

Educational Evaluation and Policy Analysis

<http://eepa.aera.net>

The Association of State Policy Attributes With Teachers' Instructional Alignment

Morgan S. Polikoff

EDUCATIONAL EVALUATION AND POLICY ANALYSIS 2012 34: 278 originally published online 1

May 2012

DOI: 10.3102/0162373711431302

The online version of this article can be found at:

<http://eepa.sagepub.com/content/34/3/278>

Published on behalf of



[American Educational Research Association](http://www.aera.net)

and



<http://www.sagepublications.com>

Additional services and information for *Educational Evaluation and Policy Analysis* can be found at:

Email Alerts: <http://eepa.aera.net/alerts>

Subscriptions: <http://eepa.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

>> [Version of Record](#) - Aug 8, 2012

[OnlineFirst Version of Record](#) - May 1, 2012

[What is This?](#)

The Association of State Policy Attributes With Teachers' Instructional Alignment

Morgan S. Polikoff

University of Southern California

States vary greatly in their implementation of standards-based accountability under No Child Left Behind, yet little evidence is available to guide policymakers on what attributes of state policy advance more tightly aligned instruction. This study uses survey data and content analyses from the Surveys of Enacted Curriculum to describe elements of state policy that are associated with tighter alignment between teacher-reported instruction and state standards and assessments. The author reports substantial variation on policy attributes across states, with increasing use of power (rewards and sanctions) and low overall consistency between standards and assessments. Several of the policy attributes are related to instructional alignment, with the strongest fit in mathematics.

Keywords: *alignment, instruction, standards, assessments, No Child Left Behind*

STANDARDS-BASED REFORM (SBR) is based on a clear, well-defined theory of change. The model begins with the intended learning outcomes of schooling—the knowledge and skills expected of students. These outcomes are to be outlined in curriculum frameworks, which represent the instructional target for teachers, and supported by aligned assessments, curriculum materials, and professional development. The hypothesis is that the multiple supports, if well aligned, will lead teachers to modify their instruction to increase alignment. Achievement should then rise on the domain expressed by the standards, a worthy goal if the standards represent appropriate, challenging academic content (Clune, 1993; Smith & O'Day, 1991). Under SBR policies such as the No Child Left Behind (NCLB) Act of 2001, the content of instruction is therefore the key mediating variable

between the policy and the outcomes of improved achievement and narrowed gaps.

In practice, NCLB is not one but 50 SBR policies. States can implement SBR with widely varying policies along multiple dimensions, including the content of standards and assessments (Polikoff, Porter, & Smithson, 2011; Porter, Polikoff, & Smithson, 2009; Reys, 2006; Reys et al., 2007), the rigor of cut scores on teacher licensure examinations (Porter & Polikoff, 2007), and the annual targets for adequate yearly progress (Porter, Linn, & Trimble, 2005). The NCLB Act grants states additional control over such policies as the types of rewards and sanctions for schools and students, and the extent to which SBR policies remain stable over time. Together, the attributes of state policies around SBR are highly variable (Porter & Polikoff, 2007).

I thank John Smithson for providing the data; Andrew Porter, Laura Desimone, Richard Ingersoll, Katharine Strunk, and Andrew McEachin for their helpful comments on earlier versions of this article; and the editors and reviewers of EEPA.

As state policy increasingly affects the classroom, several features of state policies—specificity, stability, consistency, power, and authority—have been shown to relate to the implementation of standards-based policies and, ultimately, student learning (Clune, 2001; Desimone, 2002; Desimone, Smith, Hayes, & Frisvold, 2004). However, there is little research connecting teachers' instructional choices with the nature of state policies, leaving policymakers little useful evidence about the most effective policy attributes for achieving SBR's intended outcomes. Thus, the focus here is on the content of instruction and how state policy might be better constructed to promote aligned instruction. Given that instructional alignment is an explicit goal of standards-based reforms, regardless of the quality of the instructional targets, it is important to understand the extent to which state policies promote teachers' accomplishment of that goal. Thus, the specific research questions to be addressed are as follows:

1. To what extent do state policies differ on attributes that may shape levels of instructional alignment?
2. What is the relationship between these state policy attributes and the outcome of instructional alignment with standards and assessments?

Background

Conceptual Framework for the Study of Policy Relationships on Instruction

To investigate these questions, I use a five-part policy attributes framework (for more detail, see Desimone, 2002; Porter, Floden, Freeman, Schmidt, & Schwille, 1988). *Consistency* is the extent to which the parts of the policy are coherent with one another and provide the same message to the implementer (e.g., Cohen & Spillane, 1992; Darling-Hammond, 1990; Porter et al., 1988; Spillane, 2004). Of all the attributes, consistency is the most fundamental to SBR. The consistency of policy messages is seen as being critical in K–12 education's fragmented governance system (Smith & O'Day, 1991). In this study, the consistency of state policy is operationalized as the *consistency* (alignment) of state assessments with content standards.

Specificity is the degree to which the policy is clear and promotes common understanding of the policy's intent (e.g., Cohen & Ball, 1990; Porter et al., 1988; Ravitch, 1995). The framework suggests that more specific standards are less open to misunderstanding and more likely to be implemented as intended. More specific standards should also be more validly and reliably measured and, thus, should be better able to contribute to an aligned standards-based system. In this study, specificity is operationalized by the degree of *focus* of the state standards—do they cover a few topics in great depth or many topics shallowly?

Stability is the extent to which the policy has been in place for a long time (e.g., Porter et al., 1988; Raizen, 1988; Spillane, 2004). Policies that are more stable give implementers a better opportunity to learn about, adapt to, and implement them, increasing the likelihood of achieving intended effects. Unstable policies diminish the utility of the information provided by SBR—if teachers receive feedback on student performance against some set of standards and those standards are rapidly changing, that information will be of limited value. For this study, *stability* is measured by the number of years the state standards have been in place.

Power, often called “rewards and sanctions,” is the use of resources, controlled by the superior, to motivate the subordinate (Spady & Mitchell, 1979). Power-based policies are intended to provide external motivation (Cohen & Spillane, 1992; Porter et al., 1988; Schwille et al., 1988). Power plays an important role in NCLB—schools may face reconstitution and closure if they do not meet student achievement targets. In this case, the threat of sanctions is supposed to incentivize schools to improve (Supovitz, 2009). Power can also be exerted on teachers indirectly, through the establishment of requirements for student promotion or graduation. Here, *power* is measured by the extent to which the state has policies that provide rewards or sanctions to schools, teachers, or students on the basis of certain measures of performance.

