Instructional Alignment under No Child Left Behind
Author(s): Morgan S. Polikoff
Reviewed work(s):
Source: *American Journal of Education*, Vol. 118, No. 3 (May 2012), pp. 341-368
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/664773

Instructional Alignment under No Child Left Behind

MORGAN S. POLIKOFF
University of Southern California

The alignment of instruction with the content of standards and assessments is the key mediating variable separating the policy of standards-based reform (SBR) from the outcome of improved student achievement. Few studies have investigated SBR's effects on instructional alignment, and most have serious methodological limitations. This research uses content analyses of state standards and assessments and survey data on more than 27,000 teachers' instruction in mathematics, science, and English/language arts (ELA) to investigate changes in instructional alignment between 2003 and 2009. Fixed-effects models indicate that alignment in grades K–12 mathematics increased by approximately 0.19–0.65 standard deviations, depending on the grade and target. Alignment also increased to grades K–12 standards in ELA and grades 3–8 standards in science, though the magnitudes were smaller. Multiple alternative specifications support the findings of increased alignment. Implications for research and SBR policy are discussed.

Throughout the last 25 years in American kindergarten through grade 12 education, standards-based reform (SBR) has been a guiding policy. The theory of change starts with coherence, beginning with the construction of content standards in core academic subjects to specify what students are to know and be able to do (Smith and O'Day 1991). Coherence is paired with explicit goals—most often student learning targets measured by aligned, standardized assessments. With appropriate supports and accountability measures, the theory proposes that teachers will align their instruction with the standards and assessments, and student learning will improve (e.g., Clune 1993; Smith and O'Day 1991). Thus, instructional alignment is the mediating variable between the policy of SBR and the outcome of improved student learning. The No Child Left Behind Act of 2001 (NCLB; Public Law 107-110) supplements SBR with test-based accountability for schools and districts using the carrot of federal dollars to essentially mandate the basic tenets of standards and

Electronically published March 5, 2012
American Journal of Education 118 (May 2012)

0195-6744/2012/11803-0003$10.00

aligned assessments. Improving the alignment of teachers' instruction with standards and assessments is therefore a direct goal of the theory of SBR and of the federal and state laws that moved SBR to the classroom.

Since the early years of SBR, researchers have conducted dozens of studies on the policy's effects on student achievement (e.g., Betts and Danenberg 2002; Carnoy and Loeb 2004; Center on Education Policy 2006; Dee and Jacob 2011; Grissmer and Flanagan 2001; Hanushek and Raymond 2005; Jacob 2005) and teachers' instruction (e.g., Koretz, Barron, et al. 1996; Koretz, Mitchell, et al. 1996; Koretz et al. 1994; Luna and Turner 2001; Pedulla et al. 2003; Shepard and Dougherty 1991; Smith 1991; Stecher and Barron 2001). The highest-quality studies of achievement effects have generally treated instruction as a black box (Carnoy and Loeb 2004; Dee and Jacob 2011; Hanushek and Raymond 2005). In contrast, the studies of instructional effects have generally focused on other instructional variables (e.g., cognitive demand emphasis, test preparation) and/or measured alignment using methods of dubious validity (e.g., Koretz, Barron et al. 1996; Koretz, Mitchell, et al. 1996; Koretz et al. 1994; Luna and Turner 2001; Pedulla et al. 2003; Shepard and Dougherty 1991; Smith 1991; Stecher and Barron 2001). The neglect of high-quality measures of change in instructional alignment in studies of SBR's effects is surprising because the content of instruction is central to the policy and is highly predictive of gains in student learning (Cueto et al. 2006; Gamoran et al. 1997; Sebring 1987).

A more useful approach than that taken in prior analyses is to use data on the self-reported content of teachers' instruction collected at particular points in time and investigate change over time. These data have several important advantages over data used previously: (*a*) they require no assumptions about teachers' understanding of the content messages of standards and assessments, and (*b*) they allow for direct comparison with content analyses of standards and assessments to estimate alignment. No prior investigation has used data of this quality to investigate changes in alignment.

To evaluate the degree to which instruction has become more aligned with standards and assessments, I use self-reported survey data on teachers' instruction. I compare these data with the content of standards and assessments as content analyzed using the Surveys of Enacted Curriculum (Porter 2002) to address the question, To what extent have teachers aligned their instruction with standards and assessments in English/language arts, mathematics, and

---

MORGAN S. POLIKOFF is an assistant professor of education at the University of Southern California's Rossier School of Education. He obtained his PhD in 2010 from the University of Pennsylvania. His research focuses on the design of standards-based and accountability reforms and their effects on instruction and student outcomes.

science since 2003? The results provide evidence of the extent to which teachers have changed their instruction in accordance with the theories underlying SBR over the years 2003–9.

Background: The Effects of Standards-Based Reform on Instructional Alignment

Under NCLB's model of SBR with test-driven accountability, standards represent the instructional target for teachers (NCLB Act; Smith and O'Day 1991). State assessments are intended to reinforce the content messages of the standards and provide external motivation for teachers to teach the content specified therein. Thus, instructional alignment should increase because the standards signal to teachers the important content to teach and because teachers will be motivated to teach that content to improve student performance on the assessments to avoid sanctions. Therefore, we should expect increasing instructional alignment to standards in all grades and subjects but greater increases in tested grades and subjects. Furthermore, alignment should increase more when the standards and assessments are themselves well aligned and mutually reinforcing (Polikoff 2012).

Several qualitative and quantitative studies from multiple settings have investigated the extent to which alignment is occurring. Many survey studies, primarily surveys of representative samples of teachers in individual states, indicate that teachers report increasing instructional alignment. One source of evidence is questions in which teachers are directly asked if they are aligning or have aligned their instruction. For instance, Kentucky teachers were asked how much they had focused on improving the match between the content of their instruction and the content of the state test (Koretz, Barron, et al. 1996), with 87 percent indicating a moderate amount or a great deal of focus. Though direct comparisons across studies are not possible because each study asks different questions, the proportions of teachers indicating increased alignment is large: from 76 percent or more of teachers in a national sample (Pedulla et al. 2003) to 86–90 percent in Colorado (Taylor et al. 2002) and 96 percent in California and Georgia (Hamilton and Berends 2006). Regardless of the state, subject, or grade, most teachers in tested subjects report efforts to improve alignment.

Measures of change in alignment based on survey questions of this sort have several important limitations. One substantive limitation is that it is unclear how teachers interpret the term "alignment." Survey questions asking teachers about their instructional alignment assume that teachers accurately understand the content messages of standards and assessments, which is likely not true (Hill 2001; McDonnell 2004). Thus, teachers might perceive increases in alignment to standards and assessments, but such perceptions might be based on misunderstandings of the content messages of those sources.

Methodologically, there are several problems with measures of teacher-reported change in alignment. First, it is more difficult for teachers to validly report on their own instructional changes than it is to report on instruction provided in a given time (Desimone et al. 2005). Indeed, surveys of perceived behavioral change result in larger estimates of change than other methods such as pre-post surveys (Lam and Bengo 2003). Thus, assuming teachers accurately understand "alignment" and the content messages of standards and assessments, relying on self-reported change in alignment may overestimate the actual change in alignment in practice.

Even more troublesome, research indicates that surveyed individuals are likely to report changes in behavior, real or not, when such changes are expected (Ross 1989; Ross and Conway 1986). These "theory-driven reconstructions" are a flaw of surveys or interviews that ask respondents to describe behavior changes resulting from particular programs or policies. Schwarz and Oyserman (2001) argued that survey reports about behavioral change are highly troublesome for this reason: "Asking program participants to report on how their behavior has changed over the course of the intervention . . . is likely to result in theory-driven reconstructions. These reconstructions are useless as measures of objective change, although they may be of interest as measures of participants' subjective perceptions. To assess actual change, we need to rely on before-after, or treatment-control, comparisons" (144). Though this quote was in the context of program evaluation, it is certainly applicable for evaluating a salient policy such as NCLB. That is, we can expect teachers to report efforts to increase instructional alignment because alignment is a well-known goal of SBR policy.

