

Review of Educational Research

<http://rer.aera.net>

Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research

Carol L. O'Donnell

REVIEW OF EDUCATIONAL RESEARCH 2008 78: 33

DOI: 10.3102/0034654307313793

The online version of this article can be found at:

<http://rer.sagepub.com/content/78/1/33>

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Review of Educational Research* can be found at:

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

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

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

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

>> [Version of Record](#) - Feb 25, 2008

[What is This?](#)

Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research

Carol L. O'Donnell
The George Washington University

Education researchers are being asked to conduct rigorous, scientifically based studies of K–12 curriculum interventions; therefore, the need for measuring fidelity of implementation and empirically relating it to outcomes (the chief rationale for this review) is warranted to ensure internal and external validity. The results of this review indicate that there are too few studies to guide researchers on how fidelity of implementation to core curriculum interventions can be measured and related to outcomes, particularly within efficacy and effectiveness studies, where the requirements for fidelity measures differ. This review attempts to clarify the definition, conceptualization, and measurement of fidelity of implementation and to guide future researchers in understanding how fidelity of implementation can be used to adjust or interpret outcome measures.

KEYWORDS: curriculum studies, effectiveness, fidelity of implementation, program evaluation, program integrity.

Fidelity is central to the validity of any intervention study and is closely related to the statistical power of outcome analyses. . . . Failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation.

—Dumas, Lynch, Laughlin, Smith, & Prinz, 2001, p. 39

The bridge between a promising idea and the impact on students is implementation, but innovations are seldom implemented as intended.

—Berman & McLaughlin, 1976, p. 349

Fidelity of implementation is traditionally defined as the determination of how well an intervention is implemented in comparison with the original program design during an efficacy and/or effectiveness study (Mihalic, 2002; cf. Berman & McLaughlin, 1976; Biglan & Taylor, 2000; Freeman, 1977; Fullan, 2001; Hord,

Rutherford, Huling-Austin, & Hall, 1987; Lipsey, 1999; National Research Council [NRC], 2004; Patton, 1978; Scheirer & Rezmovic, 1983; U.S. Department of Education, 2006). More specifically, it is “the extent to which the user’s current practice matche[s] the . . . ‘ideal’” (Loucks, 1983, p. 4). Fidelity of implementation is a relatively recent construct in K–12 curriculum intervention research, but its use in program evaluation dates back 30 to 35 years (Mowbray, Holter, Teague, & Bybee, 2003; Sechrest, West, Phillips, Redner, & Yeaton, 1979).

A review of the public health literature (cf. Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003) indicates that there are five criteria for measuring fidelity of implementation: (a) adherence—whether the components of the intervention are being delivered as designed; (b) duration—the number, length, or frequency of sessions implemented; (c) quality of delivery—the manner in which the implementer delivers the program using the techniques, processes, or methods prescribed; (d) participant responsiveness—the extent to which participants are engaged by and involved in the activities and content of the program; and (e) program differentiation—whether critical features that distinguish the program from the comparison condition are present or absent during implementation. These criteria can be considered divided into two groups (Mowbray et al., 2003): fidelity to structure (i.e., adherence, duration) and fidelity to process (i.e., quality of delivery, program differentiation), with participant responsiveness taking on characteristics of both (A. Ruiz-Primo, personal communication, June 2004). Hallmark studies in the health field have underscored the importance of measuring both the structure and process of implementation during efficacy and effectiveness studies (Connell, Turner, & Mason, 1985; Resnicow et al., 1992; Taggart, Bush, Zuckerman, & Theiss, 1990), and they provide evidence that variables such as fidelity to quality of implementation account for differential outcomes (Hopkins, Mauss, Kearney, & Weisheit, 1988; Mitchel, Hu, McDonnell, & Swisher, 1984; Tortu & Botvin, 1989).

Until the 1970s, there wasn’t really a need to study fidelity of implementation. Researchers assumed fidelity of implementation would be high during program adoption and that implementers would copy or imitate the innovation exactly as earlier adopters had used it (Rogers, 2003). This assumption was made because adopters were “considered to be rather passive acceptors of an innovation, rather than active modifiers of a new idea” (Rogers, 2003, p. 180). However, once researchers recognized that fidelity of implementation was not always a given and that adopters would adapt an innovation to suit their local needs, “they began to find that quite a lot of it occurred” (Rogers, 2003, p. 180).

Although seemingly well defined in the health literature (cf. Hansen, Graham, Wolkenstein, & Rohrbach, 1991; Kolbe & Iverson, 1981), fidelity of implementation is rarely reported in large-scale education studies that examine the effectiveness of K–12 core curriculum interventions, especially with regard to how fidelity enhances or constrains the effects of the intervention on outcomes (L. D. Dobson & Cook, 1980; NRC, 2004; U.S. Department of Education, 2006). Moreover, according to the NRC (2004), even less seldom is such a measure of fidelity to K–12 curriculum interventions used to adjust for or interpret outcome measures. For example, in an evaluation of the quality of K–12 mathematics curriculum material evaluations, the NRC reported that 33 of the 63 “*at least minimally*

methodologically adequate” comparative studies reviewed by the NRC did not report any measure of fidelity, and of the 30 studies that did measure fidelity, only 1 reported and adjusted for it when interpreting its outcome measures (NRC, 2004, p. 115).

Today, in an era of accountability, the call for measuring fidelity of implementation to K–12 core curriculum interventions during efficacy or effectiveness studies is receiving increased attention (NRC, 2004; Towne, Wise, & Winters, 2005; U.S. Department of Education, 2003a, 2006). For example, federal mandates such as the No Child Left Behind Act (2001) have stated that teachers must use only research-based teaching methods and select programs “proven” to be effective (Slavin, 2003). Recent requests for proposals—such as from the Institute of Education Sciences, the research arm of the U.S. Department of Education established by the Education Sciences Reform Act of 2002—ask that researchers not only study the efficacy and effectiveness of curriculum interventions, but that they expand knowledge on the conditions of education practices that improve academic achievement. The What Works Clearinghouse (U.S. Department of Education, 2003b) identified several features of intervention research designs that they stated would improve confidence in findings, including (a) the use of random assignment at the level of the unit of analysis, (b) evidence of assessing fidelity of implementation to the treatment, and (c) use of standardized measurements. In addition, research designs should permit the identification and assessment of factors affecting fidelity of implementation, including considering its effects as a mediating or moderating variable (U.S. Department of Education, 2006). Without methodological consideration of the level of fidelity during implementation, researchers may have insufficient evidence to support the internal validity of an efficacy or effectiveness study (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001).

Purpose and Organization

The purpose of this literature review is to identify studies that have used quantitative research methods to determine the relationship between fidelity of implementation to K–12 core curriculum interventions and outcomes. In 1977, Fullan and Pomfret conducted a review of the research literature on curriculum implementation. At that time, they attempted to define the general construct of implementation, addressed why studying it was important, reviewed how researchers measured implementation, and considered determinants. Therefore, building on the organization of Fullan and Pomfret’s review, the goals of this article are to (a) define fidelity of implementation as it relates to K–12 core curriculum interventions, (b) conceptualize fidelity of implementation within efficacy and effectiveness studies by providing a rationale for studying it and identifying any overlaps and disparities it may have with other similar educational constructs, (c) identify a set of studies that quantitatively measured the relationship between fidelity of implementation to K–12 core curriculum interventions and outcomes, and (d) apply these findings to future K–12 curriculum intervention research. Therefore, the purpose of this literature review is to examine how fidelity of implementation is defined, conceptualized, measured, and applied, each referring to a different section of the Results. To begin, methods for conducting the review are described.

Method

Indices

To identify K–12 studies that have examined the relationship between fidelity of implementation to core curriculum interventions and outcomes, a myriad of indices and methods were used. First, the Web search engine Google Scholar was used to assess an overall return on the exact phrase *fidelity of implementation* and to ascertain a general sense of the number of times each article had been cited by another author. This process yielded a series of 29 articles in the social sciences in which the term fidelity of implementation was used in the title; most of these were prevention studies in the public health field. Three articles synthesizing previous fidelity studies (Dane & Schneider, 1998; Dusenbury et al., 2003; McGrew, Bond, Dietzen, & Salyers, 1994) were the most cited. References within these 3 articles then yielded the first level of primary studies, for a total of 37 articles.

Indices to the scholarly literature in the area of curriculum evaluation were then used to widen the search beyond the area of public health. This included searching ERIC for education and PsycINFO. To obtain a broader and deeper coverage of the literature, citations and abstracts obtained from ERIC and PsycINFO were then used to obtain full-text journal articles and books at the local university library. (Keywords searched are listed in the next paragraph.) ALADIN's link to Ovid and ProQuest allowed for searching in a database interface to which the local university library and members of the local research library consortium subscribe. One search provided access to 36 databases, including those listed above, and a simultaneous keyword, author, title, or journal search for up to five databases. In addition to Ovid's own database of journals, JSTOR Online was also used to access full-text articles using the following search terms.

A variety of other search tools were also useful for locating manuscripts and other texts, including the Library of Congress Catalogue, the local university library catalogue, and the National Academies Press online. This resulted in a set of books addressing larger conceptual issues related to fidelity of implementation, including books on program evaluation (Weiss, 1998), teacher change (Fullan, 2001; Rogers, 2003), and curriculum materials evaluation (NRC, 2004).

Keywords Searched

The keywords searched included *program integrity*, *fidelity of implementation*, *fidelity implementation* (without the "of"), *curricular integrity*, *treatment integrity*, *enactment*, *treatment fidelity*, *implementation variation*, *intended curriculum*, *integrity verification*, and a combination of these terms in conjunction with *curriculum*, *intervention*, or *effectiveness*. The first and second terms resulted in the highest yields (239 and 128 hits, respectively). Other topics tangentially related to fidelity of implementation were used (such as adherence and program implementation), but both yielded results that were too broad (hits above the 600 range). Combinations such as *fidelity*, *intervention*, and *curriculum* together yielded limited but narrowed hits.

Literature on fidelity of implementation is not voluminous (yielding only approximately 133 studies overall in ProQuest and fewer in EBSCOhost, most in the years from 1970 to the present); therefore, there was little selectivity at this stage. All studies under the category of fidelity of implementation, including those in the

public and mental health fields, initially proved to be applicable to the focus of the proposed review. The majority of articles, however, fell into the public and mental health fields. For example, the articles by Dusenbury and colleagues (2003) and Dane and Schneider (1998) have been frequently cited in research literature on fidelity of implementation—about 30 citations combined; however, all but one of these citations are from the health field (outside of K–12 education), the exception being a teacher education program on drug abuse prevention. The use of these keywords, however, even outside of the field of K–12 programs, did inform the review in that they provided methodological support for understanding how the latent construct of fidelity of implementation has been defined, conceptualized, and measured.

Instrument Search

The title of instruments and other tools used to measure fidelity of implementation that were referenced in the literature were then typed into the *Mental Measurements Yearbook* index. This allowed a cross-reference to instruments, inventories, and questionnaires that had been used in the field to measure fidelity. Because fidelity of implementation is a relatively new construct (in use for the past 35 years or so), the UMI Dissertation Abstracts index was then used to locate the instruments and to examine more recent research (1997 to 2006) in the field. This source provided links to more than 25 citations of North American doctoral dissertations available in PDF. Dissertations from 1970 to 1996 were reviewed directly. Limitations, however, arose when instruments were removed from the dissertation for copyright purposes. In this case, the university or researcher was contacted directly for copies of the instrument.

Criteria Development and Article Selection

In the end, more than 120 articles, books, dissertations, and other publications were carefully read and analyzed to better understand how researchers (a) define and (b) conceptualize fidelity of implementation within efficacy and effectiveness studies and to understand any overlaps and disparities that fidelity of implementation may have with other similar educational constructs. Finally, to identify research that (c) quantitatively measured the relationship between fidelity of implementation to K–12 core curriculum interventions and outcomes, 23 primary (empirical) studies of the more than 120 documents reviewed were subjected to the evaluation criteria illustrated in Figure 1. Results from the review are detailed in Appendix A and show that in the end, only 5 of the 23 studies met all of the criteria. These 5 studies quantitatively examined the relationship between fidelity of implementation to core curriculum interventions and outcomes in a K–12 classroom setting. Note that due to varying research designs, sample size (and unit of analysis) was reported for each article reviewed but was not judged as a criterion being met or not met.

