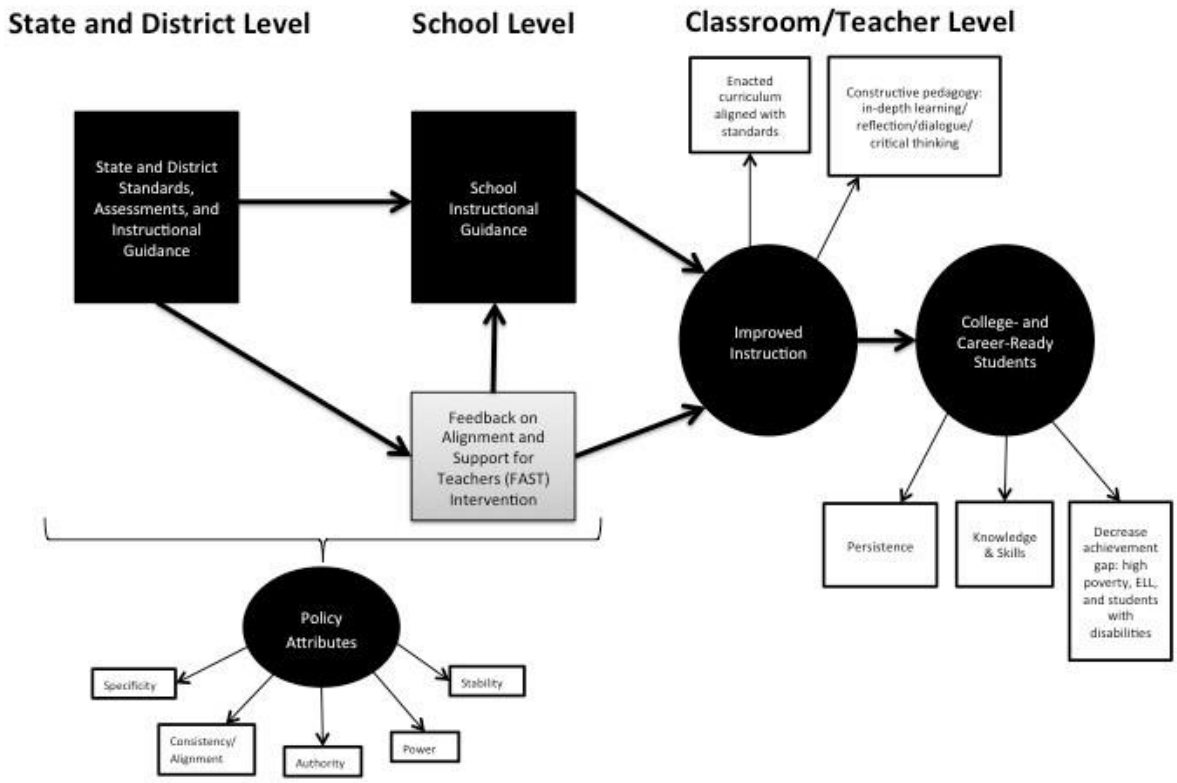
The web-based monitoring tools will be the lynchpin of an intervention that seeks to take standards-based reform to the classroom level. Teachers in our FAST RCT intervention condition will receive (a) at least weekly monitoring and feedback on the alignment of their enacted curriculum with their state standards; (b) video orientation and training to the monitoring system and how it can be used; (c) an off-site coach who reaches out on a regular basis to each teacher to provide support in understanding the feedback and taking appropriate actions to adjust instruction where necessary; (d) a school-based team where teachers interact with each other twice monthly in a school-level supportive community; and lastly, (e) a principal who is knowledgeable about the program, supportive of its implementation, provides the necessary release time for teachers to participate in the school-based team, and who holds teachers accountable for participation in the intervention, and for quality implementation of state CCR standards.

Classroom-level implementation of the standards-based reform intervention will be subjected to a rigorous RCT appropriately powered to detect effects on student achievement as mediated by alignment of instruction to CCR standards. If the intervention is as effective as we expect, the Center will provide an existence proof that standards-based reform can be a powerful, positive force for change in instruction at scale and with achievement benefits for all students.