A third methodological concern is in the use of one-item survey questions. The limitations of one-item scales in terms of reliability are obvious, and research suggests avoiding their use in measuring instruction (Mayer 1999). In short, the validity of extant survey data on teacher-reported change in alignment for representing actual increases in alignment is weak.

A second way change in alignment with standards and assessments has been gauged is through teacher-reported change in coverage of focal content from the standards or assessments. For instance, another Kentucky study asked teachers how often they covered core content areas of "numbers and computation," "algebraic ideas," and so forth, given that these were focal, tested topics in that state (Stecher and Barron 2001). Similar methods were used in a Maryland study (Koretz, Mitchell, et al. 1996). These examples indicate that 50–80 percent of teachers reported increasing coverage of these focal topics, suggesting improved instructional alignment to standards at the individual strand level. However, these findings are of limited utility because they do not report how each strand is represented in the standards and assessments, and they do not ask about the fine-grained topics in standards and assessments. Thus, teachers might be increasing their emphasis on algebra skills, a focal topic for the state assessments, but they could certainly be increasing focus on particular algebra skills that are not the focal objectives in the standards.

Despite the methodological and substantive concerns with the extant survey literature on changes in alignment, the results are quite consistent that teachers report such changes taking place. However, the consistency of the findings from survey studies is not always reflected in studies based on interviews or observations. For example, case studies of elementary teachers in Arizona revealed that they were not aligning their core instruction but were rather creating separate test-preparation classes to improve alignment (Smith 1997). In North Carolina and Kentucky, teachers were asked to submit assignments that were "most aligned" to the standards (McDonnell 2004). Analysis of the submitted assignments revealed often substantial misalignment with the instructional target. Overall, these and other qualitative findings largely support the claim that uneven implementation of educational reforms has multiple causes, including the sometimes conflicting role of district policies (Wong et al. 2003), the misinterpretation of content specifications (Hill 2001), and the tendency to graft new approaches onto existing ones (Cohen 1990). In short, while there are numerous reports of increased alignment under SBR, there are counterexamples or reasons why teachers might overstate their instructional alignment and substantive and methodological problems with the extant research.

Method and Data

To add to the literature on content effects of standards-based reform, this analysis uses secondary data from the Surveys of Enacted Curriculum (SEC) content taxonomies (Porter 2002). The taxonomies measure the content of teachers' instruction at the intersection of topics and levels of cognitive demand. The surveys are the result of a more than 20-year line of research (e.g., Porter et al. 1988). The taxonomies were developed over time with the input of content experts, arriving in their present form in the early 2000s (for more details on the development of the tools, see Porter [2002]).

Data on the topics and cognitive demand emphases of teachers' instruction come from more than 27,000 collected surveys of K–12 mathematics, science, and English/language arts (ELA) teachers. The data have been collected over the years 2003–9 for other studies or for evaluating alignment in states, districts, or schools. The year 2003 was the first year instructional data were available in an SEC language approximating the current format. The data are mainly cross-sectional; there are several thousand cases in which teachers completed the SEC in multiple years, but most of the surveys are from teachers who completed the SEC only once. Both types of data are used here, together in the main analysis and separately in alternative specifications.

Data on the content of standards and assessments also come from the SEC. Since 2002, trained coders have analyzed state standards and assessment documents in the three content areas, using the same taxonomies. As of the writing of this article, 187 standards documents (i.e., a state and grade) and 87 assessments have been content analyzed in mathematics. The figures are 166 and 89 for standards and assessments in ELA and 136 and 46 for standards and assessments in science.

The data for both instruction and content analyses are in the form of two-dimensional matrices of topics by levels of cognitive demand. For each subject area there is a list of broad topics. Underneath each broad topic are fine-grained topics. For instance, in science, one broad topic is animal biology, with fine-grained topics including nutrition, circulation, and excretion. There are 183 fine-grained topics in mathematics, 133 in ELA, and 211 in science. Perpendicular to the topics is a set of five levels of cognitive demand. These are different across content areas but generally range from memorizing/producing facts to conjecturing or proving. Thus, there are 915 "cells" in the SEC taxonomy for mathematics, 665 for ELA, and 1,055 for science. Of course, there are many alternative ways to define instructional content, and alignment would differ on the basis of the particular content language used. However, the content languages used here have been developed over time with input from experts, and they have been widely used for nearly a decade. The actual SEC surveys are downloadable on the SEC website (see http://seconline.wceruw.org/resources.asp).

When teachers complete the SEC, they are asked to think about a particular target class and time period. The time period is generally a semester or a full year. Teachers always complete enough surveys to represent the full year's instruction, and the surveys for each teacher are always aggregated to represent a full year by weighting the instructional content on each survey by the number of instructional days represented. Once the teacher has a time period and a target class in mind, the task is to identify (*a*) all the fine-grained topics he or she did not teach to the target class in the time period, (*b*) the number of lessons for each topic taught (on a scale of none, less than one lesson, one to five lessons, and more than five lessons), and (*c*) the relative emphasis on each level of cognitive demand for each topic taught (on a scale of none, less than 25 percent, 25–33 percent, and more than 33 percent). After the surveys are aggregated to represent a year, the values in each SEC cell indicate proportions of total instructional time spent on each topic by cognitive demand combination. Analyses of data quality indicate that teacher reports of content coverage on year-end surveys correlate well with content reports from daily logs and that content coverage reported by teachers correlates well with ratings by external observers (Porter et al. 1993).

Data from the content analyses of the standards and assessments are in the same form—proportions of total content in each SEC cell. Content analysts are subject matter experts in the particular subject of the document. Analysts are trained before conducting their independent content analyses. For standards, analysts categorize at the finest-grained level of detail available in the document; these are often called "objectives." Each coder places each objective into between one and six cells in the SEC taxonomy. Multiple cells are allowed because some objectives are seen as covering multiple topics and/or levels of cognitive demand. The weight for the objective is then evenly divided among the target cells. For assessments, the methods are identical except that test items are allowed to be placed in only three cells (test items are not as broad as objectives), and the number of score points for the test item is evenly divided among the target cells. As an example, if a two-point constructed response item is placed into three cells, each cell would receive two-thirds of a point.

The result of the content analysis is a matrix of proportions for each rater indicating the proportion of total standards (or assessment) content on each cell in the SEC framework. The matrices are then averaged across raters to give the final content analysis (for more detail on content analysis procedures, see Porter et al. [2008]). The quality of the content analysis data is strong; generalizability theory d-studies indicate that content analyses using four raters have reliability of 0.70 or higher in 90 percent of documents analyzed and 0.80 or higher in 75 percent of documents analyzed in mathematics and ELA (Porter et al. 2008).

One complication with the SEC data is that the content languages changed slightly between 2006 and 2007. To address this, every topic that did not appear before and after the transition was deleted from the data set, and the remaining proportions were normalized to sum to one. In so doing, an average of 9.6 percent of the typical math teacher's content, 8.7 percent of the typical ELA teacher's content, and 11.8 percent of the typical science teacher's content was lost. Thus, the alignment indices reported here are actually based on roughly 90 percent of the typical teacher's instruction. To the extent that this recentering would bias the alignment index, the effect should be to bias alignment downward over time because the presence of more topics to choose from in completing the survey (after 2006) should spread out teachers' instruction across those topics. The potential bias of this deletion of content is studied in the sensitivity analyses.