Results

Section 1: Defining Fidelity of Implementation

The term *fidelity of implementation* has been defined in various ways, depending on the type of study (e.g., action research, program evaluation, efficacy or effectiveness research, or scale-up) and the field of study (e.g., mental health, public health, or science education). A review of the literature reveals that fidelity of

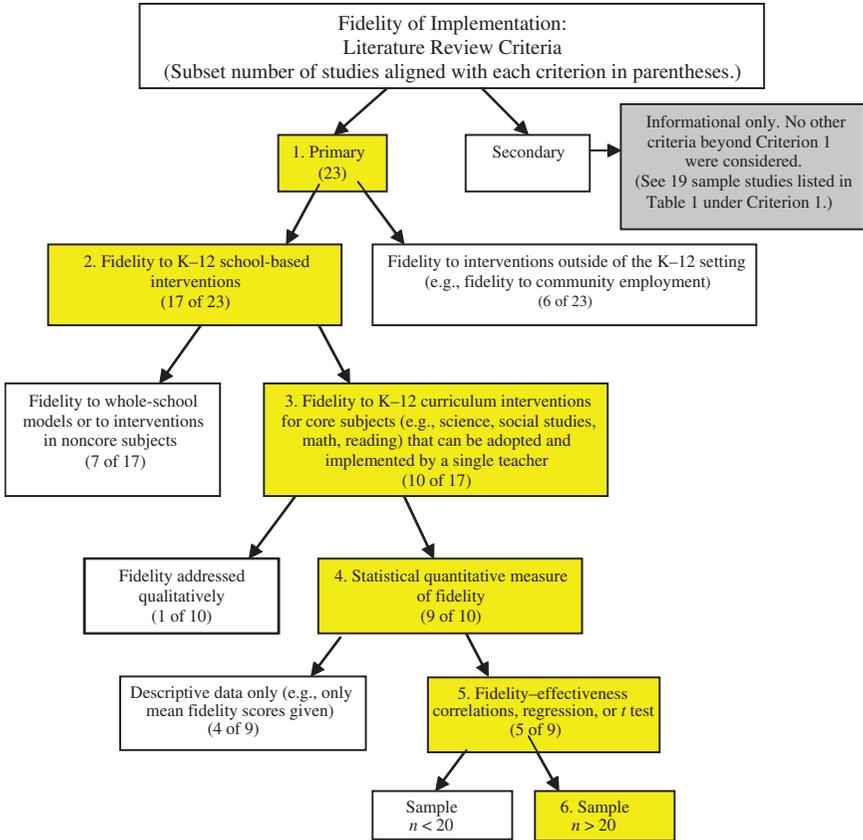


FIGURE 1. *Criteria for identifying studies that measured the relationship between fidelity of implementation to K-12 core curriculum interventions and outcomes.*

implementation has multiple but similar definitions. Because much of the fidelity literature appears to fall into either the health or K-12 core curriculum intervention fields, definitions of fidelity of implementation (or its equivalent synonym) from these two areas of study are examined here.

Definitions of fidelity of implementation to health programs include, but are not limited to, the following:

1. determination of how well the program is being implemented in comparison with the original program design (i.e., the program is being delivered as it was designed and implemented in its efficacy and/or effectiveness trials; Mihalic, 2002).
2. integrity, adherence, or quality of program delivery (Domitrovich & Greenberg, 2000).
3. adherence to the standards of a program model (Bond, Becker, Drake, & Vogler, 1997).

4. “the degree to which programs were implemented as planned” (Dane & Schneider, 1998, p. 23).
5. “the degree to which teachers and other program providers implement programs *as intended by the program developers*” (Dusenbury et al., 2003, p. 240).
6. “closeness between the program-as-planned and the program-as-delivered” (Summerfelt, 2003, p. 56).
7. the degree to which specified procedures are implemented as planned (Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993).

Definitions of fidelity of implementation (or its equivalent synonym) to K–12 core curriculum interventions include, but are not limited to, the following:

1. “the extent to which the project was implemented as proposed (or laid out)” (Loucks, 1983, p. 5).
2. “a measure of the basic extent of use of the curricular materials. It does not address issues of instructional quality. In some studies, implementation fidelity is synonymous with ‘opportunity to learn’” (NRC, 2004, p. 114).
3. “to implement it [an already developed innovation] faithfully in practice—that is, to use it as it is ‘supposed to be used,’ as intended by the developer” (Fullan, 2001, p. 40).
4. “the extent to which the project was implemented as originally planned” (Berman & McLaughlin, 1976, p. 350).
5. the extent to which the program components were implemented (Scheirer & Rezmovic, 1983).
6. the extent to which teachers enact innovations in ways that either follow designers’ intentions or replicate practices developed elsewhere, or the “extent to which the user’s current practice matched the developer’s ‘ideal’” (Loucks, 1983, p. 4).

Overall, the definitions in the health field appear to focus on implementation of program components, and they define the importance of fidelity from an experimental perspective. For example, fidelity of implementation is a means of confirming “that the manipulation of the independent variable occurred as planned” (Moncher & Prinz, 1991, p. 247). However, definitions of fidelity to K–12 core curriculum interventions often reference instructional quality. Loucks, for example, defines fidelity from a teaching perspective by stating fidelity of implementation is “related to the amount of change that occur(s) in the teacher’s practice” (1983, p. 5).

Overall, fidelity of implementation seems to be synonymous with adherence and integrity. For example, citing the work of Gresham et al. (1993), Dane and Schneider state that “integrity is defined as the degree to which specified procedures are implemented as planned,” while “fidelity has been used in this same manner” (1998, p. 23). These same authors then define adherence as “the extent to which specified program components were delivered as prescribed in program manuals” (Dane & Schneider, 1998, p. 45)—a definition nearly identical to that of fidelity of implementation. These inconsistencies make defining (and ultimately measuring) the construct—fidelity of implementation—at times, amorphous.

Fidelity of implementation seems to be well researched in the health field, with some reviews spanning 25 years (Backer, 2000; Dane & Schneider, 1998; Dusenbury et al., 2003; Mihalic, 2002; Remillard, 2005; Resnick et al., 2005; Snyder, Bolin, & Zumwalt, 1992). In most cases, definitions and measures of fidelity of implementation were found not to be consistent across studies. Therefore, reviewers either proposed new definitions for fidelity or suggested that fidelity components (Dane & Schneider, 1998; Dusenbury et al., 2003) and criteria (Mowbray et al., 2003) be established prior to measuring fidelity. Other reviewers suggested guidelines to comprehensively evaluate treatment fidelity (Resnick et al., 2005). These guidelines expanded the traditional definition of fidelity of implementation (i.e., the determination of how well an intervention is implemented in comparison with the original program design; Mihalic, 2002) to also include fidelity to training of interventionists, delivery and receipt of the intervention, and enactment of the intervention in real-life settings (Resnick et al., 2005).

Fidelity of implementation has played a central role in clinical case studies (cf. Bauman, Stein, & Ireys, 1991; K. S. Dobson & Shaw, 1988; McGrew et al., 1994) and health research (Dane & Schneider, 1998; Dusenbury et al., 2003); however, it has been struggling for more than 30 years to find its identity in K–12 core curriculum implementation research (Berman & MacLaughlin, 1976; Fullan & Pomfret, 1977; Loucks, 1983; Snyder et al., 1992). Although there are multiple definitions for fidelity of implementation across each of these fields (and as this article will show later, multiple methods for studying fidelity), it is rarely measured and related to outcomes in K–12 intervention research. Therefore, no consensus seems to exist in the K–12 curriculum intervention literature regarding what exactly fidelity of implementation is, how it is measured, or how program theory or study design (which details how the program will lead to improved outcomes) relates to fidelity of implementation (Ruiz-Primo, 2005; Summerfelt, 2003).

Section 2: Conceptualizing Fidelity of Implementation

To determine how fidelity of implementation is conceptualized in the literature, this section presents a rationale for why studying fidelity of implementation is important in both efficacy and effectiveness studies. Also discussed are the overlaps and disparities that the term *fidelity of implementation* appears to have with other related educational constructs in the literature, including curriculum-in-use, teaching, curriculum potential, and adaptation.

Fidelity of Implementation in Efficacy and Effectiveness Studies

Researchers argue that the statistical power of a study—which depends on reliable and valid measures, appropriate design and sampling, and careful assessment of fidelity—decreases as research moves from the laboratory (efficacy studies) and toward the field (effectiveness studies; Boruch & Gomez, 1977; Hohmann & Shear, 2002; Summerfelt, 2003). Therefore, the need to assess program outcomes as they relate to fidelity of implementation is crucial for both efficacy and effectiveness studies, especially when researchers want to know if poor program outcomes are due to a lack of implementation or due to a poorly designed intervention misaligned with program theory (Summerfelt, 2003).

Efficacy is defined by *Dorland's Illustrated Medical Dictionary* as “the ability of an intervention to produce the desired beneficial effect in expert hands and under

ideal circumstances” (Dorland, 1994, p. 531). Efficacy can be considered to occur temporally as the first stage of program evaluation and refers to the degree to which the program has been shown scientifically to accomplish the desired outcome in extremely favorable conditions, such as in what is equivalent to randomized controlled trials (RCTs; Aron, Raff, & Findling, 1997; Mihalic, 2002; Raudenbush, 2007; Summerfelt & Meltzer, 1998). The purpose of this stage of research is to demonstrate that the program can, under the most favorable conditions, lead to the desired outcomes. Failure to do so would then “give evidence of theory failure, not implementation failure” (Raudenbush, 2007, p. 26).

Interventions with demonstrated benefit in efficacy studies are then transferred into the effectiveness arena. *Effectiveness* is defined by Dorland as “the ability of an intervention to produce the desired beneficial effect in actual use” (1994, p. 531). Effectiveness research complements efficacy research and refers to the degree to which the program achieves its outcome in actual clinical practice or in community or field settings where mediating and moderating factors can be identified (Aron et al., 1997; Mihalic, 2002; Raudenbush, 2007; Summerfelt & Meltzer, 1998). Both efficacy (internal validity) and effectiveness (external validity)—which can be generalized) are complementary, critical aspects of evaluating interventions (Summerfelt & Meltzer, 1998).

What follows is a description of how researchers have conceptualized fidelity of implementation during each of these two stages of intervention research and a rationale for why studying fidelity of implementation to K–12 core curriculum interventions is crucial at each stage.

Efficacy studies. An efficacy study’s examination of fidelity focuses on whether a program is implemented at all (i.e., did the program get delivered?) and to what degree (i.e., what was the program’s quality of delivery?), and it uses the answers to these questions to improve the program. Fidelity of implementation, therefore, is important not only to program evaluators, but also to program developers (Weiss, 1998). Gauging fidelity in an efficacy study may help to determine whether the implementation of the program aligns with the material’s intended program theory (Weiss, 1998). In addition, studying fidelity during efficacy trials may help developers gain an understanding of how the quality of implementation can be improved when research-based programs are disseminated. Gagne and colleagues (Gagne, Wager, Golas, & Keller, 2005), for example, state that the “overall objective of instructional materials is validation” (p. 354). The goal of internal validity is to determine that the program will result in successful achievement of the instructional objectives, providing that the program is “delivered effectively as designed” (Gagne et al., 2005, p. 354). Measuring fidelity of implementation during an efficacy study might entail continuously monitoring and improving the program implementation to ensure it is implemented with the highest fidelity (Resnick et al., 2005).

