American Educational Research Journal
December 2017, Vol. 54, No. 6, pp. 1288–1315
DOI: 10.3102/0002831217718164

Accommodating Change: Relating Fidelity of Implementation to Program Fit in Educational Reforms

Eleanor R. Anderson
Northwestern University

Asking practitioners to make larger changes to their practice is often thought to lead to lower fidelity of implementation. However, salient differences between ambitious new reforms and teachers' existing practices may also facilitate processes of conceptual change and correspondingly increase fidelity of implementation. I use survey data on the implementation of two Comprehensive School Reform programs to investigate this puzzle, presenting a series of descriptive multivariate regressions that—contrary to conventional wisdom—predominantly support a positive association between larger changes and higher fidelity. I also address alternative explanations for this finding and discuss the conceptual and empirical strengths and weaknesses, implications for future research, and potential utility for practice of each interpretation.

Keywords: comprehensive school reform, conceptual change, fidelity of implementation, magnitude of change, organizational change

Leaders seeking to make change in their schools often do so through the adoption of evidence-based interventions intended to establish high-quality pedagogical practices in schools and classrooms (Desimone, 2002). Evaluations of such programs in practice, however, have shown uneven implementation of their recommendations (Corcoran, Hoppe, Luhm, & Supovitz, 2000; Desimone, 2002; Sherin & Drake, 2009; Spillane, 2004). Like the frontline workers in other sectors—arguably even more so—teachers often exert wide discretion in interpreting and carrying out the elements of a reform adopted by their school or district (Berends, 2000; Lipsky, 1980;

Eleanor R. Anderson is a postdoctoral fellow in the School of Education and Social Policy at Northwestern University, 2120 Campus Drive, Annenberg Hall 238, Evanston, IL 60208; e-mail: *eanderson@u.northwestern.edu*. Her research interests include the implementation and institutionalization of educational reforms and the cognitive, social, organizational, and political processes that facilitate and undermine high-quality educational practice.

Scott & Meyer, 1983). Given strongly entrenched ideas about the nature of teaching and learning, not to mention substantial constraints of time, resources, and organizational support, at times this can result in "new" practices that are only superficially different from old ones (Coburn, 2004; D. K. Cohen, 1988; Mehta, 2015; Spillane, 2004).

It is not enough then to simply design or select a program whose goals align with a particular vision for teaching and learning practices. Those seeking to bring their vision to life must also ask to what degree enactment in a given context is likely to match the program's goals. While the benefits of pursuing strict adherence to a program's designed elements may need to be weighed against other considerations (McLaughlin, 1987), for those decisions to be well informed, leaders and policymakers need good information about the dynamics that affect fidelity of implementation.

It is taken for granted, in most cases, that in adopting new programs, schools are asking teachers to make changes to their practice. However, the scope and scale of those changes may vary dramatically from case to case. The magnitude of change a new program asks of teachers is a reflection of its fit with existing practices, that is, it depends both on the specifications of the program itself and also on the practices teachers were engaging in previously (Coburn, 2004).

Put together, these issues present researchers, policy designers, and school leaders with a puzzle: Is high-fidelity implementation more likely in cases where a new program is similar to teachers' existing practices, minimizing the change they are asked to enact? Or is fidelity increased when newly specified practices are very different from existing practices, signaling to practitioners that real change has arrived? That is, how is program fit related to fidelity of implementation?

Conventional accounts of policy implementation and organizational change have postulated that practitioners asked to make bigger changes generally implement those changes with lesser fidelity (D. K. Cohen & Moffitt, 2009; W. M. Cohen & Levinthal, 1990; Guskey, 1991; Mazmanian & Sabatier, 1983; Van Meter & Van Horn, 1975). In this paper, I challenge that assumption, offering evidence that the opposite may often be true: Practitioners asked to make bigger changes may actually implement those changes with *greater* fidelity. Drawing on literature in teacher learning and conceptual change, I demonstrate that a cognitive account of implementation suggests plausible mechanisms for both possibilities. Using survey data on the implementation of two Comprehensive School Reform (CSR) programs, I present a series of descriptive multivariate regressions that predominantly support the latter relationship—an association between larger change and higher fidelity. I conclude by addressing a number of possible explanations for the empirical finding, discussing their conceptual and empirical strengths and weaknesses, implications for future research, and potential utility for practice.

Conceptual Framework

Terminology

In this paper, I am explicitly theorizing about organizational-level educational reforms, that is, programs adopted at the school level that specify aspects of the work of individuals within the school. I consider such programs to be school-level policies, and thus for the purposes of this paper I use the terms *program* and *policy* interchangeably.

Fidelity of Implementation

For years, social and behavioral researchers have called for increased attention to the fidelity with which programs and policies are implemented (Century, Rudnick, & Freeman, 2010; Dane & Schneider, 1998; Fullan, 1983). Scholars have taken a variety of conceptual and methodological approaches to measuring fidelity of implementation, corresponding to varying research questions.

For studies intended to measure the efficacy or effectiveness of a particular program or policy, knowing how much of an intervention was actually administered is critical to understanding its impact. Even under experimental conditions, without information about fidelity, mixed or limited results could be the result of the inherent (in)efficacy of a program or of inadequate implementation (Desimone, 2002; Dusenbury, Brannigan, Falco, & Hansen, 2003; Flay, 1986; Linder & Peters, 1987; Shadish, Cook, & Campbell, 2002). Large-scale experimental or quasi-experimental designs often use relatively coarse measures of fidelity of implementation. At the most basic level, this may simply take the form of a binary compliance variable or a dosage threshold to differentiate intent-to-treat from treatment-on-treated samples. In other cases, researchers may use a measure of dosage to estimate of the effect of increasing exposure to the intervention (Cordray & Pion, 2006; Lipsey & Cordray, 2000).

Other studies are designed to investigate the processes through which program and policy implementation occur. For example, they may seek to describe processes of sensemaking and interpretation (e.g., Coburn, 2004; Spillane, 2004), how teachers adapt a program to fit their local needs (e.g., Datnow & Castellano, 2000a), or how qualitatively different forms of the same policy take hold and interact with different organizational cultures (e.g., Lin, 2002). In these cases, a one-dimensional measure of fidelity is often wholly inadequate (Century & Cassata, 2016). Rather, these studies often use ethnographic observation in one or a small number of schools to provide rich descriptions of the many different forms enacted policy can take.

A third class of studies seeks to describe relationships between various aspects of implementation as a process and an outcome. Such studies might include investigations of the importance of teacher buy-in to program

sustainability (Berends, 2000), the relationship between program type and styles of management (Rowan, Camburn, & Barnes, 2004), or the role of teacher knowledge in program efficacy (Phelps & Schilling, 2004). It is in this spirit that I use the construct of fidelity to intervention prescriptions—not as normative requirement but as an important factor in the dynamics of implementation. While this approach cannot capture the rich variation across different cases of implementation, its simplicity makes it well suited for operationalization across a large number of observations, offering a different type of potency.

Among the limitations of the construct of fidelity across all these uses is that it assumes a program that offers specific, observable prescriptions. Many programs and policies are not of this character at all, being much more ambiguous, and in these cases, fidelity is largely undefined (D. K. Cohen & Moffitt, 2009; Majone & Wildavsky, 1979; Matland, 1995). Organizational scholars have noted that innovations transform over time as they diffuse among organizations, a process also obscured by traditional definitions of fidelity (Ansari, Fiss, & Zajac, 2010). Other objections are more philosophical: The term often carries a normative valence that seems to suggest that practitioners *should* implement programs exactly as they are designed, in spite of the fact that teachers and other frontline workers often have information about local settings that developers don't (McLaughlin, 1987).