Finally, *authority* is based on teachers' belief in the legitimacy of the policy and is classified in four types: traditional, based in history, customs, and values; charismatic, based on the affective relationship between subordinate and

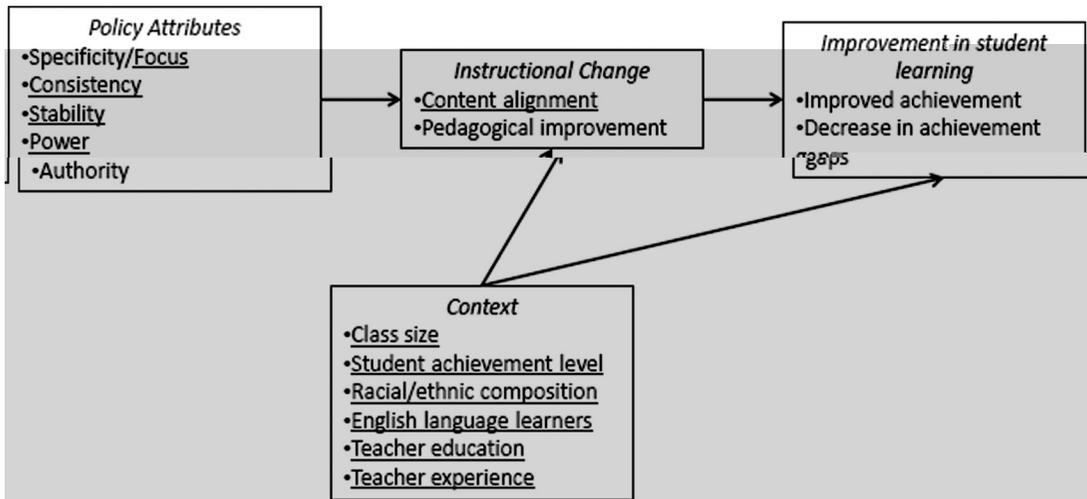


FIGURE 1. Conceptual framework for the study of standards-based reform.

superior; legal, based on rules and laws; and expert, based on possession of specialized knowledge or skills. Authority is widely seen as being important in the implementation of standards-based reform policies, as it can affect the intrinsic motivation of teachers and other educational actors (Cohen & Spillane, 1992; Porter et al., 1988; Spillane, 2004). Unfortunately, no high quality measures of authority were able to be devised for this analysis. For one, an intrinsic concept such as authority is perhaps best measured at the source of the motivation (Desimone, 2002)—in this case, the teacher him or herself. For another, the models used here required variation in attributes within states, and no measures could be derived that fit this requirement.

Figure 1 shows a model of the conceptual framework for the effect of standards-based reform on instruction and student learning. The underlined variables are measured and included in the statistical models. On the left are the features of the policy attributes framework. Based on the variation of these features, we would expect variation in the degree of implementation of standards-based practices (e.g., adoption of content standards). To the extent that the instructional targets are high quality, these instructional changes would affect the quality of student learning, which is not measured here.

Prior Research on State Policy Associations With Instructional Alignment

Few studies have focused on the relationships between specific state policies or policy attributes and instructional alignment using multi-state data. Existing studies are primarily based in individual states. In terms of content effects, consistent findings are that instruction has narrowed to focus on tested subjects (e.g., Hamilton & Berends, 2006; Pedulla et al., 2003; Stecher, Barron, Chun, & Ross, 2000) and that teachers report increased instructional alignment with standards and assessments (e.g., Hamilton & Berends, 2006; Koretz, Mitchell, Barron, & Keith, 1996; Pedulla et al., 2003; Taylor, Shepard, Kinner, & Rosenthal, 2002). These state-specific studies use surveys or qualitative designs and generally find instructional effects of standards-based reforms. However, although there are differing policy contexts identified across these studies, the lack of common methodologies and reporting metrics does not allow for comparison of effects across studies.

One study has directly compared teachers using multi-state data on the basis of state policy effects on content (Pedulla et al., 2003). The national survey of 4,000 teachers first characterized each U.S. state in terms of the stakes for teachers/schools and the stakes for students associated with state tests. Given these classifications,

teachers were randomly sampled and surveyed about their opinions of standards-based reform and its effects on their instruction. Teachers in states with high stakes for students reported greater increases in coverage of tested subjects and greater decreases in coverage of untested subjects. Furthermore, teachers in states with high stakes for schools or teachers reported greater instructional alignment than teachers in states with moderate stakes for schools/teachers. Standardized effect sizes were not reported.

Overall, the results of this study were suggestive that the state use of power was related to the degree of content effects, with greater effects in high-stakes states. However, the study was limited by the focus on only one attribute of state policy (power) and by the sorts of questions asked. The survey relied on teacher reports of instructional change, which likely suffer from weak validity (Desimone, Le Floch, Ansell, & Taylor, 2005) and may contribute to theory-driven reconstructions where teachers report alignment efforts because alignment is an obvious intended policy outcome (Ross, 1989; Ross & Conway, 1986; Schwartz & Oyserman, 2001).

Few other studies evaluate the relationship between state policy attributes and the degree of content effects. McDonnell (2004) identified that a higher proportion of Kentucky than North Carolina teachers were able to produce assignments similar to their target assessments when asked to do so, hypothesizing that one explanation for the difference was that the stakes for schools attached to assessments were higher in Kentucky. Firestone, Mayrowetz, and Fairman (1998) similarly found differences in instruction across two states differing on level of stakes attached to assessment results, attributing those differences to the stakes. In contrast, Clarke and colleagues (2003) interviewed teachers spread across high-, medium-, and low-stakes states and found no clear relationship between level of stakes and perceived instructional change.

Some evidence suggests that power may be related to differences in the content of instruction. However, the evidence is dated, and the magnitude of such effects is unclear. Furthermore, power is apparently the only policy attribute to have been investigated for relationships with instruction across states. Given that increasingly prominent role of state policy, the lack of

investigation of policy effects in research on content alignment, the key mediating variable in the theory of change for standards-based reform, is an important oversight.