Other reasons for examining fidelity of implementation during efficacy studies include helping to explain why innovations succeed and fail (Dusenbury et al., 2003). An assessment of fidelity of implementation during efficacy trials allows program developers to guide revisions or outside evaluators to identify and control what has been changed in a program (Resnick et al., 2005). During an efficacy study, implementation researchers may try to determine the critical components of the program (Mowbray et al., 2003). Such an analysis can help future implementers

to understand which features of the program are essential and require the highest level of fidelity, and which may be adapted or deleted (Mowbray et al., 2003)—an issue often raised in the fidelity-adaptation debate (Blakely et al., 1987).

Effectiveness studies. Effectiveness studies seem more interested in interpreting evidence of effectiveness of the program for generalizability (i.e., in comparative evaluations) and observing the implementation of the program in the field (Hohmann & Shear, 2002; NRC, 2004; Raudenbush, 2007). An effectiveness study is not a simple replication of an efficacy study with more subjects and more diverse outcome measures conducted in a naturalistic setting (Hohmann & Shear, 2002). Effectiveness researchers might not focus on monitoring and controlling levels of fidelity; instead, variations in fidelity are measured in a natural setting and then related to student outcomes. If results indicate that the unit was not effective, researchers might ask if the materials “could not be used in the manner intended or that the instructor did not carry out the intended procedures” (Gagne et al., 2005, p. 345).

Because effectiveness studies are concerned with the ability of the program to produce the desired effect in actual use, it seems important to study teachers' fidelity to that program in practice. Fidelity of implementation can reveal important information about the feasibility of how likely an intervention can and will be implemented with fidelity in the classroom. Dusenbury and colleagues state, “If it is difficult to achieve fidelity of implementation in practice, a program has low feasibility. Programs that are implemented with high levels of fidelity but fail to produce desired effects may need to be redesigned” (2003, p. 240). Citing the work of others, they contend that the study of fidelity of implementation is important to gain an understanding of how the quality and extent of implementation can affect program outcomes and to gain confidence that the observed outcomes can be attributed to the intervention (i.e., that positive results are due to the program). Without a measure of fidelity of implementation in an effectiveness study, researchers may not be able to account for negative or ambiguous findings, nor determine whether unsuccessful outcomes are due to an ineffective program or due to failure to implement the program and its conceptual and methodological underpinnings as intended (L. D. Dobson & Cook, 1980; Forgatch, Patterson, & DeGarmo, 2005; Hohmann & Shear, 2002).

Although the most practical reason to study fidelity of implementation during efficacy studies is to monitor programs to help assure quality, measures of fidelity during effectiveness studies promote external validity by creating a “manipulation check in treatment effectiveness research” (Mowbray et al., 2003, p. 317). Researchers claim that well-developed and valid measures of fidelity of implementation can also enhance statistical power in outcome studies (Mowbray et al., 2003) and provide a basis for excluding data from any sites that deviate too far from the experimental treatment model (Teague, Drake, & Ackerson, 1995).

Finally, fidelity measures during an effectiveness study have implications for *scale-up*—defined as the deliberate expansion of an externally developed program that has been previously proven efficacious in one or a small number of school settings to many settings (cf. McDonald, Keesler, Kaufman, & Schneider, 2006; Stringfield & Datnow, 1998). Quadriform analysis, which allows two-dimensional relations between input and output to be graphically viewed (Duyar, 2006), can be used to show that when an intervention is implemented with fidelity and outcomes

Positive outcomes (high achievement)	Do not scale up	Scale up
Negative outcomes (low achievement)	Do not scale up	Do not scale up
	Low fidelity	High fidelity

FIGURE 2. *Scale-up and fidelity during an effectiveness study.*

are high, it seems reasonable that the unit could and should be moved to wide-scale adoption. However, when there is failure to put the intervention into practice as designed (Freeman, 1977; Patton, 1978) and fidelity is low, results—whether positive or negative—cannot be attributed to the intervention, and data suggesting that it succeeded or failed to have an effect become insignificant at a practical level (Yeaton & Sechrest, 1981). Scale-up in this situation would be unproductive (see Figure 2).

Summary. Overall, the conceptualization of fidelity of implementation appears to fall into two stages of research: efficacy studies and effectiveness studies. Understanding to what extent a program is implemented can help researchers test their assumptions that the failure of an evaluated program is either the result of a material’s poor program design (and, therefore, the program needs to be revised so that implementation is more practical), or the program design works, but there is evidence of failed implementation (Forgatch et al., 2005; Resnick et al., 2005). Therefore, the methods used to study fidelity depend on each particular stage.

Although it is clear that studying fidelity of implementation in both efficacy and effectiveness studies has and will become a crucial part of K–12 curriculum intervention research, the actual conceptualization of fidelity of implementation in K–12 settings is confounded by other related and similar, but at times opposing, educational constructs.

Overlaps and Disparities With Other Applicable Constructs in the Literature

The most challenging aspect of reviewing fidelity of implementation literature in K–12 settings is the conflict and overlap that the term *fidelity of implementation* seems to have with other educational constructs. These constructs include teaching (Shulman, 1990); curriculum potential (Ben-Peretz, 1975); curriculum-in-use (Munby & Russell, 1990; Shkedi, 1998), curriculum use (Remillard, 2005), or perceived curriculum (Ben-Peretz, Katz, & Silberstein, 1982; Shkedi, 1998); and adaptation (Cho, 1998; Blakely et al., 1987; Hall & Loucks, 1978). This section examines such overlaps and disparities that exist in the education literature.

Teaching. A review of fidelity of implementation to curriculum materials reveals its conflicts with the construct of teaching. Lee Shulman, for example, in his foreword

to Miriam Ben-Peretz's book *The Teacher–Curriculum Encounter: Freeing Teachers From the Tyranny of Texts*, states, “curriculum and teaching have long been treated as opposites, akin to hot and cold, war and peace, or sadness and joy” (Shulman, 1990, p. vii). Shulman argues that on one hand, curriculum, in the form of written materials, historically manifests itself as carefully organized, concrete, rigid, and well-planned units of instruction; whereas teaching, on the other hand, is interactive, natural, and unstructured. Although Shulman portrays teaching as often planned, he explains that it is more commonly adaptive and reactive. He prepares the reader for Ben-Peretz's work by stating, “While curriculum might be the backdrop for teaching, the two are not to be confused” (Shulman, 1990, p. vii). The challenge for educational researchers then becomes how to distinguish between good teaching and fidelity of implementation to good teaching practices prompted by the curriculum materials (cf. O'Donnell, 2007, for an empirical study that examines this question).

Curriculum potential. In 1975, Ben-Peretz (p. 158) wrote of curriculum potential as she described curriculum materials as “sources for new interpretation.” Her view attacks the opposition of curriculum and instruction and points to the ways the two ideas can be mutually supportive and reinforcing. In line with most researchers of that time (Fox, 1977; McLaughlin & Marsh, 1978), Ben-Peretz and others wrote about how external experts were not capable of introducing curricular innovations into the schools without seriously involving the teachers.

Curriculum-in-use. While curriculum-in-theory is regarded as “that which is written by external experts describing what is to be taught” (Shkedi, 1998, p. 210) and the foundation for which fidelity of implementation is measured, curriculum-in-use appears to be viewed as that which is implemented by the teachers and not necessarily identical to the written curriculum (Munby & Russell, 1990; Shkedi, 1998). Shkedi reports that teachers make limited use of curriculum guides and “in most cases use the student materials themselves” (1998, p. 210); therefore, the researcher must distinguish between the written curriculum guide, which is the document specifically designed by writers for teachers, and the perceived curriculum as viewed and implemented by teachers (Ben-Peretz et al., 1982; Shkedi, 1998). In addition, curriculum use refers to the processes by which “individual teachers interact with, draw on, refer to, and are influenced by the material resources designed to guide instruction” (Remillard, 2005, p. 212). How to distinguish between the effects caused by the materials and the effects caused by the teachers' interactions with the materials, however, becomes difficult and represents a conflict when measuring the fidelity of implementation construct.

Adaptation. There is a dichotomy that exists in the literature between the two constructs of fidelity of implementation and adaptation. Notable researchers discuss the tension between fidelity and reinvention/adaptation and outline ways of resolving this tension (Blakely et al., 1987; Cho, 1998; Dusenbury et al., 2003; Hedges, 2004; Lee, 2002; U.S. Department of Health and Human Services, 2002). Hall and Loucks (1978), for example, argued that adaptation is acceptable up to the point of drastic mutations that compromise the program's integrity and effectiveness. Blakely et al. (1987) measured program fidelity, reinvention, and effectiveness in

diverse program settings. Results suggested that high-fidelity users tended to result in more effective implementations than low-fidelity users. When holding fidelity constant, additions enhanced effectiveness, but local modifications to the model were not related to greater program effectiveness (Blakely et al., 1987).

Cho looked at traditional curriculum implementation strategies and claimed that they have been divided into “two polarized perspectives: fidelity of implementation and adaptive implementation” (1998, p. 1). Like Cho, Fullan speaks of the “dilemma and tension running through the educational change literature in which two different emphases or perspectives are evident: the fidelity perspective and the mutual-adaptation or evolutionary perspective” (2001, p. 40). Cho’s theoretical argument was that curriculum implementation as a field of study should relinquish the notion of fidelity of implementation and should create a new construct where the role of the teacher can be realistically constructed. However, this argument may have been made outside of the context of intervention research, where fidelity serves as an internal validity check and helps to assure that the independent variable (the intervention) has been implemented as intended.

Rogers contends, “innovations that are more flexible and that can be more easily re-invented can be fit to a wider range of adopters’ conditions” (2003, p. 183). This explains why the rate of adoption for flexible innovations is more rapid (Backer, 2000) and leads to a higher degree of sustainability (Goodman & Steckler, 1989; Rogers, 2003). But does the adoption and sustainability of an adapted product lead to higher student outcomes? Research on rapid diffusion of the school-based drug abuse prevention program DARE (Drug Abuse Resistance Education) found that fidelity of implementation was low and adaptation to the program by local schools was high. Rogers claimed that the speedy rate of adoption of DARE occurred because “a good deal of re-invention took place” (2003, p. 182). Material was skipped and elements of the program were discarded. In the end, however, despite rapid adoption and sustainability, evaluations of the DARE program in U.S. schools showed few lasting effects on decreasing drug use by students (Ennett, Tobler, Ringwalt, & Flewellin, 1994; Lyman et al., 1999).

Summary. There are overlaps and disparities between fidelity of implementation and other constructs in the K–12 education literature. This review exposes the difficulty researchers could have studying fidelity of implementation if it is confounded with other (and often opposing) constructs, especially if these constructs guide the development of measures that must be independent from one another (see Allinder, Bolling, Oats, & Gagnon, 2000; Blakely et al., 1987; Emshoff et al., 1987; and Hall & Loucks, 1977, who examined teachers’ variations to interventions during their fidelity studies).

Section 3: Measuring Fidelity of Implementation

To better understand how fidelity of implementation to K–12 curriculum interventions has been measured and how these measures relate to outcomes, 23 primary studies of the more than 120 documents reviewed were identified and subjected to the evaluation criteria illustrated in Figure 1. Appendix A lists the results of this evaluation and describes how each of the 23 primary studies aligns with each of the evaluation criteria. The second column of Appendix A tells whether the article met all criteria (YES) or did not (NO). Table 1 summarizes articles that did not meet the evaluation criteria and explains why.