The Center also will have engaged policy makers, education practitioners, and researchers in national discussions of the Center's work and its findings. The partnerships we will build with states, districts, schools, and teachers will be the key to our outreach, leadership and communication strategies.

Appendix B

Figure B1. Conceptual Framework



The Policy Attributes define the nature of state and district policy. Our intervention addresses the policy attributes at the teacher/ classroom level.

How to Read Figure B1. Figure B.1shows how we envision our FAST intervention fitting in the overarching system of standards-based reform. We highlight feedback and guidance to teachers as the core mechanism for bringing standards-based reform successfully to the classroom. Our intervention is designed to target the classroom level, providing guidance that is specific and consistent, backed by authority, with appropriate power, and stability. The intervention influences school-level instructional guidance in its integration into the overall guidance system teachers receive. It directly influences teachers, given that teachers are the target of the intervention.

While we depict linear relationships for ease of presentation and to reflect the systematic nature of our research, we recognize the complex, nonrecursive nature of the model as well as the interactions present in real-world education policy.

The contextual factors that we know influence any school reform attempt—school and community demographics, school climate, teacher experience and content knowledge, and others—are implicit in the system. We include these potential moderators in our implementation, longitudinal, and intervention studies. We focus our figure to highlight the key malleable factors and mediators between standards and better student outcomes.

Table B1. Timeline for Implementation, Longitudinal, Measurement and Intervention Studies

	Year 1			Year 2			Year 3			Year 4			Year 5		
	Su	F	S	Su	F	S	SU	F	S	Su	F	S	Su	F	S
Implementation Study															
Draw national probability sample of principals and teachers															
Conduct State interviews															
Code and analyze interview data															
Develop and refine survey															
Administer survey															
Clean and analyze survey data															
Conduct case studies															
Write papers, briefs, etc.															
Longitudinal Study															
Develop and populate state inventory															
Collect longitudinal outcome data															
Update state inventory															
Add additional year of outcome data															
Conduct longitudinal analysis															
Write papers and research briefs															
Develop and update 50-state policy database															
Measurement Study															
Create survey/observation measures															
Pilot survey/observation measures															
Revise survey/observation measures															
Use measures in RCT pilot															
Revise measures based on RCT pilot															
Implement measures in RCT															
RCT to Test Intervention															
Pilot intervention															
Refine intervention															
Finalize intervention															
Randomize schools															
Interview district & school admin.															
Implement intervention cohort 1															
Implement intervention cohort 2															
Analyze data and write results															
Share results with school & district															

Table B2. State Partners Overview

State	Adopted CCSS	Assessment	Implementation Context	Percent Students with IEPs	Percent ELLs	Geographic Region
Kentucky	Yes	Other	Aggressive implementation/political support	14.5%	2.5%	South
Massachusetts	Yes	PARCC	Aggressive implementation/political support	17.4%	6.8%	Northeast
Missouri	Yes	Smarter – Balanced	Contested support for implementation	13.6%	2.7%	Midwest
Texas	No	Other	Has their own standards system	8.8%	14.9%	Southwest
Ohio	Yes	PARCC	Strong implementation	4.9%	2.2%	Midwest

Figure B2. Data Analysis Approach for Longitudinal Study

Our modeling approach is formulated as follows:

$$Y_{jt} = \sum_{t=0}^{T} \beta_t year_t + \beta_{trt} trt\_post_{jt} + \sum_{k=0}^{N} \beta_j s_j$$

Where
Y_{jt} is the outcomes for state *j* at time *t*;
$year_t$ is a vector of indicator variables for each year in the study period (year fixed effects for both pre- and postintervention years);
β_t is a vector of coefficients associated with each of the year fixed effects;
trt_post_{jt} is an indicator for if state *j* is implementing CCR standards or assessment in year *t* (this variable is always 0 for non0implementing states, 0 for treatment states in the pretreatment period, and 1 for treatment stats in the posttreatment period);
β_{trt} is the difference in average performance between treatment states and comparison stated in the period after the intervention was implemented, net of school and year fixed effects;
s_j is a vector of state indicator variables (state fixed effects);
β_j is a vector of coefficients associated with each of the state fixed effects.