Method

Data

I estimate two indices of instructional alignment—one each describing the alignment of teachers' instruction with state content standards and assessments. These indices are calculated by comparing the reported content of teachers' instruction with content analyses of state standards and assessments. A total of 9,920 surveys of teachers' instructional content have been collected using the Surveys of Enacted Curriculum (SEC) content taxonomies (for more information on the SEC, see Porter, 2002; Porter et al., 1988). The teachers are from grades 2–12 in mathematics, English language arts (ELA), and science. The data have been collected over the years 2003–2009 for use in various research studies; there is no evidence that data collection procedures differed systematically across settings or years. Content analyses of standards and assessments also come from the SEC. Since 2002, trained content analysts have analyzed state documents in the three content areas, using the same content taxonomies as are used for teachers' instruction. No new instructional or content analysis data were collected for this article.

The SEC defines content at the intersection of specific topics (183 in mathematics, 133 in ELA, and 211 in science) and five levels of cognitive demand, which differ slightly by content area. The same framework is used for teacher surveys as for content analyses. Teachers complete the survey for a target class by thinking about a particular time period (often a semester) and reporting the extent to which they covered each topic on the list. Then, for each topic covered, the teacher reports the proportion of instructional time at each level of cognitive demand. The data are turned into proportions of time and aggregated to represent a full academic year. Investigations of data quality support the reliability and validity of teacher responses (Gamoran, Porter, Smithson, & White, 1997; Porter, Kirst,

Osthoff, Smithson, & Schneider, 1993; Porter, Smithson, Blank, & Zeidner, 2007). Especially positive were the 0.45 correlations of alignment indices with student achievement gains in high school mathematics (Gamoran et al., 1997). Teachers also complete a “Part A” survey on their classes and professional backgrounds; these data are also used here.

The SEC is also used to content analyze standards and assessments (Porter, Polikoff, Zeidner, & Smithson, 2008). The content analysis is carried out by trained subject matter experts who make independent judgments. Each coder analyzes test items or content standards at the finest grained level of detail and places the item/objective into one or more cells in the SEC taxonomy, resulting in a matrix for each rater indicating proportions of the document’s total content in each SEC cell. The content analyses are averaged across raters, with the result again being a matrix of proportions. Generalizability coefficients estimating the reliability of the coding procedures are generally 0.75 or greater for four or more raters (Porter, 2002; Porter et al., 2008).

Because the instructional data and content analysis data are in the same format, they can be used to calculate alignment of instruction with standards/assessments or of standards with assessments (Porter, 2002). The formula is

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

where x_i is the proportion of content in cell i of matrix x (e.g., a teacher’s reported instruction) and y_i is the proportion of content in cell i of matrix y (e.g., a state’s content standards). The index ranges from 0 (misalignment) to 1 (alignment). An illustrative histogram of alignment indices for the data used in this analysis appears in Figure 2. The histogram, for the alignment of ELA instruction with standards, is quite normal, with slightly thick tails. The distributions for the other outcome variables are similar.

Also, a database of state policies on power was assembled from data provided in *Education Week’s* annual “Quality Counts” series (*Education Week*, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009) and a report on high school exit exams (Center on Education Policy, 2009). Similar databases were used in previous state policy analyses

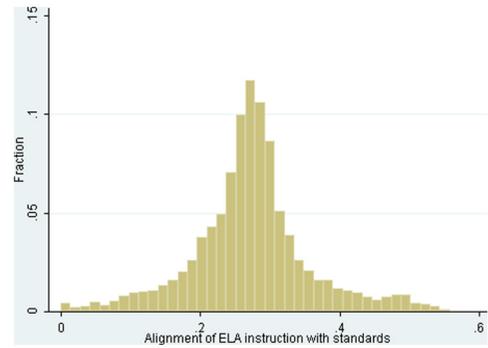


FIGURE 2. Distribution of alignment indices for alignment of instruction with standards in ELAR.

(e.g., Desimone et al., 2004; Desimone, Smith, & Phillips, 2007), and the *Education Week* data are seen as being of high quality and appropriate for use in constructing indicators of key constructs (Camilli & Firestone, 1999; Coe, 2003). The variables in the database contribute to a measure of the policy attribute of power.

Sample

The teacher data for this study have not been collected systematically and are not nationally representative. Table 1 summarizes the samples by academic subject and year of administration. Although there are more than 9,000 teachers in each subject available in the database, only a small portion of these teachers is in states and grades where both standards and assessments have been content analyzed. Thus, the teacher samples are approximately 4,000 for mathematics and ELA and 1,600 for science. In each subject, the years with the most data are 2004–2008. Most teachers are concentrated in the NCLB testing grades of 3–8, although there are 105 mathematics teachers from grades K–2 and approximately 600 mathematics and 250 each ELA and science teachers from grades 9–12. In terms of states, mathematics teachers are drawn primarily from Indiana, Montana, Ohio, Oklahoma, and Oregon; ELA teachers are drawn primarily from Indiana, New York, and Ohio; and science teachers are drawn primarily from Illinois, Oklahoma, and Wisconsin.

Because the samples are not probability samples, it is important to discuss the extent to

TABLE 1
Teacher Sample Sizes by Year

Year	Mathematics	ELA	Science
2002/3	0	0	22
2003/4	452	321	117
2004/5	668	483	338
2005/6	1,066	724	389
2006/7	1,016	1,139	296
2007/8	683	960	288
2008/9	536	280	183
Total	4,421	3,907	1,633

Note. Teachers are clustered in 8 states in mathematics, 11 states in English language arts (ELA), and 10 states in science.

which participating teachers are representative of U.S. teachers. A set of descriptive statistics on the teacher sample drawn from the SEC Part A is shown in Table 2. These are compared with national figures taken from the *Digest of Education Statistics* (Snyder, Dillow, & Hoffman, 2008). The limited set of descriptive variables presented was based on the available variables in the *Digest*. The descriptive statistics highlight that the samples are close to national averages, with few exceptions. The samples of mathematics classes and science classes both have 5% to 10% more White students than is typical of classes nationwide. In all three subjects, the class sizes in secondary classrooms are lower by two to four students than the national average. In contrast, the class sizes in elementary classrooms are very close to national averages in mathematics, slightly lower than average in ELA, and slightly higher than average in science. In ELA, the sample is more educated than is typical (i.e., fewer teachers with a top degree of bachelor's, more with a top degree of master's). Overall, the samples do not appear radically different from national figures on most indicators. Still, these and other descriptive variables are entered into the regression equations to control for sample composition.