TABLE 1
Evaluation criteria and examples of excluded literature

Criterion #	Evaluation criteria	Reason for exclusion	Examples of excluded literature
1	Primary intervention research	Literature reviews	Backer, 2000; Dane & Schneider, 1998; Desimone, 2002; Dusenbury, Brannigan, Falco, & Hansen, 2003; Loucks, 1983; McGrew, Bond, Dietzen, & Salyers, 1994; Mihalic, 2002; Moncher & Prinz, 1991; Mowbray, Holter, Teague, & Bybee, 2003; National Research Council, 2004; Remillard, 2005; Resnick et al., 2005; Scheirer & Rezmovic, 1983; Snyder, Bolin, & Zumwalt, 1992; Summerfelt, 2003
		Theoretical articles	Ball & Cohen, 1996; Cho, 1998; Fullan & Pomfret, 1977; Hohmann & Shear, 2002
2	Fidelity of implementation to K–12 school-based standards, programs, or materials	Fidelity to community programs, such as employment, or other non-K–12 programs	Becker, Smith, Tanzman, Drake, & Tremblay, 2001; Forgatch, Patterson, & DeGarmo, 2005; McHugo, Drake, Teague, & Xie, 1999; Orwin, 2000
3	Efficacy or effectiveness of curriculum interventions for core subjects (e.g., math, science, social studies, reading) that can be adopted and implemented by a single teacher	Fidelity to: Noncore subject Core subject Prevention programs Whole-school models	Mihalic, Irwin, Fagan, & Elliott, 2004 Blakely et al., 1987; Datnow & Castellano, 2001; Emshoff et al., 1987

(continued)

TABLE 1 (continued)

Criterion #	Evaluation criteria	Reason for exclusion	Examples of excluded literature
4	Statistical quantitative measure of fidelity	Qualitative measure of fidelity	Schneider, Krajcik, & Blumenfeld, 2005
5	Fidelity–effectiveness correlation	Fidelity frequencies, means, or other descriptive statistics with no quantitative relationship to outcomes Fidelity instrument validation	Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; Fuchs, Fuchs, & Karns, 2001 Bond, Becker, Drake, & Vogler, 1997; Teague, Bond, & Drake, 1998
6	Sample size reported	No exclusions	

Although this article cites the work of many authors and recognizes the vast importance of all of the studies reviewed as contributing to the understanding of how fidelity of implementation is defined and conceptualized, only 5 studies of the 23 that were subjected to the evaluation met fully all criteria. These 5 studies are primary studies that were conducted in K–12 settings and examined the efficacy or effectiveness of core curriculum interventions (e.g., math, science, reading, or social studies) that could be adopted and implemented by a single teacher in a classroom for a limited duration. Curriculum interventions could include curriculum guides, training programs, evaluation systems, curriculum-based instructional management systems, computer courseware, or the like. Whole-school reform programs and non-core subject curriculum interventions did not meet Criterion 3. These 5 studies also used quantitative research methods to measure the relationship between fidelity of implementation to the K–12 core curriculum intervention and outcomes. Appendix B details how each of the 5 studies defined fidelity of implementation and summarizes each study's theoretical framework, purpose, program studied (intervention), sample, methods or methodology, analysis, and findings.

The sections that follow highlight the theoretical frameworks that guided the examination of fidelity, the fidelity criteria, and the fidelity measures of the five selected studies (with periodic references to the other studies as needed), and they examine the methods for measuring the relationship between fidelity of implementation to K–12 curriculum interventions and outcomes. Instruments used in each study are also considered.

Theoretical Frameworks

Theoretical frameworks, or models, for studying fidelity of implementation are important, especially when a researcher attempts to theorize (and ultimately measure)

the relationship among the components of the model. Constructs—sometimes referred to as components (Dane & Schneider, 1998; Dumas et al., 2001), criteria (Mowbray et al., 2003), or factors (Becker, Smith, Tanzman, Drake, & Tremblay, 2001) in fidelity studies—serve as the building blocks of theory (Turner, 1989) and help researchers to identify measurable variables of fidelity of implementation. Theoretical frameworks and criteria for measuring fidelity of implementation varied widely in the reviewed studies, and they often included some reference to teachers' potential to adapt.

Of the five studies that met the evaluation criteria, references to theoretical frameworks varied. Hall and Loucks (1977) based their work on the concerns-based adoption model (CBAM; developed by Hall, Wallace, & Dossett, 1973) to investigate the quality of teachers' lessons and their attempts to use and adapt the innovation to the context of their classrooms. This model moves beyond the dichotomous, bipolar use/nonuse of curriculum programs by creating a hierarchy called *Levels of Use* (LoU), which measures the teacher's growth from nonuse, to mechanical use (fidelity), to creative local adaptation and refinement (constructs separate from fidelity). Penuel and Means (2004) based their fidelity studies on the theory of change, which is built on the assumptions that teachers' choices about program enactment will not undercut the scientific aims of the program, local support will enhance the likelihood of fidelity, and aligning programs to locally defined standards will increase fidelity to the program. Songer and Gotwals (2005) provided a theoretical framework for the design of the intervention, and although this design (or program) theory did not directly influence Songer and Gotwal's measure of fidelity, such program theories often do guide fidelity measures (Mowbray et al., 2003). Neither of the researchers who investigated the effects of evaluation systems on math achievement specified a theoretical framework to guide their fidelity measures (Allinder et al., 2000; Ysseldyke et al., 2003).

The classic research, development, and diffusion model is often referenced in fidelity literature. One study that did not meet the criteria stated, teachers "who believe they are involved and effective in curriculum development will show greater congruence between intended and actual use of a curriculum" (Kimpston, 1985, p. 185), with teachers being more faithful to curriculum materials they have helped to design. Emshoff and colleagues (1987), who measured fidelity of implementation in social programs within education and criminal justice, referred to diverse organizational and social change theories as the framework for fidelity studies within two contrasting models—the rational-comprehensive model (Lindblom, 1959) and the theory of bounded rationality (March, 1978). These models served to define how new programs are conceived and implemented. The rational-comprehensive model is based on Dewey's 1933 account of the problem-solving process. This theory maintains that only one definitive programmatic solution could be identified and that it "could, would, and should be implemented by faithful, rational adopters" (Emshoff et al., 1987, p. 301). On the basis of this theory, fidelity of implementation is measured by a set of observable variables. Bounded rationality, on the other hand, states that social, political, economic, and organizational factors go beyond the limits of one individual's intellectual capacity. Implementers, according to Emshoff et al. (1987), do not consider that programs have been empirically validated in efficacy and effectiveness studies and

should be implemented with fidelity. Instead, program implementers choose program components that meet the immediate needs of the organization, and they modify the program to fit the organizational constraints. Others refer to this model as the mutual adaptation model or the coconstruction perspective (Berman & McLaughlin, 1976; Blakely et al., 1987; Price, Friedland, Choi, & Caplan, 1998; Snyder et al., 1992). This model states that local variation is inevitable.

Fidelity Criteria

Rather than creating a theoretical framework ad hoc, it is important to establish a theoretical framework or conceptual model a priori. This framework can then guide a researcher in identifying critical components of the intervention and, ultimately, constructing fidelity of implementation criteria and measures.

To develop fidelity criteria, researchers often reported starting with a curriculum profile or analysis that outlined the critical components of the intervention along with an indication of the range of variations for acceptable use. The researcher or developer then outlined acceptable ranges of variation (Songer & Gotwals, 2005). A component checklist was then developed to record fidelity to these components or LoU (Hall & Loucks, 1977). Two of the five studies reported details on the construction of a fidelity instrument (Allinder et al., 2000; Hall & Loucks, 1977) and included interitem correlations and interrater reliability. All identified critical components of the curriculum intervention and then developed measures of implementation based on these components.

Mills and Ragan (2000), a study excluded from the evaluation because it focused on fidelity instrument validation, suggested that researchers consult with developers when creating component checklists using a five-step process: (a) Identify the innovation components—participant activities, behaviors of the implementer, materials—by reviewing the program materials and consulting the program developer; (b) identify additional components and variations by interviewing past implementers to ascertain ideal use and unacceptable use for each component; (c) refine the program components by going back to the developer and clarifying with him or her user discrepancies regarding which of the observed components is the most important; (d) finalize the innovation components by constructing a component checklist and a set of variations within each before piloting; and (e) collect data either in writing (i.e., through questionnaires), classroom observation, or interview.

Fidelity Measures

Methods used to collect fidelity data varied from self-report surveys and interviews (Allinder et al., 2000; Hall & Loucks, 1977; Penuel & Means, 2004) to analysis of student artifacts (Songer & Gotwals, 2005; Ysseldyke et al., 2003). Although the traditional view of fidelity is to measure a teacher's adherence to the structural components of an intervention, two of the five studies measured students' adherence to the structural components of the intervention, with Songer and Gotwal (2005) measuring students' percentage of completed student worksheets and Ysseldyke et al. (2003) measuring the number of objectives students mastered compared to the number of problems they attempted. Both of these measures are examples of what Dane and Schneider (1998) called "participant responsiveness." The remaining three studies (Allinder et al., 2000; Hall & Loucks, 1977; Penuel &

Means, 2004) measured fidelity to the teachers' quality of delivery (cf. Dane & Schneider, 1998). The criteria for quality of delivery were set prior to implementation and were based on the intervention's program theory.

In other studies, which did not meet the full criteria in Figure 1, researchers used multimethod, multisource methodologies that included use of video and audiotapes, classroom observations, questionnaires, and teacher interviews (cf. Dumas et al., 2001; Emshoff et al., 1987; Kimpston, 1985; Kutash, Duchnowski, Sumi, Rudo, & Harris, 2002; Mihalic, Irwin, Fagan, & Elliott, 2004; see also Ruiz-Primo, 2005). When self-reports were used simultaneously with field observations, researchers found that self-report data indicated higher levels of fidelity than were observed in the field (Emshoff et al., 1987).

As researchers disseminate results from efficacy trials to effectiveness studies, it is vital that policy makers and practitioners who are adopting the program understand, both qualitatively and quantitatively, what structural components and processes of the intervention are necessary to produce similar outcomes. By examining and measuring fidelity criteria using multimethods and relating these measures to student outcomes, researchers can differentiate between implementation failure and program failure (Harachi, Abbott, Catalano, Haggerty, & Fleming, 1999; Leithwood & Montgomery, 1980).

Relationships to Outcomes

All five studies consistently showed statistically significantly higher outcomes when the program was implemented with greater fidelity. Relationships ranged from causal (Ysseldyke et al., 2003), to associational (Hall & Loucks, 1977), to predictive (Songer & Gotwals, 2005). Only one of the five studies reported that fidelity of implementation was measured as the dependent variable (Allinder et al., 2000), with fidelity of implementation being caused by teachers' self-monitoring practices—the intervention of interest.

Considerable variability is often reported within the treatment condition; therefore, different results from the same program raise concerns that differential implementation may account for the variability (Hall & Loucks, 1977). The question of what components account for or predict such variability in outcome data assumes the need for multiple regression, yet only one of the five studies reported using such analytical techniques (Songer & Gotwals, 2005). Health studies that have used multiple regression to examine the variance in outcomes due to fidelity of implementation include Latimer (1999); Mills and Ragan (2000); Rohrbach, Graham, and Hansen (1993); and White et al. (2005).

Significance tests (*t* tests, analysis of variance [ANOVA], multivariate ANOVA) were often used in comparative studies to examine the differences between the treatment (intervention) and comparison conditions (or to examine the effects caused by each condition). In addition, ANOVA was also used to study the differences between levels of implementation (Hall & Loucks, 1977; Ysseldyke et al., 2003; see also Kimpston, 1985; Mills & Ragan, 2000). Teachers in two or more levels were compared to one another to determine statistically significant differences in program outcomes.

Units of analyses for the five studies also differed, but most used the teacher or classroom and aggregated student achievement scores to the classroom level; only one reported data at the school level (Penuel & Means, 2004). None of the five studies attempted to adjust teachers' levels of fidelity during implementation to

assure higher fidelity—a practice that appears to be more common in efficacy studies. Unadjusted fidelity levels result in variability, which allows researchers to relate fidelity variability to outcomes—a practice that appears to be more common in effectiveness studies.