I agree that these limitations point to the need for better constructs in this area (Century & Cassata, 2016); however, that endeavor is not one I take up in this paper. Rather, I define fidelity of implementation as the similarity between enacted practice and the benchmark of program designers' specifications (Dane & Schneider, 1998; Lewis & Seibold, 1993). I consider the constraints this imposes on which programs I am able to explore empirically to represent boundary conditions of the analysis.

Implementation and Conceptual Change

Frontline workers shape the way new programs and policies are implemented in a host of ways. Programs cannot specify literally everything a teacher is to do. Therefore, teachers must fill in the blanks of the moment-to-moment specifics of how they will implement a program (Coburn & Stein, 2010). In addition, teachers can and often do choose to make adaptations to what is written or prescribed (Corcoran et al., 2000; Datnow & Castellano, 2000b). As a rule, what makes adaptations problematic is when they undermine some important principle underlying the reform's effect (Datnow & Castellano, 2000a; Spillane, 2004; Supovitz, Poglinco, & Bach, 2002). This means that teachers' degree of understanding of key principles (as well as their willingness and ability to follow them) is critical to fidelity of implementation (Gregoire, 2003; Spillane, 2004; Spillane, Reiser, & Reimer, 2002).

Constructivist theories of learning indicate that people's understandings of new ideas are based on and built from their existing conceptions (Rumelhart, 1980; Schank & Abelson, 1977). Everyday learning means making minor changes in the organization of existing conceptions or integrating new ideas into existing cognitive structures (Carey, 1988; Fosnot, 1996; Posner, Strike, Hewson, & Gertzog, 1982; Rumelhart, 1980). However, cognitive scientists have long observed that some ideas are more difficult to learn than others. A number of examples of such challenges have been documented in science education, including the concept of force (DiSessa, 1993), processes of natural selection (Brumby, 1984), and diffusion (Chi, 2005).

Scholars of program and policy implementation have noted similar patterns with regard to changes in educational philosophy. For example, Cohen's (1990) classic study of "Mrs. Oublier" illustrated a teacher who felt that she had made significant changes to her instructional practice based on California's mathematics reform. However, in observing her teaching, Cohen found that her practices adhered to some of the reform's more superficial prescriptions but deviated significantly from other, more philosophically unfamiliar ones. Even with enthusiastic effort, reforms can be challenging to wrap one's mind around.

Several streams of thought exist as to how and under what circumstances teachers are able to make significant shifts in their understanding (Gregoire, 2003). One set of approaches draws a distinction between small and large changes in cognitive structure, often referred to as *assimilation* and *accommodation*, respectively (Carey, 1988; Fosnot, 1996; Piaget, 1977; Posner et al., 1982; Strike & Posner, 1992). When new ideas are dramatically different from existing conceptions, making only small changes in cognitive structures through the processes of assimilation will typically lead to misconceptions. Other theories emphasize differences in the type of cognitive processing, distinguishing between heuristic and systematic processing. Heuristic processing is faster but more shallow and rarely leads to lasting transformations in understanding (Eagly & Chaiken, 1993; Gregoire, 2003).

Scholars have also offered varying explanations for what prompts these different cognitive processes. For example, among other factors, Posner et al. (1982) indicate that for cognitive accommodation to occur, a learner must experience dissatisfaction with his or her existing conceptions. In the attempt to interpret new information, a conflict must become apparent between existing conceptions and the new idea that is salient enough to prompt a rethinking and ultimately reorganization of existing understandings.

Other scholars have critiqued this approach for being excessively rational and have offered models that foreground affect, motives, and goals above and beyond the rationalistic inclination for consistency (Strike & Posner, 1992). For example, Gregoire (2003) provides an integrated model beginning with teachers' assessments of whether a reform message implicates their own practices and beliefs. Only if teachers both perceive

implications for themselves and have sufficient motivation and ability will they engage with the reform's ideas through systematic processing that may lead to more substantial changes in belief.

Program Fit

This study's central question concerns a new program's fit with existing practices and thus the magnitude of change it asks teachers to enact. Many researchers have noted variation in this factor from one implementation case to the next. Policy implementation literature has often taken a macrolevel perspective, treating the ambitiousness of a reform as a feature of the policy itself (D. K. Cohen & Moffitt, 2009; Mazmanian & Sabatier, 1983; Van Meter & Van Horn, 1975). Some organizational researchers have taken a more mesolevel perspective, noting that heterogeneity in implementing contexts means that the same program or innovation may have differing degrees of fit and misfit in different organizations, thus entailing changes of differing scope (Ansari et al., 2010; W. M. Cohen & Levinthal, 1990). In the context of educational reforms, even an organizational-level analysis may not be fine-grained enough as fit with a new program may vary from individual to individual even with a single school (Coburn, 2004; Sherin & Drake, 2009). Thus, for this paper, I define the magnitude of change as the degree of difference between a new program's specifications and an individual teacher's existing practices.

Relating Fit and Fidelity

Most existing policy implementation research, both in educational settings and more broadly, suggests that—other things being equal—asking practitioners to make larger changes tends to reduce fidelity of implementation (D. K. Cohen & Moffitt, 2009; W. M. Cohen & Levinthal, 1990; Guskey, 1991; Mazmanian & Sabatier, 1983; Van Meter & Van Horn, 1975). As Guskey (1991) puts it, "if there is one truism in the vast research literature on change it is that the magnitude of change persons are asked to make is inversely related to their likelihood of making it" (p. 241).

Constraints on material resources are perhaps the most straightforward obstacle to implementation. For example, money and support provided by the policy itself may or may not be adequate to the implementation task (D. K. Cohen, Moffitt, & Goldin, 2007; Van Meter & Van Horn, 1975). In the education sector, structural opportunities for enforcing accountability may be especially limited by the highly autonomous and uncertain nature of the work itself (D. K. Cohen, 1988; Lipsky, 1980; Scott & Meyer, 1983), although recent research has suggested this may be changing (Hallett, 2010; Spillane & Burch, 2006). To the extent that larger changes require greater resources or more supervision, this suggests they will be implemented with less fidelity.

Theories of conceptual change offer some evidence for a connection between larger changes and lower fidelity as well. Programs designed for educational reform are frequently built around conceptions of subject matter, teaching, and/or learning that differ substantially from those held by many practitioners (indeed, this is what marks them as reforms). Thus, differences between educators' existing ideas and those espoused by a new program are unlikely to be very small but rather may range from moderate to quite large. New ideas that do not align with people's existing conceptions are more challenging to recognize and understand (Carey, 1988; Gregoire, 2003; Posner et al., 1982; Strike & Posner, 1992). Thus, one might argue that larger changes would be increasingly likely to prompt adaptations during implementation that undercut the principles of the reform, therefore being associated with lower fidelity.

One important source of empirical evidence supporting this prediction comes from Cynthia Coburn's (2004) in-depth case comparison of three California teachers' responses to conflicting messages about literacy instruction. Using extensive shadowing, oral histories, and document analysis, Coburn identified several categories of response to policy messages including either minor or major changes to practice.[1] She also identified several message characteristics that seemed to consistently prompt different responses, including the degree of "congruence" between a new message and teachers' existing practices and beliefs. Coburn found teachers incorporated high congruence messages by making minor changes in their practice relatively frequently. When message congruence was low, teachers were much more likely to reject those messages outright and not incorporate them at all.