Alignment.—The primary use of the SEC data and the content analyses of standards and assessments is the calculation of alignment. A simple alignment index (Porter 2002) is:

$$\text{Alignment Index 1} = 1 - \frac{\sum_i |x_i - y_i|}{2},$$

where x_i is the proportion of content in cell i of matrix x (the matrix representing a teacher's instruction) and y_i is the proportion of content in cell i of matrix y (the matrix representing a set of content standards or an assessment). Alignment is calculated at the fine-grained level of detail, that is, using data from 915, 665, and 1,055 cells, depending on subject. The index ranges from zero (perfect misalignment) to one (perfect alignment), with in-between values representing the proportion of content in common. Thus, this alignment index considers content to be aligned if it agrees on topic, cognitive demand, and proportion of total content. The distribution of instructional alignment is symmetric and roughly normal, with means below 0.50 (Porter et al. 2005). There is no absolute standard for alignment: the indices are best used for comparisons.

The alignment index has been used for descriptive analyses of state standards and assessments (Polikoff et al. 2011; Porter 2002; Porter et al. 2008), as an independent variable in estimating student achievement level and gains (Gamoran et al. 1997; Smithson and Collares 2007), and as a dependent variable in a randomized experiment (Porter et al. 2005, 2007). Perhaps the most powerful validity evidence for the SEC is based on the study by Gamoran et al. (1997). In that study, student achievement gains in high school mathematics were regressed on course indicators and an alignment index based on a previous version of the SEC. The results indicated that content alignment, when defined as above, was correlated 0.45 with value-added scores. Furthermore, content coverage explained the vast majority of between-class differences in achievement gains.

While the main alignment index has been well studied and shown to be highly predictive of student learning gains, it is not the only way to define instructional alignment. To verify that the results obtained are not merely an artifact of the particular way alignment is defined, a second alignment index considers teachers' instruction to be aligned if it focuses on any cell in the SEC framework that is covered at all on the standards or assessment. If x_i is the teacher's instruction and y_i is the standards or assessment document, the equation is

$$\text{Alignment Index 2} = \sum_{i|y_i>0} x_i.$$

In other words, this second alignment index represents the proportion of the teachers' instruction that is on SEC cells that are covered in the target document. Unlike the first alignment index, this index describes instruction as aligned even if there is not agreement on the proportion of total content represented in the target SEC cell.

Sample.—Table 1 summarizes the samples and provides descriptive statistics. Across content areas there are large samples of teachers from multiple states and grades in the full sample (the first two columns). Samples are smaller for teachers with multiple surveys but still roughly 1,000 or more per subject (the third and fourth columns). Grade bands (K–2, 3–8, 9–12) are used for all analyses presented here since (*a*) grades within each band should be similarly influenced by SBR because of the grades at which NCLB testing is required, (*b*) the power to detect effects would be low if individual grades were used, and (*c*) the interpretation of results would be overly complicated with 13 sets of regressions for each subject for each analysis.

Each teacher in the sample falls in a specific state, grade, and year of administration. Thus, to fully describe the data would require a three-dimensional table with each of those variables on an axis. It is not possible to present such a table here, but some summaries of the data are useful in understanding the nature of the teacher samples. First, the list of states represented by at least one teacher for each analysis is presented in the appendix. Teachers are not distributed evenly across states. For mathematics standards, the best-represented states are Illinois, Ohio, Oklahoma, and Oregon, constituting 56 percent of the full teacher sample. For mathematics assessments, teachers from Ohio and Oklahoma constitute 52 percent of the teacher sample. For ELA standards, Ohio teachers alone constitute 54 percent of the sample, and New York, Illinois, and Indiana teachers constitute an additional 30 percent. Ohio and New York teachers constitute 74 percent of the ELA assessment sample. In science, 59 percent of the teacher sample for standards is from Illinois and Oklahoma, along with 62 percent of the sample for assessments.

The teachers are also distributed unevenly across grades and years. The plurality of teachers in the full sample is from grades 3–8: 76 percent and 85 percent for mathematics standards and assessments, 49 percent and 94 percent for ELA standards and assessments, and 80 percent and 78 percent for science standards and assessments. In terms of survey year, most teachers completed the SEC between 2005–6 and 2007–8: 67 percent and 61 percent for mathematics standards and assessments, 72 percent and 74 percent for ELA standards and assessments, and 63 percent and 62 percent for science standards and assessments.

Because the samples are not probability samples, an important question for gauging the generalizability of the results is the extent to which the sample characteristics indicate that the teacher samples are similar in composition to the population of US K–12 teachers. To address this question, descriptive statistics on the teacher sample were compared with national figures from the *Digest of Education Statistics* (Snyder et al. 2008). The variables chosen were those that were included in both the digest and the SEC survey. Because of the large samples, nearly all sample averages are statistically different from

TABLE 1

Comparison of Sample Descriptive Statistics with National Averages

VARIABLE	FULL SAMPLE		LONGITUDINAL SAMPLE		POPULATION
	Alignment to Standards	Alignment to Assessments	Alignment to Standards	Alignment to Assessments	
Mathematics:					
Nonwhite students (%)	33.6	32.4	32.7	30.6	43.5
Elementary (K–5) class size	20.6	20.3	20.9	20.1	20.4
Secondary (9–12) class size	21.7	20.2	20.4	19.9	24.7
Female secondary (9–12) teachers (%)	63.2	63.9	68.2	69.3	54.9
Teachers with highest degree bachelor's (%)	50.1	51.7	52.6	57.6	50.8
n	10,814	5,071	3,210	1,502	
n states	23	12	18	9	
ELA:					
Nonwhite students (%)	43.8	45.2	41.0	39.5	43.5
Elementary (K–5) class size	18.7	18.6	18.7	18.2	20.4
Secondary (9–12) class size	22.0	22.0	22.4	21.3	24.7
Female secondary (9–12) teachers (%)	78.1	78.3	79.1	80.0	73.9
Teachers with highest degree bachelor's (%)	38.3	35.7	40.9	42.4	50.8
n	12,522	4,431	3,964	1,028	
n states	19	10	9	4	
Science:					
Nonwhite students (%)	37.0	33.9	34.6	34.6	43.5
Elementary (K–5) class size	22.0	22.0	21.7	22.0	20.4
Secondary (9–12) class size	22.1	22.1	21.0	22.2	24.7
Female secondary (9–12) teachers (%)	61.3	58.3	68.2	67.3	48.7
Teachers with highest degree bachelor's (%)	55.8	53.1	55.8	53.0	50.8
n	5,346	1,837	2,740	886	
n states	22	12	16	9	

NOTE.—Population data are from the *Digest of Education Statistics* (Snyder et al. 2008).

national averages; however, the descriptive statistics in table 1 indicate that the samples are close in absolute terms to national averages. While there are some imperfect matches (e.g., the samples of mathematics and science classes have 5–10 percent more white students than is typical), the results suggest that the classes in the data set are not extraordinary on these variables. Furthermore, the teachers in the longitudinal sample appear quite similar to the teachers in the full sample on these descriptive variables. Of course, it cannot be proven that a nonprobability sample is representative of a larger population, but it is clear that the samples used in these analyses are from multiple states, grades, and years and that their characteristics do not depart dramatically from national figures.

Analysis

To address the research question, the two alignment indices are calculated for each teacher, and these alignment indices are regressed on year, controlling only for sample composition with fixed effects for state and grade. The coefficient on the year variable represents the average one-year change in alignment within states and grades, averaged across states and grades:

$$\text{Align}_{ijk} = \beta_0 + \beta_1 \times \text{Year} + u_j + f_k + \varepsilon_{ijk}.$$

Here, Year is the year of SEC administration, u_j is a set of state fixed effects, and f_k is a set of grade fixed effects. Both standardized and unstandardized results are discussed. Standardized results are appropriate for the same reason that standardized effects are used in studies of the effects of NCLB on achievement. Furthermore, given the research showing correlations of 0.45 between content alignment and achievement in high school mathematics (Gamoran et al. 1997), it could be possible to extrapolate alignment increases to student achievement gains, assuming that the relationships were the same across subjects and grades, by simply multiplying the standardized regression coefficient by 0.45. However, unstandardized results are also presented to highlight the small absolute magnitude of the identified effects.