Procedure

To address the first research question, the values of the policy attributes are described with descriptive statistics, tables, graphs, and correlations. For the second research question, intraclass correlations (ICCs) are first examined; to the extent that there is significant variation in

aligned instruction, a series of standardized fixed effects regressions is used to investigate the relationships between state policy attributes and the outcomes. Separate models are run for each subject, with standard errors clustered at the state level:

$$ALIGN_{isgt} = P_{sgr} \beta + T_{isgt} \gamma + \rho_s + \rho_t + \varepsilon_{isgt} \quad (2)$$

Here, $ALIGN_{isgt}$ is the instructional alignment with standards for teacher i in state s , grade g , and year t . The four policy attributes—focus, stability, power, and consistency—are captured in the vector P_{sgr} and measured at the state-grade-year level. This vector also contains a control for spurious correlation, discussed below. The state fixed effects (ρ_s) control for time- and grade-invariant observable and unobservable characteristics of states. The time fixed effects (ρ_t) control for secular trends over time across grades and states. The coefficients on the consistency and focus variables are primarily identified off variation across grades within states, as these variables rarely change across time within a particular state and grade (i.e., there are few state-grades where standards in the database change over time). In contrast, the coefficients on the power and stability variables are identified off variation across years within state-grades, as these variables have the same values for each grade within a particular state and year. A set of teacher controls is included in the vector T_{isgt} . These are (a) teachers' years of experience teaching the target class and dummy variables for (b) medium class (21–25 students), (c) large class (26+ students), (d) medium proportion non-White

TABLE 2

Comparison of Sample Descriptive Statistics With National Averages

Variable	Sample	Population
Mathematics		
Percentage non-White students	33.3	43.5
Elementary (K–5) class size	20.2	20.4
Secondary (9–12) class size	20.3	24.7
Percentage of teachers with highest degree bachelor's	50.4	50.8
<i>n</i>	4,421	
English language arts		
Percentage non-White students	44.5	43.5
Elementary (K–5) class size	18.5	20.4
Secondary (9–12) class size	22.1	24.7
Percentage of teachers with highest degree bachelor's	34.3	50.8
<i>n</i>	3,907	
Science		
Percentage non-White students	34.9	43.5
Elementary (K–5) class size	21.9	20.4
Secondary (9–12) class size	21.8	24.7
Percentage of teachers with highest degree bachelor's	53.8	50.8
<i>n</i>	1,633	

Note. Population data are from 2008 *Digest of Education Statistics*.

students (20–50%), (e) high proportion non-White students (> 50%), (f) low class average achievement level, (g) high class average achievement level, (h) teacher has a degree above a bachelor's, and (i) teacher has a degree in the subject area taught. Six models are run—one for alignment to standards and assessments in each content area.

The primary variables of interest are the four state policy attributes contained in vector P_{sgt} . The coefficient on the consistency variable indicates the relationship between the coherence of the policy system and the outcome of instructional alignment. However, there is the possibility that this variable (the alignment of standards S with assessments A) may be systematically and spuriously correlated with the dependent variable (the alignment of instruction I with standards S) because both are calculated by comparing something with S. To investigate the possibility of spurious correlation, the target assessment A is replaced with a randomly chosen assessment A* from a different state and grade. Although there should theoretically be a correlation between IS and SA alignment, there should not be a correlation between IS and SA* unless there is a spurious correlation caused by

the inclusion of S in both variables. When this test is performed repeatedly (i.e., 20 or more times), it is clear that there is indeed a spurious correlation of approximately two-thirds the magnitude of the original correlation. Thus, there is evidence of spuriousness, and it must be accounted for or the coefficient on the consistency variable will be biased.

To address this spurious relationship, the consistency of S with all A** (i.e., the consistency of the standards with every assessment in the database except the target assessment and the randomly chosen A*) is calculated. Taking the average of these consistency indices captures the extent to which the standards document is prone to high or low consistency; call this variable ALIGNABILITY (SA**-bar). If alignability is controlled, the correlation between SA* and IS goes to 0. However, the correlation between SA and IS remains large and positive. Thus, there is no longer any evidence of a spurious relationship after controlling for alignability. The coefficient on consistency therefore represents the relationship between standards-test alignment and instructional alignment with standards, controlling for the spurious correlation caused by alignability. The approach

is completely analogous for assessments. Thus, alignability is included in the vector of state-grade-year level policy variables in the model above.

Results

Descriptive Statistics for Policy Attributes

Focus. The index for focus measures the extent to which the state's standards in a particular grade/subject are focused on a few big ideas. It is a composite of (a) the number of cells in the SEC framework that contain .556% or more (1 day in a 180-day year) of the total content, (b) the number of cells in the SEC framework required to account for 50% of the year's total content, (c) the total number of SEC cells covered at all, and (d) the proportion of total content included in the 10 most prominent SEC cells (reverse coded). Each value is standardized within subject and the four values are averaged. Cronbach's alpha of the scale is .93 for mathematics and .84 for ELA and science.

Because the FOCUS variable has no meaningful scale, it is useful to examine what high- and low-focus standards look like. The least focused state standards in mathematics are Ohio's eighth-grade standards, and the most focused are Montana's third-grade standards. Montana's third-grade standards capture half of their content in just 3 SEC cells and cover 83% of the curriculum in the 10 most prominent cells. In total, only 33 SEC cells are included in these standards. In contrast, Ohio's eighth-grade standards capture half their content in 62 cells and contain just 13% of the curriculum in the 10 most prominent cells. These standards cover a total of 182 SEC cells, more than five times as many as Montana's third-grade standards.

A standards document at the mean for each component of the focus index in mathematics would cover 113 SEC cells and would capture 42% of the curriculum in the top 10 cells. In science, an average standards document would cover 119 cells and capture 44% of the curriculum in the top 10 cells. In ELA, the figures would be 177 cells and 30%. Clearly, ELA standards are less focused than math and science standards. Table 3 shows the average focus values by state and subject. States that have highly focused

standards in some subjects do not always have highly focused standards in all subjects. Of the states appearing in multiple subjects, only Oklahoma's standards are consistently above average on focus across the subjects.

Power. The power of a state's standards-based reform policies is operationalized by an index similar to that used in Carnoy and Loeb's (2002) analysis of the effects of accountability policy, combining state accountability for schools and students into a single index. Using indicators from *Education Week's* annual "Quality Counts" series and a report on high school exit exams, a 5-point scale for power was constructed. A state received a 1 if it provided additional rewards to high-achieving schools above and beyond those called for by NCLB. A score of 2 was awarded if the state used an additional indicator other than those required under NCLB for accountability, or if the state had an exit exam requirement for graduation. A score of 3 was awarded for states that had both additional indicators and an exit exam requirement. And a score of 4 was awarded for states that had additional indicators and an exit exam based on 10th-grade standards or higher. A score of 0 was therefore awarded to states with none of these policies. Thus, the low point is for states that follow NCLB but go no further. Subsequent points on the scale are for states that provide additional rewards, levy additional sanctions to schools or students, levy additional sanctions to schools and students, and levy sanctions to schools and students using high standards. Several alternate scales were also constructed and all results presented here were consistent across scales.