Yet, a few studies have suggested that this may not be the full picture (Huberman & Miles, 1984). For example, Correnti and Rowan (2007) point out that several examples of school reforms that have been successful in bringing about substantial changes in teaching practice did so using programs that were "ambitious and represent a marked change in existing practices" (p. 302).

Indeed, theories of conceptual change also support this competing prediction. For programs representing at least a moderate change from existing approaches, the bigger the difference, the more salient and problematic it may become for practitioners and the more clearly it may implicate their own practice—making it increasingly likely that it will be interpreted through accommodation or systematic processing. This provides a substantial theoretical reason to predict that larger changes might be understood better and thus implemented with greater fidelity. Spillane et al. (2002) make a similar point, emphasizing the importance of this type of cognitive conflict: "It is key to create a sense of dissonance in which agents see the issues in their current practice rather than seeing the new ideas as achieved within their current practice" (p. 418).

Indeed, Coburn's (2004) study offers empirical evidence toward this point as well. Coburn found that when teachers did incorporate low

congruence messages into their classroom practice, they were much more likely to do so "in ways that pushed their thinking or caused them to reorganize their practice in more substantial ways" (p. 228). That is, when taken up at all, these larger asks were more likely to result in substantial changes to practice, aligned to the principles of the new approach.

Coburn's (2004) analysis drew on the responses of only three teachers. However, her findings underscore the potential significance of the magnitude of change a program asks in the process of policy implementation. The notion that large policy changes might frequently be implemented with higher fidelity than smaller ones is quite contrary to the conventional wisdom of policy implementation. Strong evidence of such a relationship would have significant implications for policy design, selection, implementation, and analysis and set an important agenda for future research in the area. To examine this possibility empirically with a large sample of teachers, I turn to the case of Comprehensive School Reform programs.

Empirical Analyses

Setting: Comprehensive School Reform

Schoolwide reform models experienced a meteoric rise in prominence and funding during the 1990s and 2000s. To qualify as a CSR, a program must be comprehensive, addressing the need for change systemically at the whole school level rather than only one subject or grade level or classroom at a time and also specify practices that are evidence-based, supported by research demonstrating their effectiveness (Borman, Hewes, Overman, & Brown, 2003; Comprehensive School Reform Quality Center, 2005; Orland, Hoffman, & Vaughn, 2010).

I focus on two CSR programs: America's Choice (Corcoran et al., 2000; Glazer, 2009; Poglinco et al., 2003; Supovitz et al., 2002; Supovitz, Poglinco, & Snyder, 2001) and Success for All (Datnow & Castellano, 2000b; Slavin et al., 1996; Slavin & Madden, 1999, 2000). Both programs are well specified, making them appropriate for an analysis of fidelity. Both programs also were designed around conceptions of achievement that differ substantially from the norm in many schools, making them appropriate for examining mechanisms related to conceptual change.

America's Choice (AC) was developed by the National Center on Education and the Economy and first implemented in a cohort of schools in 1998 (Corcoran et al., 2000; Glazer, 2009; Supovitz et al., 2001). The program came out of the standards movement and is built around a set of internationally benchmarked standards (National Center on Education and the Economy & University of Pittsburgh, 1997), which teachers are expected to use to develop their instructional strategies. The central philosophy is to hold all students to high expectations rather than comparing them to

one another. Success for All (SFA) was developed by researchers at Johns Hopkins University and first implemented in 1987 (Slavin et al., 1996). The program's central philosophy is that "every child can and must succeed in the early grades, no matter what it takes" (Slavin et al., 1996, p. 43). Teachers and administrators are to identify difficulties as early as possible and intervene intensively, working "relentlessly" to ensure that every child learns to read (Slavin et al., 1996; Slavin & Madden, 2000).

Both programs offer prescriptions for staffing, professional development, materials, assessment, and other schoolwide practices. However, because the current study is concerned with teachers' role in the implementation process, for this analysis, I focus specifically on each program's specifications for classroom practice. Differences in the ways the two programs are structured means that high-fidelity implementation has different requirements for teachers implementing each CSR.

America's Choice offers clear recommendations for how teachers should spend their time in literacy instruction, with each day including a 2- to 2.5-hour block including an hour of Readers Workshop and an hour of Writers Workshop. While the program includes training in a range of instructional strategies for implementing these workshops, teachers do not have a script to follow, nor are all materials provided. A study of the implementation of AC found substantial variation in teachers' literacy lessons, ranging from high-quality readers and writers workshops, to workshops lacking in important elements, to no discernable workshop structure at all (Supovitz et al., 2002). Thus, high-fidelity implementation of AC involves understanding the purpose of these different elements well enough to tie them together coherently and make use of the strategies and resources that America's Choice recommends.

Success for All is a much more highly regimented literacy program in which teachers are expected to follow detailed 90-minute daily lesson plans using SFA-provided materials. A study of the implementation of SFA found that despite the clarity and specificity of instructions, most teachers made modifications in their implementation of SFA lessons (Datnow & Castellano, 2000a, 2000b). Some were relatively minor or seen by developers as appropriate adaptations to variations in student population. However, many included significant deviations from the design of the program. The most common adaptations involved spending extra time on certain areas of the lesson and eliminating or making substitutions for others. Contrary to prior research on other reforms, the inclination toward adaptation did not seem closely linked with teacher characteristics, including level of experience, gender, race, or even belief in the program's efficacy. Thus, high-fidelity implementation of SFA involves using the program-provided curriculum with minimal adaptation, even when it contradicts one's own preferences as a teacher.

Data and Measures

The data used for this study are drawn from those collected by University of Michigan researchers as part of the Study for Instructional Improvement (SII). SII was a nationwide, quasi-experimental study designed to measure the effects of comprehensive school reform programs in high-poverty elementary schools (Correnti & Rowan, 2007; Rowan, Correnti, Miller, & Camburn, 2009). The data I use were collected between 2000 and 2004, at 61 schools, each of which had adopted either SFA or AC starting in 1998, 1999, or 2000.[2] For each of the three years each school participated in the study, every teacher was administered a Teacher Questionnaire that included questions about teaching practices, perceptions of the school improvement program, educational background, teaching experience, and demographic information. Data about teaching practices were also collected in the form of Language Arts Logs.[3] Logs were designed to capture the daily instruction experienced by two cohorts of focal students in each school as they progressed from Grades K–2 or 3–5; teachers completed logs only if and when they were responsible for the language arts instruction of one or more of the focal students on the assigned log day. The response rate ranged from 84% to 77% over the course of the study (Rowan et al., 2009). To measure fidelity using the Log data, I take the annual average of each teacher's fidelity score from each log he or she completed, resulting in one fidelity score per respondent per year of available data.

The sample represented in these data is shown in Table 1. A total of 556 teachers filled out both at least one log and enough of the teacher questionnaire to be usable for this analysis. Spread over the four years of the study, this corresponds to 1,267 observations. This comprises the subsample to which all subsequent analyses will refer. This subsample is similar in composition to the complete sample with respect to teacher experience, education, demographics, and employment category. The one area in which it differs dramatically is in the relative representation of America's Choice and Success for All. While in the complete sample, these are represented nearly equally, in the subsample, there are more than twice as many SFA observations as AC observations. For this reason, in addition to differences between the programs themselves, all analyses are presented separately for AC and SFA teachers.