The coefficient β_1 indicates the relationship between alignment and year. Because of the large number of regressions run and presented, only the coefficients and standard errors on the year variable are presented, though all regression output (e.g., r-squared, f-statistics) is available from the author. To aid in presentation and interpretation, the indices (and, hence, coefficients) were multiplied by 100. Thus, a coefficient of 1.0 indicates a year-to-year increase in alignment of 1 percent. Within each subject, 10 regressions were run: grades K–2 standards, grades 3–8 assessments and standards, and grades 9–12 assessments and standards for each index.

An additional concern about the sample is that its composition changes over years in ways that may bias the identified effects. Thus, it is critical to investigate the internal validity of the findings to changes in model specification and sample composition. The strategy is to run multiple, parallel sets of models and investigate the extent to which the findings change. The most conservative is a set of models using fixed effects for teachers and only those teachers who completed surveys across multiple years; these models capture only within-teacher changes in alignment rather than within- and between-teacher changes. Unfortunately, these models also require an additional assumption not needed for the cross-sectional models: that teachers' understandings of the content in the SEC do not change as they complete the survey multiple times. I ran six sets of models; to the extent that the results of the models agree with one another, it suggests that the estimated changes in alignment are robust to model specification and more likely to represent true changes in instructional alignment among sample teachers during the period 2003–9.

Results

Means and standard deviations (SDs) of each of the alignment indices are presented in table 2. In all three subjects, alignment is low on index 1, with no averages above 0.28. Alignment is slightly higher for ELA assessments than for standards, but otherwise alignment on index 1 is no higher to standards than to assessments. Thus, for standards and assessments in all subjects, the average alignment (agreement on topic, cognitive demand, and proportion of instructional time) is less than 28 percent. A comparison of index 1 and index 2 reveals differences of 0.10–0.15, indicating that 10–15 percent of teachers' instructional time is spent teaching content covered in the standards and assessments in excess of the proportion of instructional time called for by the standards and assessments, with higher proportions of time spent this way for ELA standards. The distributions of both alignment indices are roughly normal; an illustrative histogram for alignment index 1 to science assessments is presented in figure 1. The histogram illustrates that the mean, median, and modal values are near 0.2 and that there are no apparent outliers.

The results of the main fixed-effects models are presented in tables 3, 4, and 5 for mathematics, ELA, and science. The coefficients are unstandardized in the tables, though both standardized and unstandardized results are discussed in the text. The results for mathematics indicate significant increases in alignment to standards in grades K–2 and 3–8 and to assessments in grades 3–8 and 9–12, with marginal significance for the increase in alignment to grades 9–12 standards. The significant changes appear in both indices but are generally larger for alignment index 2 than for alignment index 1. Perhaps

TABLE 2

Means and Standard Deviations of Alignment Indices

	ALIGNMENT INDEX 1		ALIGNMENT INDEX 2	
GRADES	Mean	SD	Mean	SD
Mathematics standards:				
K–2	.24	.07	.37	.16
3–8	.23	.10	.35	.19
9–12	.18	.09	.29	.16
Mathematics assessments:				
3–8	.21	.06	.28	.09
9–12	.19	.06	.28	.12
ELA standards:				
K–2	.27	.07	.41	.13
3–8	.28	.10	.45	.17
9–12	.27	.07	.42	.11
ELA assessments:				
3–8	.20	.07	.29	.11
9–12	.23	.07	.34	.15
Science standards:				
K–2	.14	.07	.25	.15
3–8	.17	.06	.27	.12
9–12	.18	.08	.26	.15
Science assessments:				
3–8	.16	.05	.22	.09
9–12	.20	.09	.32	.15

surprisingly, the increases in alignment for grades K–2 standards and 9–12 assessments are as large as or larger than those for grades 3–8 standards and assessments.

The magnitudes of the statistically significant regression coefficients on year of survey administration are 0.35–0.63 for alignment index 1 (recall that the coefficients were multiplied by 100, so they are 0.0035–0.0063 in the metric of the alignment index). For the four significant coefficients, these represent standardized effect sizes of 0.05–0.09 per year. Given these coefficients, over the six-year study period, the proportion of teachers' instruction aligned to standards and assessments increased 2.1–3.8 percent, standardized increases of 0.31–0.54 SD. For index 2, the unstandardized coefficients are larger (0.50–1.74), but the standardized coefficients range from 0.03 to 0.11. The estimated six-year increases in alignment index 2 are 0.48–0.66 SD for grades K–2 standards, 3–8 assessments, and 9–12 assessments and 0.19–0.20 SD for grades 3–8 and 9–12 standards. Overall, alignment significantly increased to standards and assessments at all grades except perhaps grades 9–12 standards,

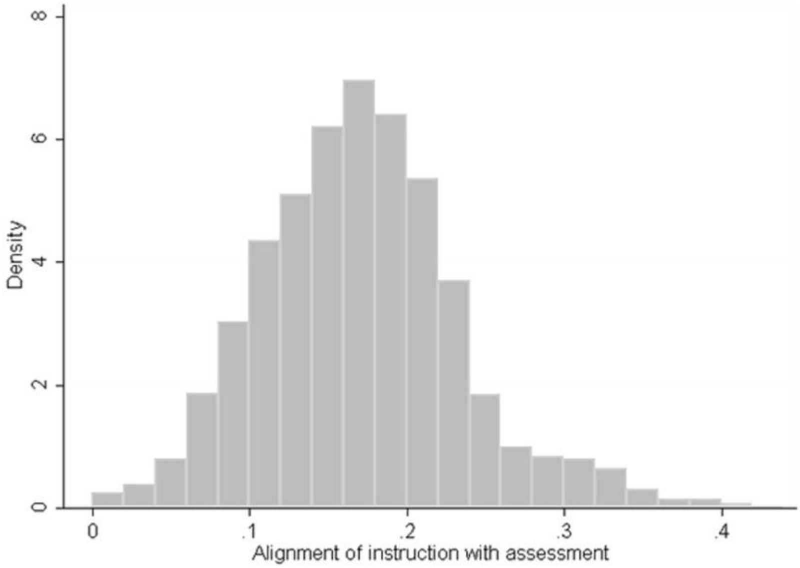

FIG. 1.—Alignment of instruction with assessment

TABLE 3

Fixed-Effects Regressions of Alignment to Mathematics Standards and Assessments

GRADES	ALIGNMENT INDEX 1 B	ALIGNMENT INDEX 1 SE(B)	ALIGNMENT INDEX 2 B	ALIGNMENT INDEX 2 SE(B)
Mathematics standards:				
K–2	.63***	.15	1.74***	.30
3–8	.51***	.05	.64***	.07
9–12	.18	.15	.50*	.25
Mathematics assessments:				
3–8	.45***	.06	.72***	.09
9–12	.35*	.17	1.11***	.29

NOTE.—Values are regression coefficients for fixed-effects regression of alignment indices on year of administration.

$^{+}p < .10$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 4

Fixed-Effects Regressions of Alignment to ELA Standards and Assessments

	ALIGNMENT INDEX 1		ALIGNMENT INDEX 2	
GRADES	B	SE(B)	B	SE(B)
ELA standards:				
K–2	.30***	.06	.57***	.08
3–8	.38***	.10	.85***	.11
9–12	.55	.34	−.42	.46
ELA assessments:				
3–8	.18*	.08	.34**	.13
9–12	−.07	.41	−.77	.57

NOTE.—Values are regression coefficients for fixed-effects regression of alignment indices on year of administration.
$^{+}p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 5

Fixed-Effects Regressions of Alignment to Science Standards and Assessments

	ALIGNMENT INDEX 1		ALIGNMENT INDEX 2	
GRADES	B	SE(B)	B	SE(B)
Science standards:				
K–2	−.77**	.27	$-.90^{+}$	.48
3–8	.21**	.07	.62***	.11
9–12	−.14	.17	−.03	.30
Science assessments:				
3–8	−.37***	.08	−.19	.15
9–12	−.71**	.22	−1.18**	.41

NOTE.—Values are regression coefficients for fixed-effects regression of alignment indices on year of administration.
$^{+}p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

with six-year increases ranging from 0.19 to 0.66 SD, depending on the particular grade span and measure of alignment.