*Section 4: Applying Fidelity of Implementation to K–12
Core Curriculum Intervention Research*

Although fidelity of implementation appears to be well defined and conceptualized in the public and mental health fields, the measure of fidelity of implementation appears to have been relatively neglected in K–12 curriculum intervention research, making this literature review that applies fidelity of implementation to K–12 core curriculum intervention research critical for the field. In this review, only 5 of the 23 studies subjected to the evaluation criteria measured the relationship between fidelity of implementation to K–12 core curriculum interventions and outcomes. Other literature reviews have shown similar patterns. Compare this to the 13 of 162 public health studies reviewed by Dane and Schneider (1998) that considered the impact of fidelity on outcomes, and the 1 of 63 studies in math evaluations that adjusted outcomes based on the fidelity measure (NRC, 2004). Therefore, although there have been several literature reviews in the past that have examined fidelity of implementation, and much can be learned from these reviews, none have examined specifically the relationship between fidelity of implementation to K–12 core curriculum interventions and outcomes. This review demonstrates the need for K–12 curriculum intervention researchers to measure and report such outcomes more readily.

After reviewing the fidelity literature, it is apparent that there is a shortage of K–12 core curriculum intervention studies that empirically measure fidelity of implementation and its relationship to outcomes. On the basis of the extensive work that has been done in the public and mental health fields, six points should be considered when applying fidelity of implementation to K–12 curriculum intervention research:

1. Fidelity of implementation to public and mental health programs appear to be guided by organizational or social theories. Although fidelity of implementation appears to be undertheorized in K–12 curriculum intervention research, program theory can be used to guide fidelity measures.

To better understand and measure the relationship between fidelity of implementation and outcomes, curriculum intervention researchers should establish a theoretical framework for studying fidelity of implementation a priori. This framework can be based on the program's theory and used to understand how fidelity is conceptualized and measured; it can also help implementers identify local adaptations that do not contradict the underlying theory.

2. There is a need for greater clarity and specificity in the articulation of the critical components or features that make up a curriculum intervention. Distinctions should be made between measuring fidelity to the structural components of a curriculum intervention and fidelity to the processes that guide its design.

Mowbray et al. (2003) stated that process criteria may be more difficult to measure reliably, but they may also be more significant in terms of program effects. Studies often dismiss in one statement that fidelity of implementation to the structural components of the unit has been monitored, and little variation in implementation is noted. However, the strongest relationships to outcomes appear to be caused by variation in fidelity to the processes supported by the curriculum intervention; therefore, researchers should measure fidelity to both the structure and processes of an intervention, and relate both to outcomes.

3. Fidelity to a whole-school reform model is different from fidelity of implementation to a curriculum intervention that can be used and implemented by an individual teacher for a limited duration.

Whole-school program implementation occurs within organizations such as schools that seek to comprehensively put a particular program in place. Fidelity to such models, which entail the collaboration of an entire system, is more complex than fidelity to interventions that can be used and implemented by teachers in an individual classroom for a limited duration. Although there may be general fidelity to the implementation of the structures and routines of a whole-school program throughout a school, individual teachers may adapt materials and routines for their particular needs in the classroom. The question, therefore, is how to isolate and attribute such adaptations to outcomes (see Desimone, 2002, for a review of the literature on fidelity to comprehensive school reform models). The studies reviewed here suggest that fidelity of implementation is more probable when an intervention manual is in place that clearly defines the critical components of the intervention and articulates a theory. In addition, the specificity of the intervention may determine the extent to which fidelity to the structural components and processes are possible. Higher specificity often leads to higher fidelity.

4. Methods for measuring and reporting fidelity of implementation patterns differ for efficacy and effectiveness studies.

Intervention researchers should not assume that fidelity to curriculum interventions will be high, nor should they dismiss the fact that teachers may adapt programs to suit their needs without capturing these effects. During an efficacy trial (or RCT), curriculum developers, intervention researchers, and teachers should identify the critical structural components and processes of the curriculum intervention and set acceptable ranges of fidelity for each a priori. Fidelity of implementation to the critical components and processes of the intervention are then constantly monitored and feedback is given to the teacher during the trial to ensure ideal conditions. When fidelity to the independent variable is assured, outcomes can be attributed to the curriculum intervention. Researchers should report methods for capturing implementation, even if the "curriculum was implemented consistently and correctly" (Woodward & Brown, 2006, p. 154). During an effectiveness study, however, where implementation occurs in a natural setting, variations in implementation are expected. Fidelity to the critical components and processes should be captured quantitatively so that levels of fidelity can be related to outcomes. Outcomes would be adjusted if fidelity falls outside of the acceptable range (e.g.,

Vaughn et al., 2006, indicated that 70% to 100% of practices were observed in classrooms; however, no acceptable range was noted and no relationships to outcomes were reported).

5. Adaptation and fidelity of implementation are different constructs and should be separately measured and related to outcomes.

Although this review does not claim that adaptation and variation are bad and fidelity to curriculum materials is good, it does state that fidelity and adaptation are separate constructs and the effects of each should be isolated and empirically related to outcomes. This review acknowledges that there are different perspectives on whether fidelity is a necessary condition for success or whether adaptation is inevitable and perhaps itself related to program success. However, of the studies that were reviewed, those that related fidelity of implementation to outcomes indicated a positive relationship, and additions to the materials were positively related to outcomes only when fidelity was held constant (in other words, fidelity was moderating the relationship between adaptation and outcomes). This was not the case with modifications, which were not related to outcomes (see Blakely et al., 1987).

6. Building on the work of the health field (cf. Resnick et al., 2005), a set of guidelines must be established to better understand how to measure fidelity of implementation to K–12 curriculum interventions and relate it to outcomes.

Therefore, this review proposes the following as a set of guidelines for future K–12 core curriculum intervention researchers:

- Establish the program theory a priori and determine what it means to implement the program with fidelity.
- Operationally define fidelity of implementation constructs and variables by specifying the critical components and processes necessary for implementing the curriculum intervention with fidelity. (In an efficacy study, inform the users of the critical components and processes and help them understand what fidelity is and is not; monitor and adjust their fidelity levels as needed.)
- Develop separate instruments for measuring the critical components and processes. If the program theory promotes adaptation, measures of fidelity to the critical components and processes should be separate from measures of the user's adaptations and variations.
- Incorporate random or full census sampling within the study in order to generalize fidelity findings to the study population.
- Measure the user's fidelity to the critical components and processes; measure fidelity to processes in both the experimental and comparison condition, and relate these measures to outcomes. If measures show that the critical processes are present in both the experimental and comparison conditions, but fidelity to these processes is related to outcomes only in the treatment condition, then fidelity of implementation to these processes is moderating the relationship between curriculum condition and outcomes (see O'Donnell, 2007, for specific methods).
- Test for and report on the reliability and validity of the fidelity data collected.

Conclusions

To ensure the integrity of scientifically based research, curriculum intervention researchers conducting efficacy and effectiveness studies are now being asked to describe how treatment fidelity is defined, conceptualized, and measured. Fidelity measures are needed to explain the degree of variation in treatment implementation and how it might affect or moderate outcomes. This means that fidelity measures must capture the critical features of the intervention's program theory. Such measures should also indicate how the intervention is maintained consistently across multiple groups over time or describe the parameters under which variations in the implementation may occur (U.S. Department of Education, 2006). Therefore, measures of fidelity of implementation are critical to determine not only if the intervention is sufficiently implemented, but also if there are critical differences between what the experimental and comparison groups receive so that one might expect a difference in student outcomes (U.S. Department of Education, 2006).

The purpose of fidelity of implementation research is to better operationalize and measure implementation criteria in practice during intervention studies. However, assuming relevant data related to the criteria can be collected, no universal data collection tools exist that can be applied across a wide variety of fidelity of implementation studies. As was demonstrated by this literature review, a universal fidelity instrument may not be possible because fidelity measures need to be designed with a specific program theory or type of program in mind. Researchers should follow a framework or set of guidelines, such as those proposed by Resnick et al. (2005) or here, to develop these measures.

To conduct more rigorous evaluations of curriculum interventions, researchers need to better understand how to define, conceptualize, and measure fidelity of implementation during efficacy and effectiveness studies, and how fidelity measures at each of these stages differ. The statistical power of a study relies on the use of reliable, valid measures; appropriate design and sampling; and careful assessment of fidelity. Each of these three decreases as research moves from the laboratory (efficacy studies) and toward the field (effectiveness studies; Boruch & Gomez, 1977; Hohmann & Shear, 2003; Summerfelt, 2003).

Further high-quality study considering the recommendations from the current review relating fidelity of implementation to K–12 curriculum interventions and outcomes is warranted. This review shows that there are too few studies to guide researchers on how fidelity to core curriculum interventions can be measured and related to outcomes. The outcomes of this review, therefore, have implications for future scientifically based studies of K–12 curriculum interventions and for practitioners' understanding of the need for valid and reliable classroom data collection. It demonstrates the need for K–12 curriculum intervention researchers to measure and report such outcomes more readily. In addition, this review, along with others, may help to clarify the definition, conceptualization, and measurement of fidelity of implementation—the lack of which has been a longstanding problem in K–12 curriculum intervention research.

APPENDIX A
Alignment of 23 primary studies to evaluation criteria

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Allinder, Bolling, Oats, & Gagnon (2000)	YES	YES	Elementary school (ES) students with learning disabilities and mild mental disabilities; midwestern school district	YES Intervention: curriculum-based measurement (CBM) formative evaluation system Subject: math computation skills	YES • Quasi experiment: teachers assigned to 1 of 3 groups: CBM, CBM with self-monitoring checklist of key components, or control (no CBM) • Question: does self-monitoring affect teachers' modification of instructional plans? • Fidelity measured with Math-Modified Accuracy of Implementation Rating Scale (M-MAIRS); 5-point Likert-type scale; $\alpha = .73$; researchers collected fidelity by analyzing students' graphs at end of project and observing students during Week 8 using M-MAIRS	YES • Fidelity = DV • MANOVA on M-MAIRS scores = no significant overall difference in CBM implementation between groups, but CBM + self-monitoring made significantly more instructional changes • CBM + self-monitoring teachers made greater modifications to instruction and their students had significantly greater math achievement gains than CBM-alone group or control group	Demographically matched teachers (12 = control, 10 = CBM only, 8 = CBM with self-monitoring)

(continued)

APPENDIX A (continued)

		Evaluation criteria					
Primary authors and publication year	Met all criteria	1	2	3	4	5	6
Becker, Smith, Tanzman, Drake, & Tremblay (2001)	NO	YES	NO	NO	YES	YES	2,639 clients with severe mental illness in 10 community mental health centers
		Primary research	Community mental health centers' fidelity of supported employment programs	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
					<ul style="list-style-type: none"> Measured presence of program components using Individual Placement Support Fidelity Scale of 15 items rated on 5-point Likert-type scale (<i>not implemented to fully implemented</i>); validated survey 	<ul style="list-style-type: none"> Pearson correlations used to compare fidelity scores with competitive employment rates (the DV) High correlation 	
					<ul style="list-style-type: none"> Preliminary validation of scale, but scale is considered exploratory Bivariate correlations Principle components analysis (PCA) used to identify components that account for variance in outcomes; 2 components 		

(continued)

APPENDIX A (continued)

		Evaluation criteria					
Primary authors and publication year	1 Met all criteria	2 Primary research	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis	
Bond, Becker, Drake, & Vogler (1997)	NO	YES	NO	YES	NO	27 sites	
		Employment sites for people with severe mental disorders	Individual placement and support (IPS) model of supported employment	YES • Semistructured questionnaire used to administer fidelity scale; items coded; mean item values calculated; interrater reliability; internal consistency reliability; compared IPS with other vocational programs	NO Focus on scale development; developed fidelity scale to measure extent to which specific programs met standards for a program model		
Datnow & Castellano (2001)	NO	YES	NO	NO	NO	6 schools	
		Success for All (SFA) comprehensive school reform	Intervention: SFA (whole-school reform) Subject: reading	• Case study • 4 site visits to each school (1997 to 1999) • 103 45-min semistructured interviews and focus groups with principals, facilitators, and teachers; interviews were taped and transcribed verbatim	Coded transcripts about how principal leadership shaped reform and successes and challenges faced by SFA facilitators in occupying a position of teacher leadership		