CSR Program Fit

As discussed previously, I conceptualize program fit as the magnitude of difference between teachers' existing practices and those specified by the program. No external measure of teachers' practices prior to program implementation exists in the SII data. Therefore, I use a teacher self-report of whether or not "The school improvement program in this school requires me to make major changes in my classroom practice." The question allowed four answers: *strongly disagree, disagree, agree*, or *strongly agree* (Study of

Table 1
Study Samples

	Complete Sample	Subsample
Total		
Observations (1/teacher/year)	4,844	1,267
Unique teachers	2,815	556
Teachers with multiple years of data	1,123	295
Intervention (%)		
America's Choice	53.39	31.81
Success for All	46.61	68.19
Teacher		
Mean experience (years)	12.85	11.93
Teacher education (%)		
Holding undergraduate degree	95.50	98.50
Holding graduate degree	68.67	71.98
Teacher demographics (%)		
Female	86.61	87.92
Non-White	48.98	46.17
Employment category (%)		
Full-time	89.24	95.66
Permanent/standard certification	77.62	81.53

Instructional Improvement, 2001, 2002, 2003, 2004). I take teachers' response to this question as a measure of the magnitude of change the program asks of them, with those who strongly disagreed experiencing the smallest changes and those who strongly agreed experiencing the largest changes.

Figure 1 shows the average reported program fit for teachers who answered the survey at least twice. It shows that on average, AC teachers report that their CSR program requires a larger change than SFA teachers do. In the full sample, both AC and SFA teachers report program misfit as progressively smaller each time that they answer the survey. The same pattern holds in the study sample for SFA teachers and AC teachers who answered the survey more than twice.

Because I use this measure to approximate the magnitude of the prescribed change at the time of the program's initial adoption, rather than the way the magnitude of a remaining change affects fidelity over time, I use individuals' earliest recorded response to this question for all subsequent analyses.

Figure 2 is a histogram representing the distribution of teachers' first reports of their CSR program's fit with existing practices. Each bar represents the proportion of teachers who strongly disagree (SD), disagree (D), agree (A), and strongly agree (SA) that their school improvement program requires major changes to their classroom practice for AC and SFA, respectively.

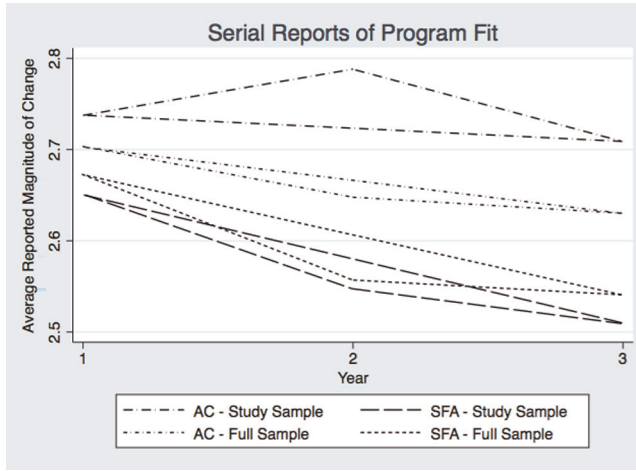


Figure 1. **Average CSR program fit reported by teachers who responded to the annual Teacher Questionnaire at least twice. Magnitude of prescribed change measured by responses to "The school improvement program in this school requires me to make major changes in my classroom practice" with *strongly disagree* = 1, *disagree* = 2, *agree* = 3, *strongly agree* = 4.**

Comparing across programs, slightly more AC teachers (58%) agree or strongly agree that AC requires major changes than the proportion of SFA teachers agreeing or strongly agreeing that SFA requires major changes (53%). Comparing within each program, for both AC and SFA, teachers' reports of the magnitude of change the program asks of them vary widely. Among teachers implementing America's Choice, 5% report that they strongly disagree and 37% that they disagree that the program requires major changes to their classroom practice, while 45% agree and 14% strongly agree that it does. Similarly, among teachers implementing Success for All, 4% report that they strongly disagree and 43% that they disagree that the program requires major changes to their classroom practice, while 38% agree and 14% strongly agree that it does. These data indicate that teachers implementing the same program differ in how large a change they experience.

To confirm the proper level at which to examine this variation, I looked for evidence that the variation in program fit shown in Figure 2 might be driven by school-level differences in preexisting teacher practices. Tables 2 and 3 show numerically how much of the variation in magnitude of change occurs among teachers within schools as compared with the variation between schools.

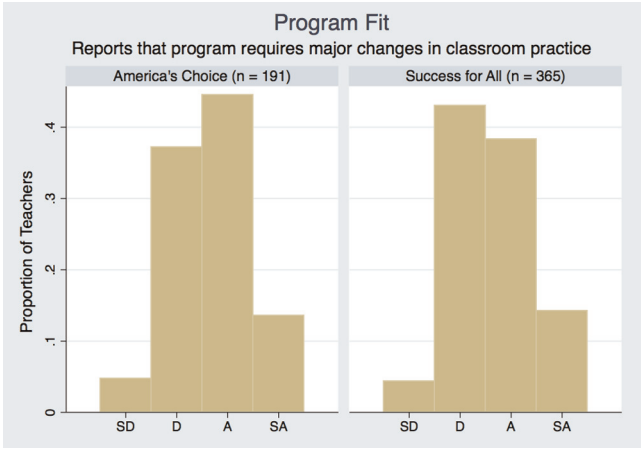


Figure 2. **Proportion of unique teachers reporting they strongly disagree (SD), disagree (D), agree (A), or strongly agree (SA) that "The school improvement program in this school requires me to make major changes in my classroom practice" in their first Teacher Questionnaire.**

Table 2
Variation in Magnitude of Change: America's Choice

Variable		Mean	*SD*	Minimum	Maximum	Observations
Major change	Total	2.695	0.775	1	4	$N = 403$
	Between		0.300	2.250	3.545	$n = 30$
	Within		0.720	0.775	4.401	$T = 13.433$

Table 3
Variation in Magnitude of Change: Success for All

Variable		Mean	*SD*	Minimum	Maximum	Observations
Major change	Total	2.624	0.778	1	4	$N = 864$
	Between		0.276	2.227	3.375	$n = 28$
	Within		0.740	0.249	4.374	$T = 30.857$

Both Table 2 and Table 3 illustrate more than twice as much variation within schools as between schools in the magnitude of change teachers report: 72% and 74% in America's Choice and Success for All, respectively.

Table 4
Fidelity of Implementation

Comprehensive School Reform	Observations	Mean	*SD*	Minimum	Maximum
America's Choice	403	0.3753	0.1511	0	0.8472
Success for All	864	0.4655	0.1729	0	1

This underscores the importance of taking the individual as the unit of analysis rather than assuming consistency within schools.

Fidelity of CSR Implementation

For each program, I constructed measures of fidelity that capture variation in teachers' implementation of its prescriptions (see Appendix A). I measured fidelity using responses to the log questions asking: "To what extent were the following topics a focus of your work with the target student in reading/language arts today?" Four check boxes were provided for each content topic listed, corresponding to answers ranging from *a major focus* to *not taught today* (Study of Instructional Improvement, 2000).