The results in table 4 for ELA indicate similar patterns. As in mathematics, there are significant increases in alignment to grades K–2 standards and 3–8 standards and assessments. However, the results indicate no significant change in alignment to grades 9–12 standards or assessments. The magnitudes of the significant coefficients indicate year-on-year changes of 0.2–0.9 percent, depending on the particular dependent variable and grade. The largest of the significant coefficients is the coefficient on index 2 for grades 3–8 standards, which is 0.05 SD. This translates to a five-year effect size of approximately 0.25 SD. The five-year effect sizes for the other significant coefficients are 0.13–0.22 SD, indicating that, while there have been increases in instructional alignment in ELA, the magnitudes of the increases are approximately one-half as large as in mathematics.

Finally, the results in table 5 illustrate that the consistent increases in alignment in mathematics and ELA do not hold in science except for grades 3–8 standards. In science, there is a significant increase in alignment to grades 3–8 standards but significant decreases in alignment to grades K–2 standards, 3–8 assessments, and 9–12 assessments. In contrast, there is no change in alignment to grades 9–12 standards. The magnitude of the increase in alignment to grades 3–8 standards is between 0.21 percent and 0.62 percent per year, or between 0.03 and 0.05 SD. This corresponds to a six-year increase of 0.20–0.30 SD, a smaller increase than was found in mathematics but equivalent to that found in ELA. The significant decreases in alignment range in magnitude from 0.06 to 0.11 SD for six-year decreases of 0.36–0.66 SD. Overall, the results in science indicate small increases in alignment to standards at grades 3–8 and moderate decreases in alignment to standards and assessments at other grades.

Sensitivity Analyses

The results of six alternative specifications are presented in table 6. All results are presented using alignment index 1 because of space, though results for the other index are similar and are available from the author. Each model tests a different concern about the internal validity of the main results.

The most important concern is that the identified effects are due to changes in sample composition over time rather than actual changes in alignment. To test this possibility, three approaches are used. Columns 1 provide results for a set of regressions that include fixed effects for the state-grade interaction (e.g., fourth-grade mathematics in Michigan) rather than the separate state and grade fixed effects used in the main models. Because the instructional

TABLE 6

Comparison of Fixed-Effects Regression Models for Alignment Index 1

GRADES	Fixed Effects for State × Grade (1)			Additional Teacher Controls (2)			Longitudinal (3)			Strictly Cross-Sectional (4)			Raw Data, 2003–6 (5)			Raw Data, 2007–9 (6)		
	B	SE(B)	*n*	B	SE(B)	*n*	B	SE(B)	*n*	B	SE(B)	*n*	B	SE(B)	*n*	B	SE(B)	*n*
Mathematics standards:																		
K–2	.72***	.15	1,339	.50**	.16	1,339	1.68***	.46	261	.62***	.17	1,078	1.83***	.41	345	−.22	.28	994
3–8	.55***	.05	7,935	.45***	.05	7,934	.41**	.12	2,589	.53***	.07	5,346	.47***	.12	3,756	.44**	.14	4,179
9–12	.20	.15	1,339	.08	.15	1,337	.28	.35	360	−.05	.19	979	.37	.30	401	−.06	.33	938
Mathematics assessments:																		
3–8	.53***	.06	4,239	.40***	.06	4,238	.10	.17	1,252	.44***	.07	2,987	.38**	.12	2,311	.16	.17	1,928
9–12	.30+	.17	664	.34+	.17	664	.16	.43	233	.11	.20	431	.29	.44	256	.38	.38	408
ELA standards:																		
K–2	.30***	.06	5,750	.23***	.06	5,750	.03	.08	2,692	.40***	.10	3,058	.60***	.11	2,865	.18	.16	2,885
3–8	.40***	.10	5,618	.27**	.10	5,618	.46**	.17	1,229	.37**	.14	4,389	.02	.22	1,924	.06	.22	3,694
9–12	.61+	.33	439	.60+	.35	439	.74	.83	43	.62	.41	396	−.19	1.22	172	.66	.70	267
ELA assessments:																		
3–8	.20**	.06	4,185	.20*	.08	4,185	−.16	.16	1,013	−.10	.10	3,172	.34+	.20	1,787	−.78***	.17	2,398
9–12	−.07	.41	248	.37	.49	248	. . .	. . .	. . .	.01	.45	233	−.95	1.65	143	−1.91*	.84	105
Science standards:																		
K–2	−.79**	.27	353	−.63*	.27	353	−.71	.63	219	−.99*	.42	134	−.95	1.03	117	−.21	.57	236
3–8	.30***	.06	4,277	.19**	.07	4,140	.27*	.12	2,301	.32**	.09	1,891	.21	.16	2,039	.47**	.18	2,238
9–12	−.17	.17	693	−.23	.20	693	.15	.42	220	−.22	.24	473	.49	.33	285	−.37	.50	408
Science assessments:																		
3–8	−.34***	.08	1,480	−.38***	.08	1,480	−.28+	.16	769	−.43***	.12	711	.44*	.19	758	.13	.23	722
9–12	−.70**	.22	406	−.65**	.22	406	−1.01	.69	117	−.66*	.27	289	.10	.41	217	.25	.52	189

NOTE.—Values are regression coefficients for fixed-effects regression of alignment indices on year of administration.
+ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

target is a state- and grade-specific set of standards or assessments, these models are based on the possibility that there are features of certain state-grade combinations that are related to instructional alignment and that the distribution of teachers across these state-grade combinations differs across time in the sample. The samples for this analysis are smaller than the sample for the main analysis because a few observations were dropped with the more extensive fixed effects. Comparison of the results of these models to the main results reveals no major differences. Again, teachers increased the alignment of their instruction with mathematics and ELA standards and assessments, especially at the early grades.

Another possible explanation for the results of the main models is that the sample composition changes over time within states and grades on observable features related to alignment. For instance, it could be the case that the classrooms represented in later years are larger, more ethnically diverse, or lower achieving than the classrooms represented in earlier years, perhaps because the studies utilizing the SEC in 2008/9 were focused on different kinds of classrooms than the earlier studies. If these characteristics were related to alignment, the coefficients in the fixed-effects models would be biased and would reflect changes in sample composition across time. To address this concern, a set of control variables available on the SEC survey were included in the fixed-effects models. These controls are class size, percentage of nonwhite students, length of the target class, average achievement level of the class, and percentage of English language learners. Each variable was answered on a survey scale, and the scales were turned into sets of dummy variables and included in the main fixed-effects models. The results of these models, shown in columns 2 in table 6, indicate that additional classroom controls do not explain away the changes in alignment across time identified in the main models.

A third possibility is that the identified changes in alignment across time are merely indicative of some other changes in sample composition not reflected in either of the two sensitivity analyses presented above. For instance, perhaps teachers who appear in the sample in later years differ from those appearing in the sample in earlier years on some unobservable time-invariant characteristics that are related to instructional alignment. Since the previous models do not control for these kinds of unobservable characteristics, the coefficients on the year variable in these models might still reflect changes in sample composition.