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Dumas, Lynch, Laughlin, & Prinz (2001)	NO	YES	YES Family, peer, and K-12 school intervention program	NO Intervention: Early Alliance	YES • 231 audio or videotaped sessions randomly selected for coding • Content and Process Fidelity Component checklists • Interrater reliability of 4-point Likert-type scale • DV: conduct disorder, substance abuse, school failure	NO Descriptive data; reported users' % of fidelity; did not correlate this use with any outcomes	Unit of analysis is implementer; 231 sessions observed in 12 schools

(continued)

APPENDIX A (continued)

Primary authors and publication year	Evaluation criteria					
	1 Met all criteria	2 Primary research	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Emshoff et al. (1987) (see also Blakely et al., 1987)	NO	YES <ul style="list-style-type: none"> National Diffusion Network NIJ Exemplary Project Program 	NO Intervention: Help One Student to Succeed (HOSTS) whole-school model for Grades 2-6 Subject: reading	YES <ul style="list-style-type: none"> Fidelity and effectiveness criteria developed for each program Multimethod data collection (phone interviews, on-site observations, archival analysis of program adopters) Intrater reliability reported Component variation rated using 3-level fidelity score: I = <i>ideal</i>, A = <i>acceptable</i>, U = <i>unacceptable</i> Developers used for face validity External validity: comparing responses from different data sources (e.g., supervisor vs. employee); convergent validity: comparing results of telephone interviews with site visits 	YES Fidelity positively correlated with effectiveness of the HOSTS adopted program (e.g., normal curve equivalency pre/post gains on nationally normed referenced reading achievement tests such as CAT, Gates-McGintie, CTBS)	129 organizations in all; 10 organizations implemented HOSTS for 2 years of use

(continued)

APPENDIX A (continued)

Primary authors and publication year	Evaluation criteria					
	1 Met all criteria	2 Primary research setting	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Forgatch, Patterson, & DeGarmo (2005)	NO	YES Oregon model of parent management training	NO Examined parenting practices	YES Fidelity of Implementation Rating System: observation-based measure of content and process using videotapes; tapes randomly selected and then coded blind • Theory-based manual; therefore, rigid adherence to manual not required; instrument included • <i>t</i> tests comparing treatment and comparison groups • Confirmatory factor analysis; structural equation modeling; predictive validity of scores in path analysis	YES Fidelity served as a predictor variable to parenting practices in mothers and stepfathers	20 families randomly selected from 110 stepfamilies (quasi experiment)

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	YES	YES	YES	YES	NO	NO
Fuchs, Fuchs, & Karns (2001)		YES	YES	YES	YES	NO	NO
		Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
		Three Title I schools and two non-Title I ESS in southeastern public school system	Peer-Assisted Learning Strategies (PALS)	<p>Intervention:</p> <ul style="list-style-type: none"> Peer-Assisted Learning Strategies (PALS) <p>Subject:</p> <ul style="list-style-type: none"> mathematics 	<ul style="list-style-type: none"> Quasi experiment; within schools, 20 classrooms randomly assigned to experimental (PALS curriculum) or control (basal textbook) Measured fidelity in treatment classrooms; teachers completed questionnaires about feasibility and effectiveness 	<ul style="list-style-type: none"> Reported descriptive statistics only (mean fidelity score 90% for 10 treatment teachers) 	<ul style="list-style-type: none"> 20 teachers (unit of analysis) in 5 schools 168 students (84 per condition)

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Hall & Loucks (1977)	YES	YES	YES	YES	YES	YES	• 11 treatment (T) (IGE) and 11 comparison (C) (non-IGE) schools
		Individually guided education (IGE)		<p>Intervention: second- and fourth-grade individualized instruction (component of IGE bundle)</p> <p>Subject: math and reading</p>	<p>• Cross-sectional study</p> <p>• Used validated interviews to collect % of distribution of overall Level of Use (LoU) for each participant at 8 levels</p> <p>• Assessed proportion of users at each LoU for a stratified sample selected according to years of experience (<i>n</i>s not equal)</p> <p>• IGE and non-IGE schools divided into nonusers (LoU 0-II) and users (LoU III-IV)</p>	<p>• ANOVA shows all users scored significantly higher on achievement tests than nonusers (except for second-grade math)</p> <p>• Curvilinear relationship between users at mechanical stage (LoU III) and reading achievement (but not the same across innovations and grade levels)</p>	<p>• 134 teachers (unit of analysis)</p>

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Hughes, Frederick, & Keel (2002)	NO	YES Middle school (MS) in northeastern Georgia	YES Intervention: constant time delay (CTD) Subject: spelling (written words)	YES Intervention: constant time delay (CTD) Subject: spelling (written words)	YES • Measured frequency of fidelity to 4 teacher behaviors (e.g., reinforcement) • Collected fidelity for 25% of sessions in each of 20 trials	NO • Fidelity measured at 100% • Established relationship between CTD and student's ability to acquire, maintain, and generalize 15 spelling words (but did not establish relationship between DV and fidelity)	• 1 special education resource teacher • 1 12-year-old boy with a learning disability

(continued)

APPENDIX A (continued)

		Evaluation criteria					
Primary authors and publication year	Met all criteria	1 Primary research	2 K-12 setting	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Kimpston (1985)	NO	YES	YES ES, MS, and high school (HS)	NO Focused on ES, MS, and HS language arts curriculum standards, but not classroom intervention	YES Objective Implementation Assessment and Curriculum Implementation System Assessment self-report questionnaires (groups or individually) • Face validity; also validated through follow-up stratified sample interviews (6 teachers total) • Teachers categorized as high or low implementers based on their location above or below mean of agreement with district objectives	NO IV = teacher beliefs about their roles in curriculum development; DV = degree of fidelity • <i>t</i> tests significant differences between 2 groups (high vs. low) and their beliefs in involvement in curriculum tasks at HS and ES levels (ES higher fidelity than HS) • No correlations tested	64 teachers in 3 ES, 1 MS, and 1 HS

(continued)

APPENDIX A (continued)

		Evaluation criteria					
	Met all criteria	1	2	3	4	5	6
Primary authors and publication year		Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Kutash, Duchnowski, Sumi, Rudo, & Harris (2002)	NO	YES	YES School-based mental health services	NO Intervention: school-based partnership program (PP) Subject: reading	YES Fidelity Form assessed degree to which concepts and principles were used (46 dichotomous items) • 72% of concepts that form the PP were implemented • Intervention for students with ED; included training program for professionals from school and community agencies; families included as partners	YES • Significant correlations (.49) between average fidelity rating and changes in reading scores; no correlations with changes in math, CBCL, or CAFAS scores	23 T students; 24 C students who volunteered for study (numbers dropped over time)

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Mihalic, Irwin, Fagan, & Elliott (2004)	NO	YES	YES School-based setting	NO 9 (8 violence + 1 drug) prevention programs	YES • Questionnaires: measured agencies' adherence to program elements, quality of implementation, and teachers' adherence to program • Checklist; unannounced classroom visits • Phone interviews • Data reported as frequencies	YES (but reverse) Fidelity was treated as a dependent variable, and training led to greater fidelity	147 sites for 2 years
Mills & Ragan (2000)	NO	YES	YES ES integrated learning system (ILS) software	YES Intervention: Successmaker ILS Subjects: math, reading, science, writing	YES • ILS Configuration Matrix; 15 implementation components with 5 variations of implementation fidelity • 45-min interview tape recorded and transcribed; checklist completed by user; no classroom observation • Determined if there were differences in operational patterns of teachers implementing ILS and which teacher practices exhibited high fidelity	NO Discriminant analysis (DA) and ANOVA used to explore differences among the configuration patterns (high, medium, and low fidelity) for each component of the fidelity instrument	29 teachers randomly selected from all ILS teachers at 4 schools recommended by tech coordinator (2nd- and 3rd-year users)

(continued)

		Evaluation criteria					
Primary authors and publication year	Met all criteria	1 Primary research	2 K-12 setting	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Orwin (2000)	NO	YES	NO	NO	YES	YES	9 sites
		Community health services	Homeless Cooperative Agreement Program	<ul style="list-style-type: none"> • Quarterly Report Form, open-ended responses coded; captured presence or absence of 39 distinct services; phone calls; site visits; fidelity scale was model dependent; assessed duration 	<ul style="list-style-type: none"> • Statistical significance between groups (based on intervention strength) • Effect sizes reported 		
Penuel & Means (2004)	YES	YES	YES	YES	YES	YES	1,500 schools
		K-12 (Note: Studied both fidelity and implementation on variation)	Intervention: GLOBE Web international program Subject: science	<ul style="list-style-type: none"> • Teachers' frequency in data reporting • 4 data reporting levels (nonreporters, periodic reporters, average reporters, steady reporters) along 2 dimensions (consistency: reporting data monthly, and persistence: reporting data each year) • Teacher surveys administered on annual or biennial basis 	<ul style="list-style-type: none"> • Associations among teachers' responses to survey items to patterns in GLOBE program data reporting • χ^2 used (e.g., no poststraining supports predicted persistence in data reporting) 		

(continued)

APPENDIX A (continued)

Primary authors and publication year	Evaluation criteria					
	1 Met all criteria	2 Primary research setting	3 K-12 classroom intervention	4 Core curriculum	5 Statistical measure of fidelity of implementation	6 Fidelity-effectiveness correlation and unit of analysis
Schneider, Krajick, & Blumenfeld (2005)	NO	YES Eighth graders using an inquiry-based curriculum unit	YES Intervention: "Why do I need to wear a bike helmet?" unit Subject: science	NO Videotaped classroom enactment; coding scheme captured 3 aspects of enactment in comparison to what was intended in materials; 7 teacher and student behaviors were observed in each video;	NO coded tapes for 8 categories (rated descriptively 1-4); used shades to show patterns of enactment (i.e., fidelity to intended enactment)	4 teachers

(continued)

APPENDIX A (continued)

		Evaluation criteria					
		1	2	3	4	5	6
Primary authors and publication year	Met all criteria	Primary research	K-12 setting	Core curriculum classroom intervention	Statistical measure of fidelity of implementation	Fidelity-effectiveness correlation	Sample size and unit of analysis
Songer & Gotwals (2005)	YES	YES (peer-reviewed AERA symposium paper)	YES Detroit public MSs (sixth grade)	YES Interventions: BioKids, Kids as Global Scientists, Big Things Subject: science	YES Documented how many lessons and student worksheets associated with those lessons teachers enact in their classrooms; created 2 groups of teachers for each curricular unit: a high-fidelity group (high amount of implementation of student worksheets) and a low-fidelity group (low amount of implementation of student worksheets)	YES Multiple regressions indicate students receiving biodiversity unit with high-fidelity score higher on posttest than students in counterpart low-fidelity classrooms ($p < .01$) when controlling for pretest scores	23 teachers with a range of experience and expertise

(continued)

APPENDIX A (continued)

Primary authors and publication year	Evaluation criteria					
	1 Met all criteria	2 Primary research	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Teague et al. (1998)	NO	YES	NO	YES	NO	50 programs expected to range in fidelity (4 levels); sizes range from 40 to 75 clients
		Community-based service helping people with severe mental disorders	Fidelity to 50 programs in ACT	<ul style="list-style-type: none"> 26-item instrument (28 variables across 3 domains to measure fidelity in T and C); multimethods; reports from supervisors; document review; interviews; daily checklists completed by workers; descriptive statistics; composite ratings (FA/PCA and hierarchical cluster analysis); correlations between scales; % of variance (and loading) on each factor 	<ul style="list-style-type: none"> Correlation between fidelity to high-intensity services and community treatment (proposed that future studies examine relationship to effectiveness) 	

(continued)

APPENDIX A (continued)