SFA and AC ask teachers to cover slightly different content and also emphasize the content areas they cover differently. Fidelity to these specifications was scored in the following way: Content that the program asked be a central component of each day's instruction was given a score of 1 if marked a major focus. Elements that the program indicated were important but not necessarily a major focus of each day's lesson were scored as 1 if marked touched on briefly or a minor focus and 0.5 if marked a major focus. Each of those elements was scored as zero if teachers reported teaching it with less emphasis than prescribed or if they did not mark a box corresponding to that content area. A weighted average of these measures was then taken, weighting daily prescriptions twice as much as the less central ones.

Summary statistics reflecting the distribution of fidelity of implementation in the sample are shown in Table 4.

Methods

To investigate the relationship between the magnitude of change teachers are asked to enact and the fidelity with which they enact that change, I used the data described previously to conduct a series of multiple regression analyses. Each one estimates the relationship between teachers' reports of the extent to which AC or SFA required major changes in their classroom practice and the fidelity with which they enacted AC or SFA's classroom-level prescriptions for literacy instruction. Each year of Log data for each teacher is

treated as a separate observation (although observations of the same individual over multiple years will share the same magnitude of change value). The estimating equation is as follows:

$$y_{ist} = \beta X_{is} + \gamma A_{st} + \delta E_{ist} + \theta D_{is} + \alpha_s + u_{ist}.$$

The outcome of interest, y_{ist}, is the fidelity of implementation of individual i in school s at time since school CSR adoption t. The X_{is} term represents a set of binary variables corresponding to each possible answer to the magnitude of change self-report. The coefficients on each of these variables, represented collectively by β, are the focus of this analysis.

Recognizing that other characteristics of schools and teachers might be correlated with both program fit and fidelity of implementation and thus might obscure this central relationship, most specifications also include a series of control variables intended to reduce these confounding effects.

To account for school characteristics that might affect the relationship between program fit and fidelity of implementation, some specifications include school fixed effects, with individual school-level intercepts represented by α_s. In effect, these specifications compare teachers to other teachers within the same school while controlling for observed and unobserved differences between the schools, including district-level differences. I also control separately for the number of years passed since the school first adopted the program (range, 1–6) using a set of dummy variables, represented by γA_{st}.

Some specifications also include controls for teacher-level characteristics, consisting of both time-varying characteristics, represented by δE_{ist}, and fixed characteristics, represented by θD_{is}. These include teacher experience, teacher education, employment category (certification and full- vs. part-time), and teacher demographics, including gender and race. They also represent some variables intended to more closely capture individuals' relationship toward their school's CSR. These include a dummy variable for whether teachers worked at the school before the CSR was implemented and responses to a series of questions that address various facets of teachers' attitudes toward the CSR: "I am capable of making the kinds of changes called for by the school improvement program," "The kinds of changes called for by the school improvement program are helping my students reach higher levels of achievement," and "I strongly value the kinds of changes called for by the school improvement program" (Study of Instructional Improvement, 2001, 2002, 2003, 2004).

Results

The results of these regression analyses are represented in Tables 5 and 6. Each column represents a separate regression with different combinations of control variables indicated with Xs. (For tables including coefficients for all control variables, please see Appendix B in the online version of the journal.) The coefficients shown in Tables 5 and 6 reflect the increase in fidelity,

Table 5

Estimated Effects on Fidelity of Implementation: America's Choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CSR requires major change							
Strongly disagree	−0.003	0.027	−0.005	−0.009	−0.008	0.008	0.040
	(0.034)	(0.032)	(0.035)	(0.035)	(0.033)	(0.034)	(0.031)
Disagree	(omitted, reference category)						
Agree	0.016	0.023	0.021	0.023	0.026	0.027	0.020
	(0.019)	(0.017)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)
Strongly agree	0.077**	0.061*	0.081**	0.082**	0.081**	0.069*	0.039
	(0.027)	(0.028)	(0.027)	(0.026)	(0.026)	(0.026)	(0.028)
Controls							
School							
Fixed effects		X					X
Years since adoption		X	X	X	X	X	X
Teacher							
Experience			X	X	X	X	X
Education				X	X	X	X
Employment category				X	X	X	X
Demographics					X	X	X
Present before CSR?						X	X
Attitudes toward CSR						X	X
Observations	403	403	403	403	403	403	403
R^2	0.027	0.025	0.055	0.066	0.082	0.109	0.083
Number of schools		30					30

Note. Robust standard errors in parentheses. CSR = Comprehensive School Reform.
*$p < .05$. **$p < .01$.

Table 6
Estimated Effects on Fidelity of Implementation: Success for All

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CSR requires major change							
Strongly disagree	−0.022	−0.024	−0.023	−0.028	−0.038	−0.038	−0.038
	(0.026)	(0.033)	(0.030)	(0.029)	(0.024)	(0.023)	(0.024)
Disagree	(omitted, reference category)						
Agree	−0.003	0.005	−0.007	−0.009	−0.007	−0.006	0.006
	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.012)
Strongly Agree	0.032	0.037*	0.026	0.026	0.032^{+}	0.031^{+}	0.034*
	(0.023)	(0.018)	(0.021)	(0.021)	(0.018)	(0.018)	(0.015)
Controls							
School							
Fixed effects		X					X
Years since adoption		X	X	X	X	X	X
Teacher							
Experience			X	X	X	X	X
Education				X	X	X	X
Employment category				X	X	X	X
Demographics					X	X	X
Present before CSR?						X	X
Attitudes toward CSR						X	X
Observations	864	864	864	864	864	864	864
R^2	0.006	0.013	0.012	0.018	0.049	0.063	0.046
Number of schools		28					28

Note. Robust standard errors in parentheses. CSR = Comprehensive School Reform.
$+p < .10$. $*p < .05$.

on a scale from 0 to 1, associated with a response of strongly disagree, agree, or strongly agree compared to a response of disagree. If teachers who are asked to make larger changes implement programs with lower fidelity, we should expect to see that the more strongly teachers agree that a CSR requires them to make major changes to their classroom practice, the smaller the coefficient associated with their fidelity of implementation. If teachers who are asked to make larger changes implement programs with greater fidelity, we should expect to see that the more strongly teachers agree that a CSR requires them to make major changes to their classroom practice, the larger the coefficient associated with their fidelity of implementation.

The results for teachers implementing America's Choice are shown in Table 5. Looking across the different specifications, the estimates are not identical, but some patterns emerge. First, the coefficient on the strongly agree response is positive and statistically significant at the .05 level or less in all but one specification, indicating that those who perceived the largest changes also implemented with the greatest fidelity. In addition, even among the coefficients not reaching statistical significance, Models 1, 3, 4, and 5 all show a similar pattern, where the point estimates indicate that fidelity increases monotonically with increasing recognition of a major change.

In Models 2, 6, and 7 the point estimates on strongly disagree are higher than in the other models and higher than for those who disagree or agree (although they still remain statistically insignificant). This could be interpreted as weak support for the prediction that teachers reporting greater changes implement programs with lesser fidelity (although it still remains statistically insignificant). However, even in these specifications, there remains a consistent pattern of increase from disagree through strongly agree, which is statistically significant at the .05 level for strongly agree in Models 2 and 6.

Table 6 presents the same set of relationships for Success for All.

In Table 6, most of the coefficients are not statistically significant. However, as in the America's Choice analysis, there are several positive and statistically significant estimates on the strongly agree response. In Models 2 and 7, the difference between individuals who disagreed (and strongly disagreed) from those who strongly agreed is statistically significant at the .05 level. In Models 5 and 6, this effect is marginally significant at the .1 level.