The third set of models in table 6 (cols. 3) tests this possibility using a restricted teacher sample that includes only teachers who completed the SEC in multiple years. Fixed effects for teachers (i.e., dummy variables for each teacher in the sample) are used, meaning that the estimated coefficients on the year variable indicate the average within-teacher change in alignment per year, controlling for all observable and unobservable time-invariant teacher characteristics. The standard errors are larger for these models because of the smaller samples and highly restrictive nature of the fixed effects. While the results of these within-teacher models are not identical to those of the main models, there is a good deal of agreement. For instance, the within-teacher models identify significant increases in alignment to grades 3–8 standards in all three subjects and grades K–2 standards in mathematics, as did the main models. The within-teacher models also identify several increases in alignment that are large but not significant, such as grades 9–12 mathematics and ELA standards. Certainly, the results of these within-teacher models support the conclusion that alignment to standards across subjects is increasing within teachers.

Columns 4 use the main models but only the data for teachers who appear in the sample once. Thus, this strictly cross-sectional sample is mutually exclusive from the sample used in the within-teacher analysis just described. To the extent that the identified changes in alignment in these two sets of analyses are similar, it provides evidence that changes in alignment are similar within and between teachers. The results in these columns strongly support this conclusion. Again, there are significant increases in alignment to grades 3–8 standards in all subjects, as well as grades K–2 standards in mathematics. The magnitudes of the significant estimated changes in alignment are quite similar between the strictly cross-sectional and within-teacher models as well, lending further support to the conclusion that the identified changes in instructional alignment are not merely artifacts of changes in sample composition across time. The similarity of the results between the longitudinal and cross-sectional analyses also suggests that the identified effects are not reflecting differences in teacher interpretation of the SEC surveys based on repeated surveying. However, it is possible that all teachers may change their views as to the nature of their content coverage over time in ways that artificially inflate alignment increases; this hypothesis is not testable with these data.

A final internal validity concern is that the identified changes in instructional alignment across time are merely reflective of the previously described changes in the SEC survey that took place between 2006 and 2007. In all analyses to this point, the SEC data were first recentered by deleting all topics that did not appear both before and after the survey change. But it is possible that the changes in alignment identified in the main model occurred mainly between 2006 and 2007 because of changes in the SEC survey in that summer. To address this possibility, columns 5 and 6 in table 6 investigate changes in alignment from 2003 to 2006 and 2007 to 2009 using the raw SEC data that were not recentered by deleting topics. While these results are not in perfect agreement with the other models, they are again largely supportive of the conclusion that alignment increased over time. The coefficients are positive both before and after the SEC change in the models for mathematics standards at grades 3–8 and assessments at grades 3–8 and 9–12, ELA standards at grades K–2 and 3–8, and science standards at grades 3–8, though only some of these are significant in one or both time periods. The two main areas of disagreement between these models and the main models pertain to science assessments at grades 3–8 and 9–12 and ELA standards at grades 3–8. In the former, the main models showed significant decreases in alignment whereas these pre- and post-SEC change models show significant increases in alignment before and after 2006–7. In the latter, the main models showed significant increases in alignment whereas these pre- and post-SEC change models show small positive and insignificant changes. These differences suggest that, in these subjects and grades, it is possible that some of the identified effects in the main model are attributable to the survey change.

One possibility that has been ignored to this point is that the changes in instructional alignment were nonlinear in nature. For instance, perhaps teachers rapidly increased alignment in the early years of NCLB as materials became more aligned but increased alignment slowly or not at all after that. Or, perhaps teachers increased their alignment at an increasing rate over time as they became more familiar with the content of the standards. To investigate nonlinear changes in alignment, a number of nonlinear models were run, including quadratic, logarithmic, and square root functions of time. The results are not presented here, but no consistent nonlinear effects were identified across models, and the inclusion of nonlinear terms rarely improved model fit.

Across all of the sets of models and the main models, there is complete agreement that teachers significantly increased the alignment of their instruction to standards in mathematics at grades 3–8. There is near complete agreement (five or six of the seven models in agreement) on significant increases in alignment to mathematics standards at grades K–2, mathematics assessments at grades 3–8, ELA standards at grades K–2 and 3–8, and science standards at grades 3–8. In addition, there is near complete agreement about the direction of changes in alignment to mathematics assessments at grades 9–12 (increase), ELA standards at grades 9–12 (increase), and science standards at grades K–2 (decrease), though the coefficients are not always significantly different from zero. The significance levels vary somewhat across models, but these differences appear to be largely attributable to differences in sample sizes and degrees of freedom. Only in a few cases—most notably science assessments at all grades—does there appear to be substantial disagreement across the models.

Summary and Discussion

Data from the Surveys of Enacted Curriculum were used to investigate changes in instructional alignment during a large portion of the NCLB era (2003–9). Overall, the results suggest that changes in alignment have taken place, with the largest and most consistent increases in mathematics. Changes over the study period were as large as 0.65 SD in mathematics but closer to 0.1–0.2 SD in ELA. The finding that alignment has increased more in mathematics than in ELA fits with recent findings about the nature of changes in student achievement from Dee and Jacob (2011). In contrast, there have mainly been decreases in instructional alignment in science, with the exception of science standards in grades 3–8.

Despite the finding of increases in alignment, the magnitudes were mainly small to moderate when compared against the near unanimity with which teachers report efforts to improve their alignment across subjects and grades in surveys (Hamilton and Berends 2006; Pedulla et al. 2003; Stecher and Chun 2001). The results were supported across multiple types of models, including models with different types of fixed effects. Within-teacher models using teacher fixed effects produced results quite similar to those of strictly cross-sectional models using state and grade fixed effects, lending support to the use of these cross-sectional data for investigating instructional change over time.

There are important nuances to the main regression findings that merit further investigation. For instance, alignment seems to have increased more in mathematics than in ELA. It is impossible from these data to know the reason for this difference, but one plausible rationale is that mathematics standards and assessments may be more concrete and easily understandable by teachers than ELA standards and assessments. Also, shifts in alignment of instruction to standards and assessments were of roughly equal magnitudes, seemingly contradicting work that highlights extensive “teaching to the test” (Herman 2004). This finding could be attributable to the fact that only one form of each test was analyzed in this study; perhaps if multiple forms across years had been analyzed, the magnitudes of the coefficients on assessment alignment would be greater. Finally, alignment in science seems to have decreased at certain grades and for assessments. As suggested previously, there are several reasons to think that aligning instruction in science should be less of a focus for most teachers than aligning instruction in math and ELA. Unlike math and ELA, science is not used for accountability under NCLB, nor is it tested in every grade 3–8. However, it is not clear why these incentives would lead to a decrease in alignment to science standards or assessments. This result is supported across nearly all of the sensitivity checks, so it merits further investigation.

The primary limitations of this research are (*a*) the lack of a nationally representative probability sample and (*b*) the lack of survey data from before NCLB was implemented. While it is not possible to make a conclusive case that the results presented here are representative of changes in alignment nationwide over the years 2003–9, the large numbers of teachers from multiple states and grades lend support to the generalizability of the findings. To be sure, substantial proportions of the teacher samples for each analysis were from a few focal states, but the state or state-grade fixed effects account for unobservable time-invariant features of states that may be associated with changes in alignment (e.g., the format of a state's content standards), so the internal validity of the findings for these teachers should be high. Thus, the reader is left to determine the generalizability of the findings, taking into account the consistency of the findings across analyses and the characteristics of teachers in the sample.

As for the second limitation, it is possible that there was some shift in instruction after NCLB was passed into law but before the data set used here began. Given the findings presented here, the most logical hypothesis would be that alignment would have increased during that time, meaning that the estimates presented here would be understating the true effects of NCLB. Even without those data, however, there are five or six years of survey data that suggest that teachers have been aligning their instruction.

Given these limitations, it is also important to emphasize what this study was able to contribute to the literature. Previous survey studies of teachers' instruction relied on problematic teacher self-report of instructional change and assumptions about teachers' understanding of the content messages of standards and assessments. Though self-reported, the data used here are stronger in their validity in representing the content of instruction than previously used data in at least three key ways: these data (*a*) require no assumptions about teachers' understanding of alignment or the content messages contained in standards and assessments, (*b*) measure reported instruction at individual points in time rather than measuring reported change, and (*c*) measure instruction at a much finer level of detail than in any previous study on this issue. While not perfect, these data certainly provide the most comprehensive, best evidence yet on the changes in instructional alignment under NCLB.