The index is calculated for each year. Of the 17 states in the teacher sample, none saw a decrease in power over the years 2003–2009, 9 stayed the same, and 8 saw an increase. Thus, for half the states in the sample, there is no variation in power over years. The coefficients on power reflect the associations of within-state changes in power, controlling for the secular trend of increasing power that is absorbed by the time fixed effects. The figures for each state are presented at the bottom of Table 3. In 2003, 7 of 17 states had power indices of zero and 3 had indices of four. By 2009, 2 states had an index of zero, and 5 had an index of four.

TABLE 3
Descriptive Statistics for State Policy Variables and Outcome Variables for States in the Sample

Measure	FL	IL	IN	ME	MN	MS	MT	NY	NH	NC	OH	OK	OR	UT	VA	WV	WI	Average	
Mathematics																			
IS alignment	—	—	0.25	—	—	0.20	0.07	—	—	—	0.34	0.18	0.27	—	—	0.16	0.15	0.21	
IA alignment	—	—	0.22	—	—	0.22	0.22	—	—	—	0.22	0.18	0.23	—	—	0.19	0.21	0.21	
Consistency	—	—	0.29	—	—	0.42	0.12	—	—	—	0.34	0.33	0.36	—	—	0.31	0.15	0.31	
Stability	—	—	5.85	—	—	4.93	7.68	—	—	—	3.22	2.97	3.89	—	—	3.98	6.99	4.10	
Focus	—	—	-0.06	—	—	0.35	1.40	—	—	—	-0.20	0.49	0.08	—	—	-0.32	0.10	0.24	
English language arts																			
IS alignment	—	—	0.25	0.10	0.35	—	0.14	0.36	—	0.22	0.28	0.18	—	0.25	0.18	—	0.20	0.29	
IA alignment	—	—	0.22	0.14	0.16	—	0.17	0.18	—	0.18	0.22	0.15	—	0.13	0.23	—	0.16	0.20	
Consistency	—	—	0.22	0.15	0.25	—	0.28	0.20	—	0.16	0.25	0.24	—	0.20	0.25	—	0.11	0.23	
Stability	—	—	0.52	8.00	4.57	—	6.95	1.53	—	4.00	3.34	3.88	—	4.00	1.00	—	6.90	2.89	
Focus	—	—	0.22	1.57	-0.80	—	1.04	-0.96	—	0.25	0.11	0.87	—	0.45	0.30	—	-0.21	-0.04	
Science																			
IS alignment	0.17	0.18	0.18	0.14	—	—	0.18	—	0.18	—	0.17	0.17	0.18	—	—	—	0.12	0.17	
IA alignment	0.14	0.18	0.20	0.11	—	—	0.19	—	0.11	—	0.19	0.19	0.15	—	—	—	0.15	0.18	
Consistency	0.25	0.31	0.28	0.27	—	—	0.25	—	0.22	—	0.17	0.31	0.30	—	—	—	0.15	0.28	
Stability	6.00	4.10	6.03	7.72	—	—	9.32	—	8.00	—	3.55	2.99	4.73	—	—	—	7.47	4.49	
Focus	-0.48	-0.31	-0.63	-0.12	—	—	-0.99	—	-1.26	—	-0.26	0.26	-1.49	—	—	—	0.64	-0.14	
Power (2003)	4	0	2	0	3	4	0	4	0	3	3	2	2	0	2	0	0	1.71	
Power (2009)	4	0	3	1	3	4	0	4	1	3	4	2	2	2	4	2	1	2.35	

Note. IS alignment indicates the alignment of instruction with standards. IA alignment indicates the alignment of instruction with assessments.

Consistency. The alignment of a state's standards with its assessments defines the consistency of the standards-based reform policy. As seen in Table 3, the consistency of standards and assessments for teachers in the sample is generally low on the 0 to 1 scale, yet higher in mathematics and science than in ELA. The range is .01 to .51 in mathematics, .07 to .36 in ELA, and .09 to .37 in science. Because not all standards and assessments have been content analyzed using the SEC, it is impossible to compare these to national figures. However, a recent analysis of all available content-analyzed standards and assessments (a somewhat larger sample than is used here, because the present analysis is constrained to states and grades where teachers have also completed the SEC) found average alignments of .19 for ELA, .27 for mathematics, and .26 for science (Polikoff et al., 2011). Thus, the standards and assessments for the teachers in the sample used here are slightly more aligned than is typical.

Stability. The stability of the standards is measured as the length that the standards document has been in place. For the teachers in this sample, stability ranges from 0 to 9 years in ELA and from 0 to 10 years in mathematics and science. Average stability was approximately 3 to 5 years across subjects, with science and mathematics standards more stable than ELA during the study period. Overall, the sample of standards analyzed here is similar in average stability to national figures (4.1 to 4.7 years, depending on subject; American Federation of Teachers, 2008), and the stability figures suggest that the typical teacher is working with standards that are roughly 4–5 years old. Of the teachers in this study, Maine, Montana, and Wisconsin teachers are working with the oldest standards across subjects, as seen in Table 3.

Correlations among policy attributes. Given the descriptive information, one important question is the extent to which the available policy attributes are correlated. To the extent that they are highly correlated, it might be evidence that the attributes may be caused by one another or by some omitted variable. Examining correlations among the policy

attributes within states and years (i.e., correlations of the attributes after partialing out dummy variables for states and years) reveals that the attributes cluster in two pairs. In mathematics, the partial correlation of power with stability is .40, likely due to the increase in both attributes across time within states. The partial correlation of consistency with focus is .39. The other partial correlations among the attributes are all less than .07. In ELA, the partial correlation of power with stability is .87, and the partial correlation of consistency with focus is .20, whereas the other partial correlations are .09 or less. In science, the partial correlation of power with stability is .46, and the partial correlation of consistency with focus is .39, whereas the other partial correlations are .07 or less. Thus, the main likely concern about endogeneity among the attributes is between power and stability in ELA; this concern is investigated in the main analysis.