In addition to the year and state-fixed effects models, analyses that model a linear pretreatment trend will also be implemented to examine the robustness of results to modelling assumptions.

Table B3. Validity Threats to Our Interrupted Time Series Design

Validity Threat	Potential Problem	Robustness to the Threat/How We Will Address It
History	Forces other than the standards introduction, such as district-wide policy changes, might have influenced outcomes at the same time as the intervention was introduced (Shadish, Cook, & Campbell, 2002).	We use **timing of implementation** and **how challenging previous standards were** to address the threat of history. The variation in timing of implementation across states strengthens the design to the threat of history by examining whether the same effect is present when the policy is implemented at different points in time. Presumably, alternative explanations for perceived effects, such as other changes in federal or state policy and economic shifts, would not consistently co-occur with implementation across states. In addition, dividing the states based on whether they had high or low standards prior to CCR standards implementation allows us to create a comparison group within the implementing states. The comparisons alleviate the threat of history in so far as the alternative explanation for any perceived effect could be explained by forces that would have affected the comparison schools as well (e.g., other changes in national policy or economic conditions).
Instrumentation	If there was a change in how administrative data were collected that coincided with the implementation of the program, it could pose a threat that could limit interpretation of the effect of the program as causal. This is a concern if we relied on state assessments, which are generally changing as a result of CCR implementation.	The study relies on measures that are consistently collected nationally from year to year, so we do not expect instrumentation to be an issue.
Selection	With ITS the pre-intervention time series serves as the counterfactual for what achievement would have been if the program had not been implemented; therefore, selection bias can be a problem in these designs if there is an abrupt change in the population composition at the time of the intervention (Shadish, Cook, & Campbell, 2002).	Given the scale of the implementation, we do not expect this to be a problem, but we cannot rule it out. Therefore, we will include state-level, time-varying composition variables in the analytic models as appropriate.

Appendix C

Table C1. Description of the Surveys of Enacted Curriculum (SEC)

History	Since 2002-03, over 40,000 SEC surveys have been completed by K-12 teachers for use in research studies or by districts or schools to examine the enacted curriculum. The SEC is a well-studied instrument that has been used in its present form in dozens of research studies. It is based on nearly 3 decades' research, originating with efforts by researchers at Michigan State University's Institute for Research on Teaching to understand teachers' content decisions in elementary school math (for a history of the SEC, see Porter, 2002; Porter et al., 1988, 1993). The surveys are available online at http://seconline.wceruw.org/secWebHome.htm
Content	The content portion of the SEC asks teachers to report the content they taught during a given time period (most often a full year). For this portion, teachers first identify from a list of 133 to 211 topics, depending on academic subject, all the topics they taught in the previous academic year in a target class. For each topic taught, they indicate the number of lessons on a scale of no lessons, less than one lesson, one to five lessons, and more than five lessons. They then allot the instructional emphasis for each topic among five levels of cognitive demand; the cognitive demand levels are listed and defined below. The year's instruction is turned into a matrix of proportions, with each proportion indicating the percent of the year's instruction dedicated to each topic-by-cognitive demand combination (in SEC language, a "cell").
Data	Teacher data from the SEC can be compared with content analyses of state standards, assessments, or curriculum materials. Comparisons of instruction with the content of these documents are for the purpose of calculating instructional alignment. The content analysis procedures are as follows (Porter, 2002; Porter et al., 2008): Content analysts are trained subject matter experts. They analyze each document at the finest-grained level of detail possible (for standards, these are usually objectives; for tests, these are usually items). Working independently, content analysts examine each objective or item and place it into between 1 and 6 cells in the SEC framework. Multiple cells are allowed because objectives or items often tap multiple topics and/or cognitive demand levels—in the case of multiple cells, the weight of the objective is evenly divided among the target cells. Each objective or item is weighted evenly unless the document indicates otherwise, as this is the most replicable and defensible approach.
Ratings	After each rater has analyzed each objective or item, the ratings are converted into proportions indicating the percent of the total standards (or test, or textbook) content in each SEC cell. These proportions are then averaged across raters, to arrive at the final content analysis. Generalizability theory studies indicate that the content analyses are reliable (generalizability coefficients greater than .75) with 3 to 4 raters (Porter, 2002; Porter et al., 2008); all content analyses used here will have at least that many raters.
Research Support	Existing research lends support to the instructional measures derived from the SEC surveys. Research conducted during the development of the instrument indicated that teachers were comfortable making distinctions at the fine-grained level of topic-by-cognitive demand, that teacher ratings of content coverage of individual lessons correlated moderately with ratings by external observers, and that teacher reports of content coverage over a semester or year correlated highly with aggregated daily logs (Porter et al., 1993). Furthermore, one study of instructional alignment based on the SEC found significant correlations with value-added to student achievement ($r = 0.45$) (Gamoran et al., 1997). Together, these studies suggest that the SEC surveys measure important elements of teachers' content coverage. Compared to other methods of measuring the content of instruction over a full year and estimating alignment, the SEC has the most validity and reliability evidence to support it.

