A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge

Jeanne Century, Mollie Rudnick and Cassie Freeman

American Journal of Evaluation 2010 31: 199 originally published online 11 May 2010
DOI: 10.1177/1098214010366173

American Journal of Evaluation
31(2) 199-218

DOI: 10.1177/1098214010366173
http://aje.sagepub.com

A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge

Jeanne Century[1], Mollie Rudnick[1], and Cassie Freeman[1]

Abstract

There is a growing recognition of the value of measuring fidelity of implementation (FOI) as a necessary part of evaluating interventions. However, evaluators do not have a shared conceptual understanding of what FOI is and how to measure it. Thus, the creation of FOI measures is typically a secondary focus and based on specific contexts and programs. This article describes a project that holds the development of FOI measures as its primary goal and has developed a suite of data collection tools designed to be used across multiple programs. It describes the foundation of the suite—a conceptual framework for clearly and specifically describing FOI and the need for the framework. It also describes where the framework resides in existing literature and how it can be used to support measurement of interventions in education and other fields.

Keywords

fidelity of implementation, measuring use, measuring effectiveness, instructional materials

"Does an intervention work?" is one of the fundamental questions that drives evaluation. Although we all may engage with different approaches, in the end, we want to know whether a program did work, is working, or can work. The simplicity of these basic questions is deceptive. To answer them fully and usefully, we must also ask some fundamental and increasingly complex questions such as "what is the program?" "what does program implementation look like?" and "has the program changed from the original intent?" These questions move us from merely knowing *if* a program works toward understanding *why*, *how*, and *under what conditions*? And all point to what has long been the "black box" of evaluating effectiveness of interventions: fidelity of implementation (FOI).

[1]Center for Elementary Mathematics and Science Education, Physical Sciences Division, University of Chicago, IL, USA

Corresponding Author:
Jeanne Century, Center for Elementary Mathematics and Science Education, Physical Sciences Division, University of Chicago, 5640 South Ellis, EFI Box 15, Chicago, IL 60637, USA
Email: jcentury@uchicago.edu

Today's political environment with expectations for increasingly rigorous evaluations is calling attention to what Petersen (as cited in Gresham, Gansle, & Noell, 1993) called a "curious double standard." When particular behaviors and activities are dependent variables or outcomes, they are rigorously defined and measured. However, when those same behaviors serve as the treatment, or independent variable, valid and reliable measures are rarely offered. Without specific, clear measurement of implementation, it is impossible to know whether disappointing outcomes are due to an inadequate program model of change or due to poor or incomplete implementation (Fullan, 1983; Lynch & O'Donnell, 2005; Ruiz-Primo, 2005; Wang, Nojan, Strom, & Walberg, 1984).

Still, despite the growing recognition of the value of understanding FOI, we, as evaluators, have yet to develop a shared conceptual understanding of what FOI is and how to measure it. Thus, we create FOI measures based on their particular contexts and programs of interest, leaving the field with a collection of disparate measures and ad hoc theories about FOI. With no shared basis for measuring and discussing FOI, we are unable to compare findings across studies of particular interventions or accumulate knowledge on FOI itself.

The Center for Elementary Mathematics and Science Education (CEMSE) at the University of Chicago is addressing this problem with work that focuses on the development of FOI measures as its primary goal. For 3 years, CEMSE has received support from the National Science Foundation for its "Applied Research on Science Materials Implementation: Bringing Measurement of Fidelity of Implementation to Scale" project. Through this project, the CEMSE team has produced common instruments for measuring FOI of five science and mathematics instructional materials programs[1] as well as a *User's Guide* that describes procedures for using the instruments and adapting them for use with other instructional materials.

Because this project sought to create measures that could be used across multiple programs, it required that we begin with a sound conceptual framework for clearly and specifically describing FOI that could serve as a foundation for FOI measurement. As the project proceeded, it became evident that the conceptual framework we developed and its approach using "critical components" as measurable constructs was applicable to instructional materials in other subject areas and it appeared to be applicable, with some adaptation, to other kinds of interventions in education and other fields. We came to see that this framework and approach to FOI could be a useful foundation for FOI measurement in many contexts. This article focuses on the development of that framework, where the framework resides in existing FOI literature and how the framework and the instruments derived from it can help evaluators in a range of fields systematically and rigorously measure FOI.

Building on Previous Work

The literature on FOI resides in a range of areas including mental health, prevention, education, at-risk youth, criminal justice, health, manufacturing, organizational change, public policy, software development, technology, and transportation. Early conceptualizations of FOI drew important attention to the study of program use but characterized implementation in broad terms (Fullan & Pomfret, 1977; Hall & Loucks, 1977). Since then, researchers, particularly those working in health, have developed more refined approaches for characterizing and assessing FOI that account for the fact that it is complex and multidimensional (Fullan, 1983; Gersten & Carnine, 1980; Wang et al., 1984). This literature provided us with several starting points for developing our conceptual framework.