The Relationship of Policy Attributes With Aligned Instruction

An examination of the ICCs indicates that a substantial proportion of the variation in each of the outcomes is found at the state-grade-year level (an individual state-grade-year combination; the level at which the policy attributes are measured). The ICCs are as follows: mathematics standards, .62; mathematics assessments, .30; ELA standards, .43; ELA assessments, .50; science standards, .20; science assessments, .57. All ICCs are significantly different from zero. It is not clear why the ICCs vary so much across subjects and targets (tests, standards). However, given that previous research suggests that teachers' instruction across states at a particular grade level is similar (Porter, 2002), these high ICCs are suggestive that much of the difference in aligned instruction within states across grades and years is due to differences in the nature of the instructional targets.

The results of the six standardized regressions are presented in Table 4. The results for mathematics illustrate that the relationships of the policy attributes with the outcomes are quite consistent across the models. Consistency significantly predicts alignment to both standards ($\beta = .15$) and assessments ($\beta = .37$), controlling

TABLE 4
Fixed Effects Models of the Alignment of Teacher-Reported Instruction to Standards and Assessments (Standardized Betas and Standard Errors)

	Mathematics						English language arts						Science					
	Alignment to standards		Alignment to assessments		Alignment to standards		Alignment to assessments		Alignment to standards		Alignment to assessments		Alignment to standards		Alignment to assessments			
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE		
Policy variables																		
Focus	-.10	.08	-.11	.06	-.35***	.07	.04	.19	-.26	.19	-.48	.43						
Power	.14*	.05	.31	.13	-.03	.04	-.06	.20	.21**	.04	.00	.11						
Consistency	.15***	.01	.37***	.04	-.10*	.04	.38	.21	.05	.07	.49	.23						
Stability	.04	.03	.12	.07	.05	.04	-.05	.18	.06	.09	.13	.11						
Alignability	.12***	.01	.17***	.02	.23***	.04	.46***	.09	.35*	.13	.31**	.10						
Teacher controls																		
Experience	-.008	.005	-.031**	.007	-.029	.029	-.019	.013	.000	.016	-.004	.018						
Med. class	.021***	.005	.023	.017	.052**	.014	.006	.014	.073*	.028	.040	.030						
Large class	.013	.005	.000	.019	.066**	.018	.020	.020	.085**	.026	.055	.059						
Med. minority	.013**	.003	.013	.007	.005	.010	.006*	.002	.037	.034	.024	.019						
High minority	-.002	.004	-.030	.021	.028	.020	.038	.021	.047	.031	-.012	.023						
High achieving	.020	.014	.028	.041	.011	.005	.013	.012	.004	.012	-.015	.014						
Low achieving	-.014**	.004	-.014	.011	-.112***	.015	-.027	.024	-.036	.020	.005	.010						
Advanced degree	-.008	.007	-.024	.020	-.016	.007	-.045**	.013	-.043	.021	-.061	.043						
Content degree	.004	.014	.023	.031	.006	.011	.019	.012	.033*	.014	.070	.056						
n	4,421		4,421		3,907		3,907		1,633		1,633							
R^2	.56		.20		.46		.37		.19		.48							

Note. Models include state and year fixed effects. Standard errors are clustered at the state level.
* $p < .05$. ** $p < .01$. *** $p < .001$.

for alignability (also a significant predictor). That is, controlling for the extent to which standards or assessments tend to be aligned with other documents, the more aligned the assessments are with their target standards within states across grades and years, the more aligned instruction is with both standards and assessments. Stability and power are positively associated with both outcomes, although the only significant coefficient is for standards alignment ($\beta = .14$). Focus is negatively related to alignment to standards ($\beta = -.10$) and assessments ($\beta = -.11$), although neither coefficient is significant.

Second, the models explain a substantial proportion of the variation in the outcomes. Approximately 56% of the variation in standards-aligned instruction is explained by the model. The state policy attributes are particularly influential: consistency, focus, power, and alignability uniquely explain 4.7%, 0.5%, 0.3%, and 2.5%, respectively, of the variation in standards-aligned instruction after controlling for state and year fixed effects. Smaller proportions of variation in test-aligned instruction are explained by the policy attributes, but consistency remains the strongest predictor of alignment, uniquely explaining 7% of the variation in the outcome above and beyond that explained by the fixed effects.

The results for ELA are somewhat similar to the mathematics models. Again, focus, power, and stability are not significantly associated with test-aligned instruction, and stability is not significantly associated with standards-aligned instruction. The coefficient on the consistency variable in the model for test-aligned instruction is nearly the same magnitude as in the mathematics model ($\beta = .38$), although in this case it is not significant. Also, alignability is significantly positively associated with both standards- ($\beta = .23$) and test-aligned instruction ($\beta = .46$). In contrast to the mathematics models, consistency is *negatively* related to standards-aligned instruction in ELA ($\beta = -.10$), and power is negatively related to test-aligned instruction, although not significantly so.

The results for science are largely in line with those from the mathematics models, with 9 of the 10 coefficients on policy variables in the same direction in each of the two subjects

(only power for assessment-aligned instruction has a different sign, but this coefficient is not different from zero in either model). As was true for mathematics, power is associated with higher standards alignment ($\beta = .21$) and stability is not significantly associated with either form of aligned instruction. The degree of focus of state standards is also negatively associated with alignment to state standards ($\beta = -.26$) and assessments ($\beta = -.48$), although neither of these coefficients is significantly different from zero. Finally, consistency is positively associated with standards-aligned ($\beta = .05$) and test-aligned ($\beta = .49$) instruction, but not significantly.

Summary. There are some important patterns in the coefficients and statistical significance for the four policy attributes across the six models. The average of the six standardized coefficients for consistency is .22. Interestingly, consistency appears to be related to test-aligned instruction (average $\beta = .41$) but not standards-aligned instruction (average $\beta = .03$). On average, a 1 standard deviation difference in the degree of alignment between state standards and assessments (a difference of .07 to .12 on the alignment scale, depending on subject) within states across grades and state-years is associated with approximately a 0.4 standard deviation difference in teachers' instructional alignment to tests (a difference of .015 to .03 on the alignment scale). One plausible explanation for the different effects is that teachers increase their instructional alignment to the assessment only when that assessment is reflective of the content in the standards. It is important to note that just three of the six coefficients are significant despite their sometimes large magnitudes—two positive and one negative.

Another consistent predictor across the six models is the degree of focus of state standards. This variable is a significant negative predictor of alignment to ELA standards and is non-significantly negatively associated with all outcomes except alignment to ELA assessments. The average of the six coefficients is $-.21\sigma$, an effect size equal in magnitude but in the opposite direction to the average effect for consistency. Within states across grades and state-years, 1 standard deviation more focused

standards are associated with approximately one-fifth of a standard deviation less aligned instruction.