In addition, although the other coefficients do not reach statistical significance, the point estimates do display a consistent pattern: Those who strongly disagree implement SFA with the lowest fidelity, those who disagree or agree are about the same, and those who strongly agree implement the program with the highest fidelity. In other words, overall, there is a pattern of greater fidelity among those who agree more strongly that the CSR they are asked to implement requires them to make major changes to their classroom practice. Thus, although inconsistent in their precision, the regression coefficients offer modest evidence of a positive relationship between the magnitude of change asked and fidelity of implementation.

Discussion

Overall, the results for America's Choice are stronger than those for Success for All. However, both programs show similar trends. Across all specifications, in both AC and SFA, support for the dominant prediction that teachers asked to make larger changes would implement them with lower fidelity ranged from minimal to none at all. In no case was agreeing more strongly that a CSR required major changes in classroom practice associated with a statistically significant drop in fidelity of implementation. This indicates that traditional policy implementation accounts predicting lower fidelity in cases of greater change are inadequate.

On the contrary, in both programs, several specifications showed statistically significant positive effects for those who strongly agreed that the CSR required major changes compared with those who disagreed with that statement. Across both programs, the coefficient on strongly agree was always positive and reached statistical significance at the .1 level or below for 10 out of the 14 specifications presented. Coefficients on the other responses did not reach statistical significance. However, across both programs, they did indicate a fairly consistent trend toward a positive relationship between magnitude of change and fidelity of implementation. In other words, although the lack of statistical significance suggests that the analysis may have been underpowered, the results suggest that in these cases, practitioners asked to make larger changes generally did so with relatively greater fidelity.

These data would be difficult to interpret under traditional policy implementation accounts focusing on material resources or practitioner resistance as the causes of low fidelity implementation. By contrast, the theories discussed previously linking implementation to processes of conceptual change offer a more plausible explanatory mechanism. Still, the results of this study do not foreclose other possible explanations. Therefore, in the spirit of better understanding the relationship between educational reforms' fit with existing practices and the fidelity with which teachers implement such reforms, it is worth considering multiple explanations for the results presented here.

Conceptual Change

Theories of conceptual change indicate that the conditions under which people are introduced to ambitious new ideas make a difference in whether they are able to overcome the challenge of substantially changing their understandings. Teachers who do not experience a change as sufficiently salient or perceive its implications for their own practice are unlikely to engage with that idea in a transformative way (Gregoire, 2003). As prior studies of implementation have demonstrated, a lack of deep understanding of ambitious reforms can lead to lowered fidelity (e.g., Spillane, 2004). If the magnitude of change the CSR posed for some of the teachers in this study was too small to provoke cognitive accommodation, we might well expect

to see lowered fidelity for teachers experiencing smaller changes and correspondingly higher fidelity for teachers experiencing larger changes. This is exactly the pattern this analysis revealed. Moreover, we would expect to see this effect most strongly in settings where deep conceptual understanding was especially important to high-fidelity implementation. For example, for a program relying more heavily on teachers' own planning, judgment, and discretion, fidelity of implementation would likely show more sensitivity to a teacher's depth of understanding of the reform and in turn more sensitivity to the magnitude or salience of the change teachers experience. Again, this is exactly the pattern found in these data as America's Choice showed a stronger positive relationship between magnitude of change and fidelity than that found in Success for All. If this explanation is correct, it suggests that districts and schools might do well to choose boldly in selecting potential new programs and policies as larger changes might help teachers implement new practices with greater fidelity.

At the same time, it is important to weigh the possibility of increased fidelity against the other contextual factors influencing implementation, which years of research have demonstrated are manifold (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010; Honig, 2006; Huberman & Miles, 1984). For example, both AC and SFA were well-specified, well-funded programs; individual teachers' fidelity of implementation as measured here took place within a larger school context that included substantial training and hiring of additional support staff. Future research might do well to examine the external validity of this finding, in particular considering these external factors, by looking across a wider range of programs.

Perception

An alternative explanation, also plausible, turns the direction of causality described previously on its head. Given that the data used to measure the magnitude of change being asked of teachers are drawn from self-reports, one might argue that they primarily reflect teachers' perceptions or understanding of the magnitude of change a program requires rather than the "true" magnitude or degree of difference between teachers' prior practices and those espoused by the program. If the measure of fit I use is mostly capturing perception and understanding, it might be that the type of teacher who more accurately perceives how ambitious a program is is also more likely to implement it with high fidelity. Or, the measurement may even be capturing the *effects* of conceptual change rather than its causes. Especially for a CSR like SFA that provides very detailed prescriptions, it might be relatively straightforward to implement the program with high fidelity even without a deep conceptual understanding of the distinctions between new and old practices. In this case, a positive statistical relationship could appear not because asking for larger changes leads to higher fidelity

but because high fidelity retrospectively leads people to recognize the magnitude of change they have enacted. If true, this would mean that these analyses don't tell us much about the relationship between "true" program fit and fidelity of implementation.

Without an experimental research design, this ambiguity can't be clarified with certainty. If it were the case that big changes in practice subsequently lead to the perception of large changes having been asked, we would expect that individuals' perceptions of the magnitude of change asked of them would increase after implementation. In general, however, when the same individual answered the Teacher Questionnaire in more than one year, their report of the magnitude of change tended to fall over time (see Figure 1).

Still, future research might address the possibility that features relating to the framing or salience of changes—how perceptible they are or how clearly they implicate teachers' existing practice independent of "true" magnitude—might be especially important in predicting fidelity of implementation. If true, this would have meaningful implications for practice as it suggests that manipulating relatively superficial characteristics might be a disproportionately powerful lever. While school leaders may or may not be able to choose which policies their schools must abide by, they have greater discretion about how new policies are framed in relation to existing practices.

In addition, future research and theorizing on this issue might explore the notion that the relationship between perceived magnitude of change and fidelity of implementation may not be strictly unidirectional. Perceiving and enacting larger changes may go hand in hand, mutually reinforcing one another. Longitudinal research designs using repeated interviews of the same individuals over time would help to uncover such a pattern.

Other Explanations

A third type of explanation is that the statistical relationship between constructs may be real but driven by a different mechanism than the one I have theorized. That is, there may be a real predictive positive relationship between the magnitude of change AC or SFA requires of a teacher and the fidelity with which he or she implements the program, but it may not be driven by processes of conceptual change. For example, perhaps there is something about the kind of teacher who initially teaches very differently from AC or SFA that nonetheless makes him or her more likely to fully implement a new program. Or, perhaps the relationship between measures of magnitude of change and fidelity are largely incidental and dwarfed by an independent trend in how fidelity changes over time.[4] If true, these explanations leave unclear whether magnitude of change is a useful feature to consider in program selection or not.

Without an experimental design, it is impossible to know for certain whether unobserved differences between teachers might be driving the

observed effects. Many of the specifications I used include a large number of teacher-level control variables. However, as Datnow and Castellano (2000a) found, differences in implementation are not always correlated with readily measurable teacher characteristics but rather may stem from differences in pedagogical philosophy or other harder to capture facets. Most specifications also include control variables for each year of program implementation. These coefficients actually indicate that fidelity of implementation tended to rise over time for America's Choice teachers while falling for Success for All teachers (see Appendix B in the online version of the journal). This suggest that the relationship between magnitude of change and fidelity is insensitive to the difference between increasing and decreasing time trends in overall fidelity. Still, the best approach to establishing the robustness of this relationship is in future research testing for its presence in a variety of settings.