		Evaluation criteria					
Primary authors and publication year	Met all criteria	1 Primary research	2 K-12 setting	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis
Vaughn et al. (2006)	NO	YES	YES First graders from 3 sites in Texas with high % ELL students	YES Intervention: proactive reading Subject: reading	YES • Validity checks = ratings (2.73-2.9) and field notes • Classroom observations across 8 observable teacher behaviors (max score = 3) conducted at beginning, middle, and end of year; interrater reliability of observations	NO • Descriptive data only • Range of 74%-100% presence of 9 instructional practices (frequency of occurrence)	10 schools, 42 classrooms
Vincus (2004)	NO	YES	K-12	NO ALERT drug prevention school-based program	NO Classroom observations; once per school; 11 lesson-specific checklists; general form for assessing classroom management, teaching techniques, and students' responses; instrument provided	NO • Frequencies only	11 schools; observed 1 lesson per school

(continued)

APPENDIX A (continued)

		Evaluation criteria					
Primary authors and publication year	1 Met all criteria	2 Primary research	3 Core curriculum classroom intervention	4 Statistical measure of fidelity of implementation	5 Fidelity-effectiveness correlation	6 Sample size and unit of analysis	
Ysseldyke et al. (2003)	YES	YES ESs in large urban school district in the Midwest	YES Intervention: accelerated math computerized instructional management system Subject: math	YES • Level of implementation determined by mean number of objectives mastered and mean number of problems attempted by each class; classrooms coded as partial implementers (1.6–21 mean range for 8 treatment classes) to full implementers (22.7–90.2 mean range for 10 classes)	YES Full implementers demonstrated more growth than partial implementers or those with no access to treatment software, $F(2, 459) = 4.126$, $p < .02$, $d = .13$	18 classrooms in study; $T = 397$ students in 4 sites; $C1 = 484$ students Grades 3–5; $C2 = 429$ students	

Note: Total studies meeting Criterion 1 (primary study) = 23. Total studies meeting all criteria = 5. AERA = American Educational Research Association; ANOVA = analysis of variance; CAFAS = Child & Adolescent Functional Assessment Scale; CAT = California Achievement Test; CBCL = Child Behavior Checklist; CTBS = Comprehensive Test of Basic Skills; DV = dependent variable; ED = emotional disturbance; ELL = English language learner; FA = factor analysis; IV = independent variable; MANOVA = multivariate analysis of variance; NIJ = National Institute of Justice; SESAT = Stanford Early School Achievement Test.

APPENDIX B
Summary of studies meeting full criteria

Citation and title	Definition of fidelity of implementation	Theory	Research purpose/intervention	Methods (sample, unit of analysis, data analysis, measurement used)	Results/findings
Allinder, Bolling, Oats, & Gagnon (2000): <i>Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities</i>	None specified; appears to focus on fidelity to modifications to enhance student progress	None specified	<p>Purpose: does self-monitoring affect teachers' modification of instructional plans?</p> <p>Intervention: curriculum-based measurement (CBM), formative math evaluation system</p>	<p>Sample: elementary students with LD and mental disabilities from midwestern school district</p> <p>Methods: quasi experiment (CBM, CBM + self-monitoring, control); MANOVA</p> <p>Sources of data: Self-checklist of key components; Implementation Rating Scale (M-MAIRS); 5-point Likert-type scale; $\alpha = .73$; observing students</p>	<ul style="list-style-type: none"> • Fidelity = DV • Teachers who used CBM + self-monitoring made greater modifications to instruction and their students had significantly greater math achievement gains than CBM-alone group or control group

(continued)

APPENDIX B (continued)

Citation and title	Definition of fidelity of implementation	Theory	Research purpose/intervention	Methods (sample, unit of analysis, data analysis, measurement used)	Results/findings
Hall & Loucks (1977): <i>A developmental model for determining whether the treatment is actually implemented</i>	Faithfulness to an innovation's characteristics; level of use	Concerns-based adoptions model	<p>Purpose: to assess implementation characteristics to understand treatment effects</p> <p>Interventions: individually guided education (IGE) bundle</p>	<p>Population: Teachers of Grades 2–4 teachers in 11 treatment (T) and 11 comparison (C) schools</p> <p>Sample: 134 teachers in cross-sectional study</p> <p>Source of data: interviews; assessed % of users at each of 8 Levels of Use (LoU) for a stratified sample selected according to years of experience</p>	<ul style="list-style-type: none"> • High correlation between users at mechanical stage (LoU III) and high reading achievement • No differences between non-IGE and IGE groups; wide variation in LoU in both T and C
Penuel & Means (2004): <i>Implementation variation and fidelity in an inquiry science program; analysis of GLOBE data reporting patterns</i>	Variations in the implementation of educational innovations	Change theory	<p>Purpose: to examine pattern variations in enactment and contextual factors related to large-scale K–12 science inquiry program implementation</p> <p>Intervention: GLOBE, an international environmental science program that involves students in the collection of data for real scientific investigations</p>	<p>Sample: 1,500 schools</p> <p>Methods: data reporting and survey</p> <ul style="list-style-type: none"> • 4 data reporting levels: nonreporters and periodic, average, and steady reporters • Teacher surveys administered on an annual or biennial basis • Data reporting measured along 2 dimensions: consistency and persistence 	<ul style="list-style-type: none"> • Significant relationships between specific barriers and supports, and levels of data reporting • Schools' data reporting levels influenced by access to posttraining support, barriers to implementation, and years of experience in the program

(continued)

APPENDIX B (continued)

Citation and title	Definition of fidelity of implementation	Theory	Research purpose/intervention	Methods (sample, unit of analysis, data analysis, measurement used)	Results/findings
Songer & Gotwals (2005): <i>Fidelity of implementation in 3 sequential curricular units</i>	Extent to which delivery of an intervention adheres to the protocol or program model originally developed	Scientific inquiry	<p>Purpose: to determine if fidelity of implementation of various components predict student outcomes</p> <p>Interventions: BioKIDS, Kids as Global Scientists, Big Things</p>	<p>Sample: 2,000 sixth-grade students from 16 Detroit public schools; 23 teachers with range of experience</p> <p>Method: high-fidelity group (high amount of implementation of student worksheets) and a low-fidelity group (low amount of worksheets)</p> <p>Analysis: multiple regressions; relation between the DV (posttest scores) and IVs</p>	Pretest scores and level of fidelity (high number of worksheets used) accounted for high percentage of variance in student outcomes

(continued)

APPENDIX B (continued)

Citation and title	Definition of fidelity of implementation	Theory	Research purpose/intervention	Methods (sample, unit of analysis, data analysis, measurement used)	Results/findings
Ysseldyke et al. (2003): <i>Using a curriculum-based instructional management system to enhance math achievement in urban schools</i>	Propensity of teachers to implement interventions with varying degrees of classroom level implementation	None specified	<p>Purpose: to examine the impact of the level of implementation on student performance</p> <p>Intervention: accelerated math computerized curriculum-based measurement system</p>	<p>Sample: 18 classrooms T = 397 students in treatment group across 4 sites C1 = 484 Grades 3–5 control students in same school C2 = 429 students randomly selected district-wide</p> <p>Method: counted mean number of problems attempted by each class and objectives mastered; 8 classes coded as partial implementers and 10 as full</p>	ANCOVA, using pretest as covariate, showed full implementers had more growth on achievement test than partial or no-treatment control, $F(2, 459) = 4.126, p < .02, d = .13$

Note. ANCOVA = analysis of covariance; DV = dependent variable; IV = independent variable; LD = learning disability; MANOVA = multivariate analysis of variance; M-MAIRS = Math-Modified Accuracy of Implementation Rating Scale.

Note

I would like to thank Sharon Lynch of The George Washington University, who served as my advisor and dissertation chair during the writing of this article and provided guidance on a previous version. I would also like to thank the anonymous reviewers for their comments and feedback.

References

- Allinder, R. M., Bolling, R. M., Oats, R. G., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education, 21*, 219–227.
- Aron, D. C., Raff, H., & Findling, J. W. (1997). Effectiveness versus efficacy: The limited value in clinical practice of high dose dexamethasone suppression testing in the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. *Journal of Clinical Endocrinology & Metabolism, 82*, 1780–1785.
- Backer, T. E. (2000). *Finding the balance: Program fidelity and adaptation in substance abuse prevention*. Washington, DC: Center for Substance Abuse Prevention.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9), 6–8, 14.
- Bauman, L. J., Stein, R. E., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology, 19*, 619–639.
- Becker, D. R., Smith, J., Tanzman, B., Drake, R. E., & Tremblay, T. (2001). Fidelity of supported employment programs and employment outcomes. *Psychiatric Services, 52*, 834–836.
- Ben-Peretz, M. (1975). The concept of curriculum potential. *Curriculum Theory Network, 5*(2), 151–159.
- Ben-Peretz, M. (1990). *The teacher–curriculum encounter: Freeing teachers from the tyranny of texts*. Albany: State University of New York Press.
- Ben-Peretz, M., Katz, S., & Silberstein, M. (1982). Curriculum interpretation and its place in teacher education programs. *Interchange, 13*(4), 47–55.
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *Educational Forum, 40*, 345–370.
- Biglan, A., & Taylor, T. K. (2000). Increasing the use of science to improve child-rearing. *Journal of Primary Prevention, 21*, 207–226.
- Blakely, C. C., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253–268.
- Bond, G. R., Becker, D. R., Drake, R. E., & Vogler, K. M. (1997). A fidelity scale for the individual placement and support model of supported employment. *Rehabilitation Counseling Bulletin, 40*, 265–284.
- Boruch, R. R., & Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluation. *Professional Psychology, 8*, 411–433.
- Cho, J. (1998, April). *Rethinking curriculum implementation: Paradigms, models, and teachers' work*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

- Connell, D. B., Turner, R. R., & Mason, E. F. (1985). Summary of findings of the School Health Education Evaluation: Health promotion effectiveness, implementation, and costs. *Journal of School Health, 55*, 316–321.
- 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*, 23–45.
- Datnow, A., & Castellano, M. E. (2001). Managing and guiding school reform: Leadership in Success for All Schools. *Educational Administration Quarterly, 37*, 219–249.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*, 433–479.
- Dobson, K. S., & Shaw, B. F. (1988). The use of treatment manuals in cognitive therapy: Experience and issues. *Journal of Consulting and Clinical Psychology, 56*, 673–680.
- Dobson, L. D., & Cook, T. J. (1980). Avoiding Type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning, 3*, 269–276.
- Domitrovich, C., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs for school-aged children. *Journal of Educational and Psychological Consultation, 11*, 193–221.
- Dorland, W. A. (1994). *Dorland's illustrated medical dictionary* (28th ed.). Philadelphia: W. B. Saunders.
- Dumas, J., Lynch, A., Laughlin, J., Smith, E., & Prinz, R. (2001). Promoting intervention fidelity: Conceptual issues, methods and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(1S), 38–47.
- 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 Theory and Practice, 18*, 237–256.
- Duyar, I. (2006, August 26). Analyzing education productivity: An essay review. *Education Review, 9*(4). Retrieved September 8, 2006, from <http://edrev.asu.edu/essays/v9n4>
- Emshoff, J. G., Blakely, C., Gottschalk, R., Mayer, J., Davidson, W. S., & Erickson, S. (1987). Innovation in education and criminal justice: Measuring fidelity of implementation and program effectiveness. *Educational Evaluation and Policy Analysis, 9*, 300–311.
- Ennett, S. T., Tobler, N. S., Ringwalt, C., & Flewellin, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project D.A.R.E. outcome evaluations. *American Journal of Public Health, 84*, 1394–1401.
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy, 36*, 3–13.
- Fox, L. H. (1977). Sex differences: Implications for program planning for the academically gifted. In J. C. Stanley, W. C. George, & C. H. Solano (Eds.), *The gifted and the creative: A fifty-year perspective* (pp. 113–138). Baltimore: John Hopkins University Press.
- Freeman, H. E. (1977). The present status of evaluation research. In M. Guttentag (Ed.), *Evaluation Studies Review Annual II* (pp. 17–51). Beverly Hills, CA: Sage.
- Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal, 101*, 495–510.