The other two policy variables show weaker or nonexistent effects across the models. Power is positively associated with five of six dependent variables, but only two are significantly different from zero. The average of the six coefficients is $.10\sigma$, indicating that a 1 standard deviation difference in power within states over time is associated with one-tenth of a standard deviation difference in aligned instruction. This is a smaller effect than those found for consistency and focus. Stability does not appear to be related to alignment—the average of the six standardized coefficients is $.06\sigma$, with no significant coefficients. Additionally, exploratory analyses involving interaction terms of policy attributes identified no consistent effects.

Alternate specification. The results summarized above are generally insensitive to alternate specifications. The main concern to be investigated was the moderate to high correlations among the policy attributes within states—focus and consistency were moderately correlated with one another, as were power and stability. By far the strongest correlation among the attributes was between stability and power in ELA. To investigate the effect of this correlation on the regression coefficients, the ELA regressions were re-run with stability and power alternately removed. With stability removed, the coefficient on power in the models for standards and assessments changed from $-.03$ to $.05$ and $-.06$ to $-.04$, respectively, with all coefficients not significantly different from zero. With power removed, the coefficient on stability in the models for standards and assessments changed from $.05$ to $-.05$ and $-.05$ to $-.05$, respectively, with all coefficients not significantly different from zero. Despite the high correlation between these two attributes, there is little evidence that the results are affected by the presence of one variable or another in the models in ELA.

To investigate more generally the effects of the correlations among the policy attributes on the regression coefficients, all the regressions were re-run one at a time with each focal independent variable removed from the equation. Removing stability from the models did not

affect the average regression coefficient on the power variable—the mean coefficient changed from $.10$ to $.11$. Similarly, removing power from the regression equation changed the mean coefficient on stability from $.06$ to $.02$. Neither were the coefficients on consistency and focus appreciably affected by removing one or the other from the model. Removing consistency from the model changed the average coefficient on the focus variable from $-.21$ to $-.26$, whereas removing focus changed the average coefficient on the consistency variable from $.22$ to $.25$. Despite the moderate correlations among the attributes, removing one variable from a correlated pair did not meaningfully affect the average coefficients on the other variable in the pair.

Discussion

Over the past 20 years, states and the federal government have increasingly pursued standards-based educational reforms as a solution to low achievement and the large achievement gaps in U.S. education. Although NCLB has effectively mandated SBR, the law has left considerable discretion as to the specific features of SBR to be implemented in each locale. As a result, previous research has documented wide variation in the implementation of SBR along several key dimensions (Finn, Petrilli, & Julian, 2006; Polikoff et al., 2011; Porter et al., 2005; Porter et al., 2009; Porter et al., 2007). Given these differences, the purposes of this article were to (a) describe how state standards-based reform policies vary on the key policy attributes that may shape teachers' instructional responses and (b) describe the relationship between these policy attributes and teachers' instructional alignment. The results indicate great variation in attributes across states and several small-to-moderate relationships of policy attributes with instructional alignment.

There are several important limitations of this work. First, the results would have more external validity if the sample were a population-based probability sample as opposed to being a convenience sample of teachers. Furthermore, there were a limited number of states in these analyses, which reduced the power to detect effects. Second, the data did not allow for estimation of changes in alignment at the individual teacher level or group level. It would be a

worthwhile project, moving forward with the adoption of the Common Core Standards, to follow a set of teachers to investigate how they change their instruction during implementation. Third, the analyses of the relationships of policy attributes with instructional alignment were necessarily correlational. Thus, it is impossible to support claims about causal relationships between the policy attributes and the content of instruction. There is also the potential for omitted variable bias. There are many potential omitted state policy variables that could be included in the model, and there is no sure way to rule out the influence of omitted variables. The purpose of the state fixed effects is control for time- and grade-invariant state characteristics, but perhaps there are other things changing over time or across grades within states that are correlated with either the state policies or the outcomes.

Finally, this study cannot say anything about the relationship between policy attributes and the *quality* of teachers' instruction. Indeed, it is possible that more consistent standards and assessments would produce greater instructional alignment, but this alignment might lead to lower quality instruction than the status quo (e.g., rote). This study was based on the notion that alignment is a core mediating variable in the theories of standards-based reform; thus, standards-based reform policies should promote alignment regardless of the resultant quality of instruction. Although it might be possible to measure the relationship between policy attributes and instructional *quality* across states and grades, such a study would require extensive national data collection—well beyond the scope of the present study. Nonetheless, it would be a worthwhile investigation.

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 were limited in relying on teacher self-report of alignment and/or self-report of instructional change. Furthermore, those studies rarely investigated the relationships between state policies or policy features and instruction. The data used here are weaker than national probability samples in their generalizability, but they are strong in terms of their validity in representing the content of instruction.

Additionally, the ability to directly compare instruction with the content of standards and assessments is a strength of this study, as the study therefore does not rely on problematic assumptions about teachers' understanding of standards and assessments.

The results presented here may have implications for policy, as standards-based reform continues to mature. The most obvious implication is that the alignment of standards with assessments may be an important feature related to teachers' content decisions. Although this idea was one of the key underlying concepts guiding the creation of standards-based reform, the low average standard-assessment alignment found here indicates that state SBR systems are not as tightly coupled as was originally intended. In many cases, teachers are receiving vastly different messages about what to teach depending on whether they are looking at standards or assessments. To maximize the effectiveness of the standards-based reform policy, one potentially fruitful approach is to seek better alignment of standards with assessments. Even if the results presented here do not provide the causal justification for supporting such a policy recommendation, research indicates that standards assessment is unacceptably low in many states (Polikoff et al., 2011). Thus, improving standards-test alignment should be a worthy goal regardless, and the results presented here provide at least cautious hope that such changes might also result in improvements in aligned instruction. Should the instructional targets be better (as many believe the new Common Core State Standards are, as compared with many states' standards), such instructional changes might indeed result in improved student outcomes as well.

A second implication for policy is in thinking about the focus of teachers' instruction. A persistent complaint about U.S. education, especially in mathematics (Schmidt et al., 2001), is that it is too broad and shallow. Although it might be hoped that more focused standards and assessments would lead to instructional narrowing, the results presented here do not suggest that alignment is high when standards are focused. If policymakers want to focus instruction on a few core ideas, it may be unlikely that standards alone will motivate these desired changes. It is quite likely that more work will have to be done

in terms of planning, curriculum development, and professional development. Based on these correlational analyses, highly focused standards seem to be difficult for teachers to implement with fidelity based on their current practices alone.