The results also shed some light on results of earlier survey studies, which generally indicated that 80–90 percent or more of teachers across states, grades, and subjects reported increases in instructional alignment over time. The results presented here indicate that, even if these reports from teachers accurately represent the proportion of teachers who have improved their instructional alignment, these increases are small to moderate in magnitude and highly dependent on subject and grade. One possible explanation for this potentially discrepant finding is that previous studies may have represented substantially different teachers or times than those identified here (though even recent studies [e.g., Hamilton and Berends 2006] identify large proportions of teachers indicating alignment efforts). Or, it could be that teachers think of alignment in a different way than is measured here or believe that their instruction is aligned because their school or district tells them so (e.g., through "aligned" textbooks, pacing guides, or curriculum materials). However, it might also be the case that the earlier findings were not accurate and simply represent theory-driven reconstructions (Schwarz and Oyserman 2001).

While the results presented here indicate that teachers are responding somewhat to the content of standards and assessments and aligning their instruction, these results do not indicate whether or not instruction is improving in the sense of shifting to content and methods that promote deeper student understanding. To investigate such a question, measures of instructional quality would need to be taken across a wide range of classrooms and tracked over time, a truly daunting undertaking. Indeed, it is quite possible that alignment could increase but instructional quality could decrease, particularly if the instructional targets were of poor quality or if the underlying nature of the tasks did not change (Boston and Smith 2009). Sustained instructional improvement of the kind necessary to improve student opportunity to learn ambitious content certainly requires more than simply aligning instruction: it requires extensive support for teachers (Borko 2004; Cobb et al. 2003; Coburn 2003; Franke et al. 2001), powerful leadership (Goldring et al. 2009; Leithwood et al. 2004), and clear, transparent, and well-aligned goals (Clune 1993; Smith and O'Day 1991).

Of course, these results leave many important questions to be answered. One such question is the extent to which the effects identified here are localized in certain kinds of classrooms or are more uniform in nature. For instance, it might be supposed that teachers in schools that have failed adequate yearly progress would be more attentive to the standards and (especially) the assessments, in order to avoid sanctions. A second important research issue is the extent to which instructional alignment and changes in alignment are influenced by more proximal content messages, such as those contained in textbooks and district pacing guides. The current study can say nothing about whether teachers are working on their own to increase alignment or whether alignment is increasing because teachers are using better-aligned texts or curriculum materials. Finally, it would be useful to further investigate the relationship between aligned instruction and value added to student achievement, the other component of the theory of change underlying standards-based reform.

The results here indicate that at least part of NCLB's theory of change is holding true, to a moderate extent. Given the findings of the literature on achievement (Carnoy and Loeb 2004; Dee and Jacob 2011; Hanushek and

Raymond 2005; Jacob 2005), it appears that aligned instruction can lead to improved achievement. However, the magnitude of effects on instruction and achievement is certainly not as large as was hoped for when NCLB was authorized, perhaps owing to the surprisingly poor levels of coherence of state standards and assessments, the key instruments intended to drive teachers' instruction (Polikoff et al. 2011). Indeed, recent research suggests that instructional alignment tends to be higher when the standards and assessments are more aligned, as expected from the theory of action underlying SBR (Polikoff 2012). Thus, as policy makers think about strengthening SBR through the Common Core State Standards and state assessment consortia, it is imperative that greater attention be paid to closely aligning sources of instructional influence, including standards, assessments, textbooks, and curricula. SBR may well remain the main policy solution at the state and federal levels; if so, its current level of effectiveness will have to improve. Modest changes in teachers' instruction, while impressive in light of historical resistance to change, will not be enough to address the educational challenges we face.

Appendix

List of States Included in Each Teacher Sample

Mathematics Standards

Alabama, California, Delaware, Idaho, Illinois, Indiana, Kansas, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Montana, New Hampshire, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Texas, Vermont, and Wisconsin.

Mathematics Assessments

Idaho, Indiana, Kansas, Maine, Mississippi, Montana, New Hampshire, Ohio, Oklahoma, Oregon, West Virginia, and Wisconsin.

ELA Standards

Delaware, Idaho, Illinois, Indiana, Maine, Massachusetts, Michigan, Minnesota, Montana, New Hampshire, New York, North Carolina, Ohio, Oklahoma, Oregon, Utah, Vermont, Virginia, and Wisconsin.

ELA Assessments

Indiana, Maine, Minnesota, Montana, New York, North Carolina, Ohio, Oklahoma, Utah, and Wisconsin.

Science Standards

California, Delaware, Florida, Idaho, Illinois, Indiana, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, New Hampshire, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Texas, and Wisconsin.

Science Assessments

Florida, Illinois, Indiana, Iowa, Maine, Montana, New Hampshire, Ohio, Oklahoma, Oregon, Texas, and Wisconsin.

Note

Thank you to Andy Porter, Laura Desimone, Richard Ingersoll, and John Smithson, as well as several anonymous reviewers, for their feedback and support on earlier versions of this article.

References

Betts, Julian R., and Anne Danenberg. 2002. "School Accountability in California: An Early Evaluation." In *Brookings Papers on Educational Policy 2001*, ed. D. Ravitch. Washington, DC: Brookings Institution.

Borko, Hilda. 2004. "Professional Development and Teacher Learning: Mapping the Terrain." *Educational Researcher* 33 (8): 3–15.

Boston, Melissa, and Margaret S. Smith. 2009. "Transforming Secondary Mathematics Teaching: Increasing the Cognitive Demands of Instructional Tasks Used in Teachers' Classrooms." *Journal for Research in Mathematics Education* 40 (2): 119–56.

Carnoy, Martin, and Susanna Loeb. 2004. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." In *Redesigning Accountability Systems for Education*, ed. S. H. Fuhrman and R. F. Elmore. New York: Teachers College Press.

Center on Education Policy. 2006. "From the Capital to the Classroom: Year 4 of the No Child Left Behind Act." Report, Center on Education Policy, Washington, DC.

Clune, William H. 1993. "Systemic Educational Policy: A Conceptual Framework."

In *Designing Coherent Educational Policy*, ed. S. H. Fuhrman. San Francisco: Jossey-Bass.

Cobb, Paul, Kay McClain, Teruni de Silva Lamberg, and Chrystal Dean. 2003. "Situating Teachers' Instructional Practices in the Institutional Setting of the School and District." *Educational Researcher* 32 (6): 13–24.

Coburn, Cynthia E. 2003. "Rethinking Scale: Moving Beyond Numbers to Deep and Lasting Change." *Educational Researcher* 32 (6): 3–12.

Cohen, David K. 1990. "A Revolution in One Classroom: The Case of Mrs. Oublier." *Educational Evaluation and Policy Analysis* 12 (3): 311–29.

Cueto, Santiago, Cecilia Ramirez, and Juan Leon. 2006. "Opportunities to Learn and Achievement in Mathematics in a Sample of Sixth Grade Students in Lima, Peru." *Educational Studies in Mathematics* 62 (1): 25–55.

Dee, Thomas, and Brian A Jacob. 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management* 30 (3): 418–46.

Desimone, Laura M., Kerstin C. Le Floch, Susie Ansell, and James Taylor. 2005. "Are We Asking the Right Questions on National Surveys about Teacher's Professional Development? Insights from Cognitive Interviews with Teachers." Paper presented at the annual meeting of the American Educational Research Association, Montreal, April 11–15.

Franke, Megan L., Thomas P. Carpenter, Linda Levi, and Elizabeth Fennema. 2001. "Capturing Teachers' Generative Change: A Follow-up Study of Teachers' Professional Development in Mathematics." *American Educational Research Journal* 38 (3): 653–89.