- Fullan, M. (2001). *The meaning of educational change*. New York: Teachers College Press.
- Fullan, M., & Pomfret, A. (1977). Research on curriculum instruction implementation. *Review of Educational Research, 47*, 335–397.
- Gagne, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design*. Belmont, CA: Wadsworth.
- Goodman, R. M., & Steckler, A. (1989). A model for the institutionalization of health promotion programs. *Family and Community Health, 11*(4), 63–78.
- Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22*, 254–272.
- Hall, G. E., & Loucks, S. F. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal, 14*, 263–276.
- Hall, G. E., & Loucks, S. F. (1978). *Innovation configurations: Analyzing the adaptations of innovations*. Austin: University of Texas, Research and Development Center for Teacher Education.
- Hall, G. E., Wallace, R. C., & Dossett, W. D. (1973). *A developmental conceptualization of the adoption process within educational institutions*. Austin: University of Texas, Research and Development Center for Teacher Education.
- Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the Adolescent Alcohol Prevention Trial. *Journal of Studies on Alcohol, 52*, 568–579.
- Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., & Fleming, C. B. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology, 27*, 711–731.
- Hedges, L. (2004, April). *Designing studies for evidence-based scale up in education*. Invited address presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the “noise” of real life in study design. *American Journal of Psychiatry, 159*, 201–207.
- Hopkins, R. H., Mauss, A. L., Kearney, K. A., & Weisheit, R. A. (1988). Comprehensive evaluation of a model alcohol education curriculum. *Journal of Studies on Alcohol, 49*, 38–50.
- Hord, S. M., Rutherford, W. L., Huling-Austin, L., & Hall, G. E. (1987). *Taking charge of change*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hughes, T. A., Frederick, L. D., & Keel, M. C. (2002). Learning to effectively implement constant time delay procedures to teach spelling. *Learning Disability Quarterly, 25*, 210–222.
- Kimpton, R. D. (1985). Curriculum fidelity and the implementation tasks employed by teachers: A research study. *Journal of Curriculum Studies, 17*, 185–195.
- Kolbe, L. J., & Iverson, D. C. (1981). Implementing comprehensive school health education: Educational innovations and social change. *Health Education Quarterly, 8*, 57–80.
- Kutash, K., Duchnowski, A. J., Sumi, W. C., Rudo, Z., & Harris, K. M. (2002). A school, family, and community collaborative program for children who have emotional disturbances. *Journal of Emotional and Behavioral Disorders, 10*(2), 99–108.

- Latimer, E. A. (1999). Economic impacts of assertive community treatment: A review of the literature. *Canadian Journal of Psychiatry, 44*, 443–455.
- Lee, O. (2002). Science inquiry for elementary students from diverse backgrounds. In W. G. Secada (Ed.), *Review of research in education* (Vol. 26, pp. 23–69). Washington, DC: American Educational Research Association.
- Leithwood, K. A., & Montgomery, D. J. (1980). Evaluating program implementation. *Evaluation Review, 4*, 193–214.
- Lindblom, C. (1959). The science of “muddling through.” *Public Administration Review, 19*, 79–88.
- Lipsey, M. W. (1999). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy & the Law, 6*, 611–641.
- Loucks, S. F. (1983, April). *Defining fidelity: A cross-study analysis*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Lyman, D. R., Zimmerman, R., Novak, S. P., Logan, T. K., Martin, C., Leukefeld, C., et al. (1999). Project D.A.R.E.: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology, 67*, 590–593.
- March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics, 9*, 587–608.
- McDonald, S. K., Keesler, V. A., Kaufman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher, 35*(3), 15–24.
- McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health model. *Journal of Consulting and Clinical Psychology, 62*, 670–678.
- McHugo, G. J., Drake, R. E., Teague, G. B., & Xie, H. (1999). Fidelity to assertive community treatment and client outcomes in the New Hampshire Dual Disorders Study. *Psychiatric Services, 50*, 818–824.
- McLaughlin, M. W., & Marsh, D. D. (1978). Staff development and school change. *Teachers College Record, 80*, 69–93.
- Mihalic, S. (2002, April). *The importance of implementation fidelity*. Boulder, CO: Center for the Study and Prevention of Violence.
- Mihalic, S., Irwin, K., Fagan, D. B., & Elliott, D. (2004, July). Successful program implementation: Lessons from Blueprints (NCJ 204273). *Juvenile Justice Bulletin*. Retrieved October 2, 2005, from <http://www.ojjdp.ncjrs.gov/publications/PubResults.asp>
- Mills, S. C., & Ragan, T. J. (2000). A tool for analyzing implementation fidelity of an integrated learning system. *Educational Technology Research and Development, 48*(4), 21–41.
- Mitchel, M. E., Hu, T.-W., McDonnell, N. S., & Swisher, J. D. (1984). Cost-effectiveness analysis of an educational drug abuse prevention program. *Journal of Drug Education, 14*, 271–291.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247–266.
- Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340.
- Munby, H., & Russell, T. (1990). Metaphor in the study of teachers’ professional knowledge. *Theory Into Practice, 29*, 116–121.

- National Research Council Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials, Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.
- No Child Left Behind Act, Title IX, General Provisions, Part A, Section 9101, 20 U.S.C. § 7801 et seq. (2001). Retrieved December 20, 2006, from <http://www.ed.gov/policy/elsec/leg/esea02/pg107.html>
- O'Donnell, C. L. (2007). Fidelity of implementation to instructional strategies as a moderator of curriculum unit effectiveness in a large-scale middle school science quasi-experiment. *Dissertation Abstracts International*, 68(08). (UMINo. AAT 3276564)
- Orwin, R. G. (2000). Methodological challenges in study design and implementation: Assessing program fidelity in substance abuse health services research. *Addiction*, 95, S309–S327.
- Patton, M. Q. (1978). *Utilization—Focused evaluation*. Beverley Hills, CA: Sage.
- Penuel, R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: Analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching*, 41, 294–315.
- Price, R. H., Friedland, D. S., Choi, J., & Caplan, R. D. (1998). Job-loss and work transitions in a time of global economic change. In X. Arriaga & S. Oskamp (Eds.), *Addressing community problems: Psychological research and interventions* (pp. 195–222). Thousand Oaks, CA: Sage.
- Raudenbush, S. W. (2007). *Designing field trials of educational innovations*. In B. Schneider & S.-K. McDonald (Eds.), *Scale-up in education: Vol. II. Ideas in Practice* (pp. 23–40). Lanham, MD: Rowman & Littlefield.
- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75, 211–246.
- Resnick, B., Bellg, A. J., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., et al. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine*, 29, 46–54.
- Resnicow, K., Cohn, L., Reinhardt, J., Cross, D., Futterman, R., Kirschner, E., et al. (1992). A three-year evaluation of the "Know Your Body" program in inner-city schoolchildren. *Health Education Quarterly*, 19, 463–480.
- Rogers, E. (2003). *Diffusion of innovations*. New York: Free Press.
- Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*, 22, 237–260.
- Ruiz-Primo, M. A. (2005, April). A multi-method and multi-source approach for studying fidelity of implementation. In S. Lynch (Chair) & C. L. O'Donnell, "Fidelity of implementation" in *implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Scheirer, M. A., & Rezmovic, E. L. (1983). Measuring the degree of program implementation: A methodological review. *Evaluation Review*, 7, 599–633.
- Schneider, R. M., Krajick, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching*, 42, 283–312.

- Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), *Evaluation studies review annual* (Vol. 4, pp. 15–35). Thousand Oaks, CA: Sage.
- Shkedi, A. (1998). Can the curriculum guide both emancipate and educate teachers? *Curriculum Inquiry*, 28, 209–231.
- Shulman, L. (1990). Foreword. In M. Ben-Peretz, *The teacher–curriculum encounter: Freeing teachers from the tyranny of texts* (pp. vii–ix). Albany: State University of New York Press.
- Slavin, R. E. (2003). A reader’s guide to scientifically based research: Learning how to assess the validity of education research is vital for creating effective, sustained reform. *Educational Researcher*, 60(5), 12–16.
- Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402–435). New York: Macmillan.
- Songer, N. B., & Gotwals, A. W. (2005, April). Fidelity of implementation in three sequential curricular units. In S. Lynch (Chair) & C. L. O’Donnell, “Fidelity of implementation” in *implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Stringfield, S., & Datnow, A. (1998). Scaling up school restructuring designs in urban schools. *Education and Urban Society*, 30, 269–276.
- Summerfelt, W. T. (2003). Program strength and fidelity to evaluation. *Applied Developmental Science*, 7(2), 55–61.
- Summerfelt, W. T., & Meltzer, H. Y. (1998). Efficacy vs. effectiveness in psychiatric research. *Psychiatric Services*, 49, 834–835.
- Taggart, V. S., Bush, P. J., Zuckerman, A. E., & Theiss, P. K. (1990). A process evaluation of the District of Columbia “Know Your Body” project. *Journal of School Health*, 60(2), 60–66.
- Teague, G. B., Bond, G. R., & Drake, R. E. (1998). Program fidelity in assertive community treatment: Development and use of a measure. *American Journal of Orthopsychiatry*, 68, 216–232.
- Teague, G. B., Drake, R. E., & Ackerson, T. H. (1995). Evaluating use of continuous treatment teams for persons with mental illness and substance abuse. *Psychiatric Services*, 46, 689–695.
- Tortu, S., & Botvin, G. J. (1989). School-based smoking prevention: The teacher training process. *Prevention Medicine*, 18, 280–289.
- Towne, L., Wise, L. L., & Winters, T. M. (Eds.), for the Committee on Research in Education, National Research Council. (2005). *Advancing scientific research in education*. Washington, DC: National Academies Press.
- Turner, J. H. (Ed.). (1989). *Theory building in sociology: Assessing theoretical culmination*. Newbury Park, CA: Sage.
- U.S. Department of Education. (2003a, December). *Identifying and implementing educational practices supported by rigorous evidence: A user-friendly guide*. Washington, DC: Institute of Education Sciences. Retrieved August 29, 2005, from <http://www.ed.gov/print/rschstat/research/pubs/rigorousvid/guide.html>
- U.S. Department of Education. (2003b). *What Works Clearinghouse study review standards*. Washington, DC: Institute of Education Sciences. Retrieved January 10, 2005, from <http://ies.ed.gov/ncee/wwc/overview/review.asp>

- U.S. Department of Education. (2006). *Education research grant request for application: CFDA Number 84.305*. Washington, DC: Institute for Education Sciences. Retrieved May 15, 2006, from <http://ies.ed.gov/ncer/funding>
- U.S. Department of Health and Human Services. (2002). *Finding the balance: Program fidelity and adaptation in substance abuse prevention, a state-of-the-art review*. Washington, DC: Federal Center for Substance Abuse Prevention.
- Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., et al. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*, 449–487.
- Vincus, A. (2004, Spring). Assessing fidelity. *ALERT Educator*. Retrieved July 10, 2005, from <http://www.projectalert.com>
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall.
- White, M. C., Tulskey, J. P., Menendez, E., Arai, S., Goldenson, J., & Kawamura, L. M. (2005). Improving tuberculosis therapy completion after jail: Translation of research into practice. *Health Education Research*, *20*, 163–174.
- Woodward, J., & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in middle grade mathematics. *Journal of Special Education*, *40*, 151–159.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, *49*, 156–167.
- Ysseldyke, J., Spicuzza, R., Kosciulek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8*, 247–265.

Author

CAROL L. O'DONNELL was a research scientist at The George Washington University at the time of this writing and can be reached at The George Washington University, Department of Physics, Corcoran Hall, Room 105, 725 21st Street, N.W., Washington, DC, 20051 (202) 208-3749; e-mail codonnel@gwu.edu. She is a science educator whose research interests include science education reform, curriculum development, implementation, and student learning.