Although the results described here do not provide causal evidence, they are suggestive of ways that policies may be influencing teachers' decisions about what to teach. Standards-based reform has been a prominent policy for roughly two decades, but there is work to be done if its theory of change is to produce the expected outcomes in terms of teacher performance and student learning.

References

- American Federation of Teachers. (2008). *Sizing up state standards 2008*. Washington, DC: Author.
- Camilli, G., & Firestone, W. A. (1999). Values and state ratings: An examination of the state-by-state education indicators in Quality Counts. *Educational Measurement: Issues and Practice, 18*(4), 17–25.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*(4), 305–331.
- Center on Education Policy. (2009). *State high school exit exams: Trends in test programs, alternate pathways, and pass rates*. Washington, DC: Author.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy.
- Clune, W. H. (1993). Systemic educational policy: A conceptual framework. In S. H. Fuhrman (Ed.), *Designing coherent educational policy* (pp. 125–140). San Francisco, CA: Jossey-Bass.
- Clune, W. H. (2001). Toward a theory of systemic reform: The case of nine NSF statewide systemic initiatives. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the state. One hundredth yearbook of the National Society for the Study of Education, Part II* (pp. 13–38). Madison, WI: University of Chicago Press.
- Coe, C. K. (2003). A report card on report cards. *Public Performance and Management Review, 27*(2), 53–76.
- Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis, 12*(3), 233–239.
- Cohen, D. K., & Spillane, J. P. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education, 18*, 3–49.
- Darling-Hammond, L. (1990). Instructional policy into practice: “The power of the bottom over the top.” *Educational Evaluation and Policy Analysis, 12*(3), 339–347.
- Desimone, L. M. (2002). How can comprehensive school reform models be successfully implemented? *Review of Education Research, 72*(3), 433–479.
- Desimone, L. M., Le Floch, K. C., Ansell, S., & Taylor, J. (2005). *Are we asking the right questions on national surveys about teachers' professional development? Insights from cognitive interviews with teachers*. Paper presented at the annual meeting of the American Educational Research Association.

- Education Week*. (2009). *Quality counts 2009: Portrait of population*. Bethesda, MD: Editorial Projects in Education.
- Finn, C. E., Petrilli, M. J., & Julian, L. (2006). *The state of state standards 2006*. Washington, DC: Thomas B. Fordham Foundation.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95–113.
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325–338.
- Hamilton, L. S., & Berends, M. (2006). *Instructional practices related to standards and assessments* (RAND Working Paper WR-374-EDU). Santa Monica, CA: RAND.
- Koretz, D. M., Mitchell, K. J., Barron, S. I., & Keith, S. (1996). *Final report: Perceived effects of the Maryland School Performance Assessment Program*. Los Angeles, CA: CRESST.
- McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge, MA: Harvard University Press.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (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, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal, 48*(4), 965–995.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.
- Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws & T. J. Cooney (Eds.), *Perspectives on research on effective mathematical teaching* (pp. 96–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms*. Madison, WI: Wisconsin Center for Education Research.
- Porter, A. C., Linn, R. L., & Trimble, C. S. (2005). The effects of state decisions about NCLB adequate yearly progress targets. *Educational Measurement: Issues and Practice, 24*(4), 32–39.
- Porter, A. C., & Polikoff, M. S. (2007). NCLB: State interpretations, early effects, and suggestions for reauthorization. *Social Policy Report, 21*(4), 1–15.
- Porter, A. C., Polikoff, M. S., & Smithson, J. (2009). Is there a de facto national intended curriculum? Evidence from state content standards. *Educational Evaluation and Policy Analysis, 31*(3), 238–268.
- Porter, A. C., Polikoff, M. S., Zeidner, T., & Smithson, J. (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, A. C., Smithson, J., Blank, R., & Zeidner, J. (2007). Alignment as a teacher variable. *Applied Measurement in Education, 20*(1), 27–51.
- Raizen, S. A. (1988). *Increasing educational productivity through improving the science curriculum*. Washington, DC: OERI.
- Ravitch, D. (1995). *National standards in American education: A citizen's guide*. Washington, DC: Brookings Institution.
- Reys, B. (Ed.). (2006). *The intended mathematics curriculum as represented in state-level curriculum standards: Consensus or confusion?* Charlotte, NC: Information Age Publishing.
- Reys, B., Chval, K., Dingman, S., McNaught, M., Regis, T. P., & Togashi, J. (2007). Grade-level learning expectations: A new challenge for elementary mathematics teachers. *Teaching Children Mathematics, 14*(1), 6–11.
- Ross, M. (1989). The relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341–357.
- Ross, M., & Conway, M. (1986). Remembering one's own past: The construction of personal histories. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 122–124). New York: Guilford.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco, CA: Jossey-Bass.
- Schwartz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*(2), 127–160.
- Schwille, J., Porter, A. C., Alford, L., Floden, R., Freeman, D., Irwin, S., et al. (1988). State policy and the control of curriculum decisions. *Educational Policy, 2*(1), 29–50.
- Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: Politics of*

- Education Association yearbook* (pp. 233–267). Bristol, PA: Falmer Press.
- Snyder, T. D., Dillow, S. A., & Hoffman, C. M. (2008). *Digest of education statistics, 2007* (No. NCES 2008–022). Washington, DC: National Center for Education Statistics.
- Spady, W. G., & Mitchell, D. E. (1979). Authority and the management of classroom activities. In D. L. Duke (Ed.), *Classroom management: The seventy-eighth yearbook of the National Society for the Study of Education* (pp. 75–115). Chicago, IL: University of Chicago Press.
- Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.
- Stecher, B. M., Barron, S. I., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms*. Los Angeles, CA: CRESST.
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*, 211–227.
- Taylor, G., Shepard, L. A., Kinner, F., & Rosenthal, J. (2002). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost*. Los Angeles, CA: Center for the Study of Evaluation.

Author

MORGAN S. POLIKOFF, Ph.D., is an assistant professor of education at the University of Southern California, Waite Phillips Hall 904D, Los Angeles, CA 90089, USA; polikoff@usc.edu. His research focuses on standards, assessment, and accountability policies and their influence on teachers' instruction and student achievement.

Manuscript received September 16, 2010

Revision received September 29, 2011

Accepted October 24, 2011