Conclusions

As schools and districts continue to explore additional models of reform, it is important for leaders and policymakers to be well informed about factors that are likely to affect the fidelity with which such reforms are implemented. A long history of implementation research has pointed to the importance of frontline workers in enacting such reforms. One important factor that has not received sufficient attention to date is how the amount of change that those frontline workers are asked to undertake is related to how fully and faithfully they implement the program. In this paper, I demonstrate that traditional policy implementation accounts suggesting that larger changes will be implemented with lesser fidelity are too simple. Instead, I argue that for reforms that ask teachers to think about their work in significantly different ways, the salience of very large changes may actually help teachers shift their thinking in ways that promote high-fidelity implementation. More moderate changes may be easier to implement in some senses but not necessarily with high fidelity as they may be more readily subsumed into existing ideas, leading to "new" practices that are only superficially different from the old. This account is supported by data on the implementation of comprehensive school reform programs, where by and large, those teachers who reported being required to make larger changes also implemented the reforms with higher fidelity, especially in the program that requires more interpretation to implement.

From a practical standpoint, it is important to acknowledge that processes of conceptual change operate alongside a whole host of other factors that are likely to affect fidelity. I do not wish to claim that resource allocation or individuals' resistance, for example, play no role at all in how programs and policies are implemented. However, I argue that the significance of cognitive mechanisms has been underestimated. This analysis offers a step forward in building theory in this area and points toward important areas of future research.

Appendix A

I calculated fidelity of implementation each year for each teacher by comparing the elements of literacy instruction that they reported using, with those recommended by their Comprehensive School Reform (CSR).

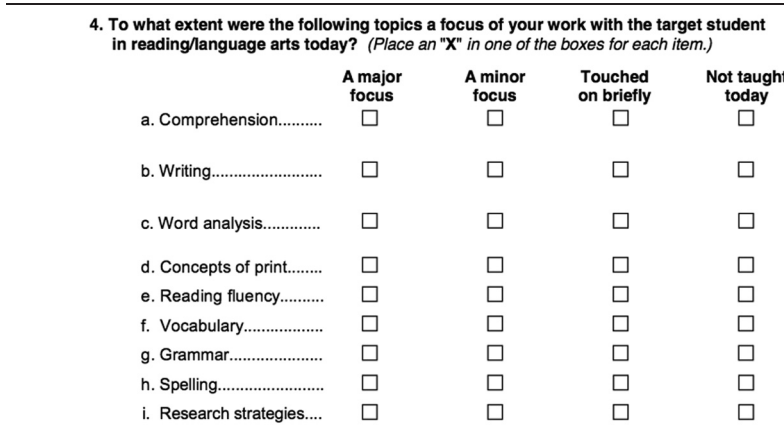
Teachers were asked about their literacy instruction in the teacher log using the prompt shown in Figure A1.

Both Success for All and America's Choice recommend slightly different program elements for younger and older students, so each teacher's reported practice was compared to the program prescriptions corresponding to their CSR and students' grade level.

To calculate fidelity, each content element received a score of 1 if the teacher reported focusing on it to a degree that aligned with their CSR program's recommendations as reflected in Table A1, a score of .5 if they reported a greater focus than called for, and a score of 0 otherwise, including for missing values. Fidelity scores ranged from 0 to 1 as a weighted average of the scores for each content element. Elements that the CSR emphasized (i.e., where the program asked for a "major" focus) were weighted more heavily. Instruction in content elements not included in the prescriptions for a given teacher's program and grade level did not affect that teacher's fidelity score.

Table A1
Measures of Fidelity

	Success for All		America's Choice	
	Grades K–1 (Reading Roots)	Grades 2–5 (Reading Wings)	Grades K–3	Grades 4–5
Comprehension	Major	Major	Major	Major
Writing		Minor/brief	Major	Major
Word analysis	Major		Minor/brief	
Concepts of print	Minor/brief			
Reading fluency	Minor/brief	Minor/brief	Minor/brief	Minor/brief
Vocabulary		Minor/brief	Minor/brief	Minor/brief
Grammar				
Spelling				
Research strategies				



4. To what extent were the following topics a focus of your work with the target student in reading/language arts today? *(Place an* **"X"** *in one of the boxes for each item.)*

	A major focus	A minor focus	Touched on briefly	Not taught today
a. Comprehension.....	☐	☐	☐	☐
b. Writing.....	☐	☐	☐	☐
c. Word analysis.....	☐	☐	☐	☐
d. Concepts of print.....	☐	☐	☐	☐
e. Reading fluency.....	☐	☐	☐	☐
f. Vocabulary.....	☐	☐	☐	☐
g. Grammar.....	☐	☐	☐	☐
h. Spelling.....	☐	☐	☐	☐
i. Research strategies....	☐	☐	☐	☐

Figure A1. **Language Arts Log item used to capture teachers' degree of focus on each element of literacy instruction.**

Notes

Many thanks to Jonathan Guryan and Jeannette Colyvas for their extensive support and assistance throughout the development of this manuscript. Thank you as well to Jim Spillane, Cynthia Coburn, and four anonymous reviewers who each provided feedback on earlier drafts. Earlier formulations of this paper were also presented at the Annual Meeting of the American Educational Research Association and the Structuring Work in and around Organizations Workshop. This research was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B080027 to Northwestern University. The opinions expressed are those of the author's and do not represent views of the Institute or the U.S. Department of Education. Finally, thanks to Deborah Ball, David Cohen, and Brian Rowan, who made data from the Study for Instructional Improvement publicly available, making these analyses possible.

[1]Coburn actually uses the terms *assimilation* and *accommodation* (among others) to characterize the changes teachers make to their classroom practices. These terms directly reference processes of cognitive change, underscoring the importance of teacher thinking in implementation processes. Nonetheless, since for the remainder of the article, I use the terms *assimilation* and *accommodation* to refer exclusively to cognitive processes, for the sake of clarity I use the simplified language of *minor* and *major* changes to characterize Coburn's categories of practice change here instead of her original terms.

[2]The Study for Instructional Improvement also includes data on a third Comprehensive School Reform, Accelerated Schools Project; however, the program specifications for this model were not sufficiently concrete to measure fidelity.

[3]Mathematics Logs were also administered, but these were not used in the current analysis.

[4]Thanks to Anonymous Reviewer 1 for this point.

References

Ansari, S. M., Fiss, P. C., & Zajac, E. J. (2010). Made to fit: How practices vary as they diffuse. *Academy of Management Review*, *35*(1), 67–92.

Berends, M. (2000). Teacher-reported effects of New American School designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, *22*(1), 65–82.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive School Reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125–230. doi:10.3102/00346543073002125

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education*, *68*(4), 493–503.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Carey, S. (1988). Reorganization of knowledge in the course of acquisition. *Ontogeny, Phylogeny, and Historical Development*, *2*, 1.

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. doi:10.1177/1098214010366173

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, *14*(2), 161–199.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*(3), 211–244.

Coburn, C. E., & Stein, M. K. (2010). Key lessons about the relationship between research and practice. In C. E. Coburn & M. K. Stein (Eds.), *Research and practice in education: Building alliances, bridging the divide* (pp. 201–226). Lanham, MD: Rowman & Littlefield.