Table C2. Cognitive Demand Levels in the SEC Framework

Mathematics

B. Memorize

- Recite basic mathematics facts; Recall mathematics terms and definitions; Recall formulas and computational processes

C. Perform Procedures

- Use numbers to count, order, or denote; Do computational procedures or algorithms; Follow procedures/instructions; Make measurements, do computations; Solve equations/formulas, routine word problems; Organize or display data; Read or produce graphs and tables; Execute geometric constructions.

D. Demonstrate understanding

- Communicate new mathematical ideas; Use representations to model mathematical ideas; Explain findings and results from data analysis; Develop/explain relationships between concepts; Explain relationship between models, diagrams, and other representations.

E. Conjecture, generalize, prove

- Determine the truth of a mathematical pattern or proposition; Write formal or informal proofs; Analyze data; Find a mathematical rule to generate a pattern or number sequence; Reason inductively or deductively; Use spatial reasoning.

F. Solve non-routine problems, make connections

- Apply and adapt a variety of appropriate strategies to solve problems; Apply mathematics in contexts outside of mathematics; Recognize, generate, or create patterns; Synthesize content and ideas from several sources.

English Language Arts (ELA)

B Memorize/recall

- Reproduce sounds or words; Provide facts, terms, definitions, conventions; Locate literal answers in text; Identify relevant information; Describe.

C. Perform procedures/explain

- Follow instructions; Give examples; Check consistency; Summarize; Identify purpose, main ideas, organizational patterns; Gather information.

D. Generate/create/demonstrate

- Create/develop connections among text, self, world; Recognize relationships; Dramatize; Order, group, outline, organize ideas; Express new ideas (or express ideas newly); Develop reasonable alternatives; Integrate with other topics and subjects.

E. Analyze/investigate

- Categorize/schematize information; Distinguish fact and opinion; Compare and contrast; Identify with another's point of view; Make inferences, draw conclusions; Predict probable consequences.

F. Evaluate

- Determine relevance, coherence, internal consistence, logic; Assess adequacy, appropriateness, credibility; Text conclusions, hypotheses; Synthesize content and ideas from several sources; Generalize; Critique.

Table C3. Calculating Alignment

We will use the SEC data to calculate several alignment indices. Alignment is calculated by comparing any two content matrices, for instance comparing the matrix representing teachers' instruction with the matrix representing the content of standards. The formula for alignment (Porter, 2002) is:

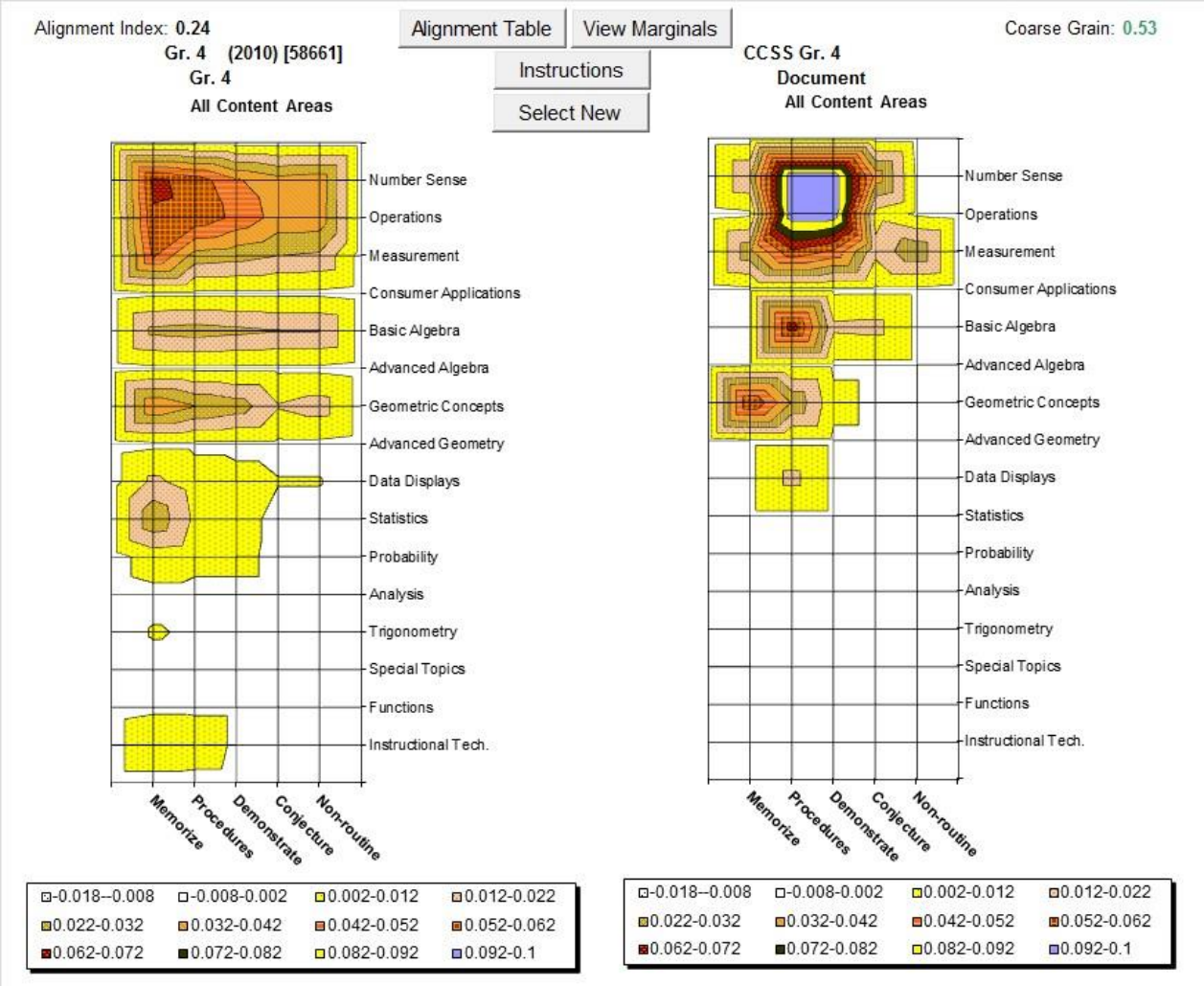
$$\text{Alignment} = 1 - (\Sigma_i |x_i - y_i|) / 2$$

Here, x_i and y_i represent the proportion of content in cell i of document x (e.g., teachers' instruction) and document y (the standards), respectively. Mathematically, this formula is equivalent to the sum of the cell-by-cell minima.