Gamoran, Adam, Andrew C. Porter, John Smithson, and Paula A. White. 1997. "Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth." *Educational Evaluation and Policy Analysis* 19 (4): 325–38.

Goldring, Ellen, Andrew C. Porter, J. Murphy, S. N. Elliott, and Xiu C. Cravens. 2009. "Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices." *Leadership and Policy in Schools* 8 (1): 1–36.

Grissmer, David, and Ann Flanagan. 2001. "Searching for Indirect Evidence of the Effects of Statewide Reforms." In *Brookings Papers on Educational Policy 2001*, ed. D. Ravitch. Washington, DC: Brookings Institution.

Hamilton, Laura S., and Mark Berends. 2006. "Instructional Practices Related to Standards and Assessments." Working Paper no. WR-374-EDU, RAND Corp., Santa Monica, CA.

Hanushek, Eric A., and Margaret E. Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24 (2): 297–327.

Herman, Joan L. 2004. "The Effects of Testing on Instruction." In *Redesigning Accountability Systems for Education*, ed. S. H. Fuhrman and R. F. Elmore. New York: Teachers College Press.

Hill, Heather C. 2001. "Policy Is Not Enough: Language and the Interpretation of State Standards." *American Educational Research Journal* 38 (2): 289–318.

Jacob, Brian A. 2005. "Accountability, Incentives, and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89:761–96.

Koretz, Daniel M., Sheila I. Barron, Karen J. Mitchell, and Brian M. Stecher. 1996. "Perceived Effects of the Kentucky Instructional Results Information System." Report, RAND Corp., Santa Monica, CA.

Koretz, Daniel M., Karen J. Mitchell, Sheila I. Barron, and Sarah Keith. 1996. "Final

Report: Perceived Effects of the Maryland School Performance Assessment Program." CSE Technical Report no. 409, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Koretz, Daniel M., Brian M. Stecher, Stephen Klein, and Daniel F. McCaffrey. 1994. "The Vermont Portfolio Assessment Program: Findings and Implications." *Educational Measurement: Issues and Practice* 13 (3): 5–16.

Lam, Tony C. M., and Priscilla Bengo. 2003. "A Comparison of Three Retrospective Self-Reporting Methods of Measuring Change in Instructional Practice." *American Journal of Evaluation* 24 (1): 65–80.

Leithwood, K., K. S. Louis, S. Anderson, and K. Wahlstrom. 2004. *How Leadership Influences Student Learning*. Minneapolis: University of Minnesota.

Luna, Catherine, and Cara L. Turner. 2001. "The Impact of the MCAS: Teachers Talk about High-Stakes Testing." *English Journal* 91 (1): 79–87.

Mayer, Daniel P. 1999. "Measuring Instructional Practice: Can Policymakers Trust Survey Data?" *Educational Evaluation and Policy Analysis* 21 (1): 29–45.

McDonnell, Lorraine M. 2004. *Politics, Persuasion, and Educational Testing*. Cambridge, MA: Harvard University Press.

Pedulla, Joseph J., Lisa M. Abrams, George F. Madaus, Michael K. Russell, Miguel A. Ramos, and Jing Miao. 2003. "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers." Chestnut Hill, MA: National Board on Educational Testing and Public Policy.

Polikoff, Morgan S. 2012. "The Association of State Policy Attributes with Teachers' Instructional Alignment." *Educational Evaluation and Policy Analysis*, forthcoming.

Polikoff, Morgan S., Andrew C. Porter, and John Smithson. 2011. "How Well Aligned Are State Assessments of Student Achievement with State Content Standards?" *American Educational Research Journal* 48 (4): 965–95.

Porter, Andrew C. 2002. "Measuring the Content of Instruction: Uses in Research and Practice." *Educational Researcher* 31 (7): 3–14.

Porter, Andrew C., Rolf Blank, John Smithson, and Eric J. Osthoff. 2005. "Place-Based Randomized Trials to Test the Effects on Instructional Practices of a Mathematics/Science Professional Development Program for Teachers." *Annals of the American Academy of Political and Social Science* 599:147–75.

Porter, Andrew C., Robert Floden, Donald Freeman, William Schmidt, and John Schwille. 1988. "Content Determinants in Elementary School Mathematics." In *Perspectives on Research on Effective Mathematics Teaching*, ed. D. A. Grouws and T. J. Cooney. Hillsdale, NJ: Erlbaum.

Porter, Andrew C., Michael W. Kirst, Eric J. Osthoff, John Smithson, and Susan A. Schneider. 1993. "Reform Up Close: An Analysis of High School Mathematics and Science Classrooms." Report, Wisconsin Center for Education Research, Madison.

Porter, Andrew C., Morgan S. Polikoff, Timothy Zeidner, and John Smithson. 2008. "The Quality of Content Analyses of State Student Achievement Tests and State Content Standards." *Educational Measurement: Issues and Practice* 27 (4): 2–14.

Porter, Andrew C., John Smithson, Rolf Blank, and Timothy Zeidner. 2007. "Alignment as a Teacher Variable." *Applied Measurement in Education* 20 (1): 27–51.

Ross, Michael. 1989. "Relation of Implicit Theories to the Construction of Personal Histories." *Psychological Review* 96:341–57.

Ross, Michael, and Michael Conway. 1986. "Remembering One's Own Past: The Construction of Personal Histories." In *Handbook of Motivation and Cognition*, ed. R. M. Sorrentino and E. T. Higgins. New York: Guilford.

Schwarz, Norbert, and Daphna Oyserman. 2001. "Asking Questions about Behavior:

Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation* 22 (2): 127–60.

Sebring, Penny A. 1987. "Consequences of Differential Amounts of High School Coursework: Will the New Graduation Requirements Help?" *Educational Evaluation and Policy Analysis* 9 (3): 257–73.

Shepard, Lorrie A., and Katharine C. Dougherty. 1991. "Effects of High-Stakes Testing on Instruction." Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 3–7.

Smith, Marshall S., and Jennifer A. O'Day. 1991. "Systemic School Reform." In *The Politics of Curriculum and Testing: Politics of Education Association Yearbook*, ed. S. H. Fuhrman and B. Malen. Bristol, PA: Falmer.

Smith, Mary L. 1991. "Put to the Test: The Effects of External Testing on Teachers." *Educational Researcher* 20 (2): 8–11.

Smith, Mary L. 1997. "Reforming Schools by Reforming Assessment: Consequences of the Arizona Student Assessment Program (ASAP): Equity and Teacher Capacity-Building." CSE Technical Report no. 425, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Smithson, John, and Ana C. Collares. 2007. "Alignment as a Predictor of Student Achievement Gains." Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 3–7.

Snyder, Thomas D., Sally A. Dillow, and Charlene M. Hoffman. 2008. *Digest of Education Statistics, 2007*. Washington, DC: National Center for Education Statistics.

Stecher, Brian M., and Sheila I. Barron. 2001. "Unintended Consequences of Test-Based Accountability When Testing in 'Milepost' Grades." *Educational Assessment* 7 (4): 259–81.

Stecher, Brian M., and Tammi Chun. 2001. "School and Classroom Practices during Two Years of Education Reform in Washington State." CSE Technical Report no. 550, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Taylor, Grace, Lorrie A. Shepard, Freya Kinner, and Justin Rosenthal. 2002. "A Survey of Teachers' Perspectives on High-Stakes Testing in Colorado: What Gets Taught, What Gets Lost." CSE Technical Report no. 588, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Wong, Kenneth K., Dorothea Anagnostopoulos, Stacey Rutledge, and Claudia Edwards. 2003. "The Challenge of Improving Instruction in Urban High Schools: Case Studies of the Implementation of the Chicago Academic Standards." *Peabody Journal of Education* 78 (3): 39–87.