Cohen, D. K. (1988). *Teaching practice: Plus ça change*. Ann Arbor, MI: Institute for Research on Teaching, College of Education, Michigan State University.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, *12*(3), 311–329.

Cohen, D. K., & Moffitt, S. L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Pres.

Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, *113*(4), 515–548. doi:10.1086/518487

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*, 128–152.

Comprehensive School Reform Quality Center. (2005). *CSRQ Center report on elementary school comprehensive school reform models*. Washington, DC: Author.

Corcoran, T., Hoppe, M., Luhm, T., & Supovitz, J. A. (2000). *America's Choice comprehensive school reform design first-year implementation evaluation summary*. Philadelphia, PA: Consortium for Policy Research in Education.

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103–124). Washington, DC: American Psychological Association.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three Comprehensive School Reform programs. *American Educational Research Journal*, *44*(2), 298–339. doi:10.3102/0002831207302501

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.

Datnow, A., & Castellano, M. (2000a). *An "inside" look at Success for All: A qualitative study of implementation and teaching and learning*. Retrieved from http://www.jhucsos.com/wp-content/uploads/2016/04/Report45.pdf

Datnow, A., & Castellano, M. (2000b). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal*, *37*(3), 775–799. doi:10.2307/1163489

Desimone, L. (2002). How can Comprehensive School Reform models be successfully implemented? *Review of Educational Research*, *72*(3), 433–479.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2/3), 105–225.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. New York, NY: Harcourt Brace Jovanovich College Publishers.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, *15*(5), 451–474. doi:10.1016/0091-7435(86)90024-1

Fosnot, C. T. (1996). *Constructivism: Theory, perspectives, and practice*. New York, NY: Teachers College Press.

Fullan, M. (1983). Evaluating program implementation: What can be learned from follow through. *Curriculum Inquiry*, *13*(2), 215–227.

Glazer, J. L. (2009). How external interveners leverage large-scale change: The case of America's Choice, 1998-2003. *Educational Evaluation and Policy Analysis*, *31*(3), 269–297. doi:10.3102/0162373709336745

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*(2), 147–179.

Guskey, T. R. (1991). Enhancing the effectiveness of professional development programs. *Journal of Educational & Psychological Consultation*, *2*(3), 239–247.

Hallett, T. (2010). The myth incarnate: Recoupling processes, turmoil, and inhabited institutions in an urban elementary school. *American Sociological Review*, *75*(1), 52–74.

Honig, M. I. (2006). Complexity and policy implementation: Challenge and opportunities for the field. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1–25). Albany, NY: SUNY Press.

Huberman, A. M., & Miles, M. B. (1984). *Innovation up close: How school improvement works*. New York, NY: Plenum Press.

Lewis, L. K., & Seibold, D. R. (1993). Innovation modification during intraorganizational adoption. *Academy of Management Review*, *18*, 322–354.

Lin, A. C. (2002). *Reform in the making: The implementation of social policy in prison*. Princeton, NJ: Princeton University Press.

Linder, S. H., & Peters, B. G. (1987). A design perspective on policy implementation: The fallacies of misplaced prescription. *Review of Policy Research*, *6*(3), 459–475. doi:10.1111/j.1541-1338.1987.tb00761.x

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology*, *51*(1), 345–375.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York, NY: Russell Sage Foundation Publications.

Majone, G., & Wildavsky, A. (1979). Implementation as evolution. In J. L. Pressman & A. Wildavsky (Eds.), *Implementation* (3rd ed., pp. 163–180). Berkeley, CA: University of California Press.

Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, *5*(2), 145–174.

Mazmanian, D. A., & Sabatier, P. A. (1983). *Implementation and public policy*. Glenview, IL: Scott Foresman.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, *9*(2), 171–178.

Mehta, J. (2015). *The allure of order: High hopes, dashed expectations, and the troubled quest to remake American schooling*. Oxford, UK: Oxford University Press.

National Center on Education and the Economy, & University of Pittsburgh. (1997). *New standards performance standards* (Vol. 1). Washington, DC: Author.

Orland, M., Hoffman, A., & Vaughn, E. S. (2010). *Evaluation of the Comprehensive School Reform program implementation and outcomes: Fifth-year report*. Washington, DC: U.S. Department of Education Office of Planning, Evaluation and Policy Development Policy and Program Studies Service.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal*, *105*(1), 31–48. doi:10.1086/428764

Piaget, J. (1977). *The essential Piaget*. New York, NY: Basic Books.

Poglinco, S. M., Bach, A. J., Hovde, K., Rosenblum, S., Saunders, M., & Supovitz, J. A. (2003). *The heart of the matter: The coaching model in America's Choice schools*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227.

Rowan, B., Camburn, E., & Barnes, C. (2004). Benefiting from comprehensive school reform: A review of research on CSR implementation. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1–52). Washington, DC: National Clearinghouse for Comprehensive School Reform.

Rowan, B., Correnti, R., Miller, R. J., & Camburn, E. M. (2009). *School improvement by design: Lessons from a study of Comprehensive School Reform programs*. Philadelphia, PA: Consortium for Policy Research in Education.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. *Theoretical Issues in Reading Comprehension*, *1*, 33–58.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scott, W. R., & Meyer, J. W. (1983). *The organization of societal sectors*. London: Elsevier.

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

Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: Investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies*, *41*(4), 467–500.

Slavin, R. E., & Madden, N. (1999). *Disseminating Success for All: Lessons for policy and practice* (No. 30). Baltimore, MD: Center for Research on the Education of Students Placed at Risk.

Slavin, R. E., & Madden, N. A. (2000). *One million children: Success for All*. Thousand Oaks, CA: Corwin Press.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S., Smith, L., & Dianda, M. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, *1*(1), 41–76.

Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy* (Vol. 43). Cambridge, MA: Harvard University Press.

Spillane, J. P., & Burch, P. (2006). The institutional environment and instructional practice: Changing patterns of guidance and control in public education. In H.-D. Meyer & B. Rowan (Eds.), *The new institutionalism in education* (pp. 87–102). Albany, NY: SUNY Press.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147–176). Albany, NY: SUNY Press.

Study of Instructional Improvement. (2000). *Language Arts Log*. Retrieved from http://www.sii.soe.umich.edu/data/

Study of Instructional Improvement. (2001). *Teacher Questionnaire 2000–2001*. Retrieved from http://www.sii.soe.umich.edu/data/

Study of Instructional Improvement. (2002). *Teacher Questionnaire 2001–2002*. Retrieved from http://www.sii.soe.umich.edu/data/

Study of Instructional Improvement. (2003). *Teacher Questionnaire 2002–2003*. Retrieved from http://www.sii.soe.umich.edu/data/

Study of Instructional Improvement. (2004). *Teacher Questionnaire 2003–2004*. Retrieved from http://www.sii.soe.umich.edu/data/

Supovitz, J. A., Poglinco, S. M., & Bach, A. (2002). *Implementation of the America's Choice literacy workshops*. Philadelphia, PA: Consortium for Policy Research in Education.

Supovitz, J. A., Poglinco, S. M., & Snyder, B. A. (2001). *Moving mountains: Successes and challenges of the America's Choice comprehensive school reform design*. Philadelphia, PA: Consortium for Policy Research in Education.

Van Meter, D. S., & Van Horn, C. E. (1975). The policy implementation process. *Administration & Society*, *6*(4), 445–488. doi:10.1177/009539977500600404

Manuscript received October 5, 2015
Final revision received May 7, 2017
Accepted May 10, 2017