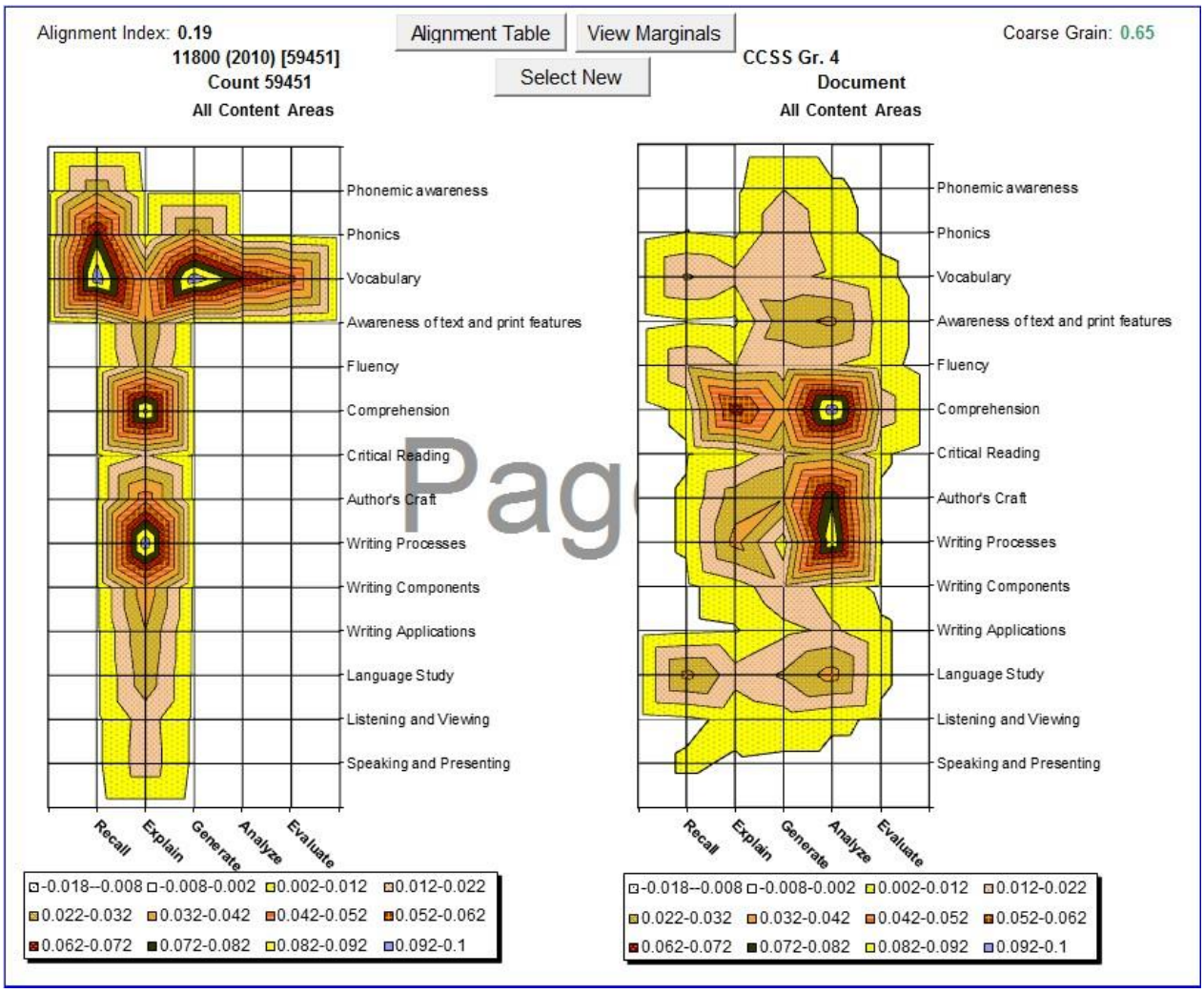
Alignment ranges from 0 to 1 and indicates the proportion of content in exact agreement at the cell level. Previous results indicate that alignment indices for the alignment of instruction with standards or assessments are normally distributed with means below 0.50 (Polikoff, 2012a; Porter et al., 2007). Recent work has also developed approaches to conducting hypothesis tests using alignment data, and we will use those as appropriate (Fulmer & Polikoff, 2014; Polikoff & Fulmer, 2013).

SEC data can also be used to create content maps to visualize instructional alignment or misalignment (See Figure C.1 below). These maps are generally created in Microsoft Excel software, and they take the form of either topographical maps or tile charts. Sample maps for both mathematics and ELA are shown below. The peaks and valleys in each map indicate relative areas of emphasis, with the darkest sections indicating the most emphasis. Note that Excel assumes these are continuous variables when creating the content maps; thus, while the maps are correct at the intersection of a topic and level of cognitive demand, they are not interpretable "between" topics or levels of cognitive demand. Teachers often find the maps to be powerful tools to understand their alignment or misalignment.

Figure C1. Content Maps



Sample content maps for fourth grade math teacher (left) and Common Core grade 4 math standards (right)



Sample content maps for fourth grade ELA teacher (left) and Common Core grade 4 ELA standards (right)

Table C4. Example Description of Intervention Feedback

A teacher is working in a district that is implementing the CCSSM. She has been teaching her students to multiply multi-digit whole numbers by a one-digit number. She is addressing standard 4.NBT.5:

- Ø The following describes what she asks her students to do:
  1) Multiply a whole number of up to four digits by a one-digit whole number, and multiply two two-digit numbers, using strategies based on place value and the properties of operations. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.
- Ø Teacher instruction and data reporting activities:
  1) She covered one-digit by one-digit multiplication on Monday, one-digit by two-digit multiplication on Tuesday, one-digit by three-digit multiplication on Wednesday, one-digit by four-digit multiplication on Thursday, and reviewed this work on Friday. Each day she taught students the rules for using the traditional algorithm. The problems assigned to students during the lesson and during classwork required students to use the multiplication algorithm to solve decontextualized problems.
  2) She completed the weekly log and was video recorded on Tuesday. Her video-recorded lesson began with review problems in which students added multi-digit numbers and shared their answers, continued with a presentation in which the teacher showed students how to use the multiplication algorithm to multiply one-digit by two-digit numbers and ended with students practicing one-digit by two-digit multiplication using the algorithm at their desks.
- Ø The content-focused feedback report for that week included:
  1) A content map displaying her coverage of standard 4.NBT.5 compared to what ideal coverage would look like (see Figure C1 for an example of the map).
  2) An easy to understand text-based explanation that her instruction focused on the multiplication algorithm, and that students used equations to represent the multiplication procedure as an equation. The feedback would explain that she did not address all aspects of the standard. At no point were place-value and/or properties of operations (in this, case distributive property) discussed. Rectangular arrays and area models were not used to illustrate the multiplication of the two numbers. As a result, the emphasis was on memorization.
  3) The report would direct the teacher's attention to the following online resources, which demonstrate examples of teaching the standard, addressing the areas where the teacher showed a lack of alignment:
     1. Support to understand the mathematics and representations involved in this standard: http://secc.sedl.org/common_core_videos/grade.php?action=view&id=612
     2. Resources to use in teaching: https://learnzillion.com/lessonsets/360-multiply-multidigit-whole-numbers; file:///C:/Users/tsmith/Downloads/math-g4-m3-topic-c-overview.pdf

Table C5. Data collection Schedule for RCT

	Spring Year 2	Fall Year 3	Spring Year 3	Fall Year 4	Spring Year 4
Intervention		X	X	X	X
Survey of Enacted Curriculum (SEC)	X		X		X
Student Scores on State Assessment	X		X		X
Student Scores on Center-Administered Assessment			X		X

Figure C2. Model for Assessing Student Impact

The Model

$$Y_{ijk} = \sum_m \sum_n \gamma_{0mn} B_{mnk} + \sum_m \gamma_{1m} T_k D_{mk} + \gamma_2 Y\text{-}1_{ijk} + \gamma_3 Y\text{-}1_k + \sum_l \alpha_l X_{lijk} + \mu_k + \upsilon_{jk} + \varepsilon_{ijk}$$

Where:

Y_{ijk} = achievement measurement for student i from class j in school k,

B_{mnk} = 1 if school k is in block n in district m (m = 1 to 10) and 0 otherwise,

D_{mk} = 1 if school k is in district m (m = 1 to 10) and 0 otherwise,

T_k = 1 if school k is assigned to receive the FAST treatment and 0 otherwise,

$Y\text{-}1_{ijk}$ = baseline score for student i from teacher j in school k,

$Y\text{-}1_k$ = average baseline state assessment score for school k,

X_{lijk} = demographics for student i from teacher j in school k,

μ_k , υ_{jk} , ε_{ijk} = school-, class-, and student-level random errors.

The weighted average γ_1 of the estimated γ_{1m} coefficients for the 10 districts (using the number of treatment schools in each district as weight) is the estimated effect of the FAST intervention on student achievement for the average treatment school in the study sample.