Five Dimensions Approach

Many efforts to develop FOI measures reference a seminal piece by Dane and Schneider (1998). In this piece, the authors review studies of primary and early secondary prevention programs for

measures of what they refer to as "program integrity" (what we refer to as FOI). They found that only 39 of the 162 studies measured program integrity and that in those studies that did, the authors defined it along one or more of five dimensions: adherence, exposure, quality of delivery, participant responsiveness, and program differentiation. The identification of these dimensions by Dane and Schneider, however, was only a first step in moving toward more coherence in FOI measurement. They wrote, "five aspects of fidelity have been identified in the literature ... though the definitions and labels assigned to these aspects vary considerably and are often not consistent with the terms used in the present review"(p. 39). Although they offered provisional definitions (detailed later in this article), they suggested that the adoption of uniform definitions would be necessary to move forward. As we looked at these dimensions, we saw that while they represent a range of FOI components, the dimensions themselves are incommensurable. Thus, while recognizing the need to account for the dimensions in our framework, we did not use them as a starting point.

Critical Components

Another approach to describing implementation focuses on program components. Hall and Hord (1987), for example, note that to analyze different instantiations of a program, "the *components* or building blocks of the innovation must be identified" (p. 117). Others refer to these essential features of programs as "critical program dimensions," "model dimensions," "fidelity criteria," "essential characteristics," and "critical parts" (Bond et al., 2000; Huntley, 2005; Mowbray, Holter, Teague, & Bybee, 2003; Sabelli & Dede, 2001; Wang et al., 1984). Notwithstanding the different names, the works suggest that programs consist of essential features that must be measured to determine whether a program is present or not. We agreed, and adopted the phrase "critical components" (Bauman, Stein, & Ireys, 1991; Ruiz-Primo, 2005; Wang et al., 1984) to refer to these essential program elements.

Structure and Process Approach

With a critical component approach in mind, we looked at the work of others who used critical components and considered those works in the context of an approach that we (and others) characterized as "structure and process." For example, Mowbray et al. (2003) discuss "fidelity criteria" and organize them into two groups: "framework for service delivery" (structure) and "the ways in which services are delivered" (process). Similarly, Wang et al. (1984) looked at the program elements "related to resources" (structural) and those related to "roles and behaviors" (process). This work and the work of others illustrate that vital aspects of programs reside in both composition (structure) and in the human interactions that take place during delivery (process).

Combining Frameworks Approach

We also looked at ways others had adopted combinations of the dimensions of fidelity, critical component, and structure–process approaches. Ruiz-Primo (2005), for example, explored implementation of a program by developing a matrix-like framework of the five dimensions and structure/process elements. Lynch and O'Donnell (2005) also combined the structure and process framework with the five dimensions of fidelity by Dane and Schneider (1998). In their approach, rather than connect them using a matrix (like Ruiz-Primo), they embedded the dimensions within the structure–process framework. In doing so, they associated *adherence*, *exposure*, and *differentiation* with "structure" and *quality* with "process." Then, in addition to structure and process, they created a third category that they called "self-perceived effects of participants" and placed *participant responsiveness* there.

Although we benefited from and built on the thinking of those who combined frameworks, we did not adopt their approaches. Ruiz-Primo's approach built on the work of Dane and Schneider and,

while we knew that work to be conceptually interesting and informative, because the dimensions were incommensurate with one another, they were not a sound starting point for a coherent FOI framework. In addition, while the approach we took was similar to that of Lynch and O'Donnell (2005) and Lastica and O'Donnell (2007), we expanded on their work by creating an organizational framework that could apply to multiple programs and delineating the structural and process domains into more specific categories.

These most recent approaches that combine critical components, the five-dimension approach, and the structure–process approach suggest that there are many commonalities in FOI work. However, there has been no attempt to explore the extent to which any approach developed for a particular intervention can work conceptually with other interventions. Our work has taken this step by drawing from all three approaches in the past to create a conceptual framework that can be applied across multiple programs and operationalizing it into a suite of instruments that measure FOI. It is our hope that this work can respond to the recommendations of past works that call for increased clarity in FOI measures, specific descriptions of interventions, and the development of a theoretical framework (Dane & Schneider, 1998; Leithwood & Montgomery, 1980; O'Donnell, 2008).

FOI Conceptual Framework

Throughout the framework development process, we were guided by our goal to enable rigorous, specific, and systematic analysis of intervention use by providing useful measurement categories in a sound organizational structure. This was an iterative process, simultaneously informed by the critical component and structure–process approaches to FOI measurement. We kept the dimensions of Dane and Schneider in mind as well, both because they captured elements that many others considered important in FOI measurement and because the dimensions themselves were useful ideas, even if not for organizing the overall framework. In the section that follows, we describe the framework development process that took place iteratively with the critical component identification process and evolved with the pilot and field tests of our FOI instruments.

Defining FOI

To begin, it was important to articulate a definition of FOI. Although there is no agreement in the field on an exact definition of FOI, there is a general consensus about what FOI is—the relationship between an intended and an enacted program. We chose to represent this consensus with the following basic definition: *The extent to which an enacted program is consistent with the intended program model*.

To elaborate on this definition, we looked at others' operational definitions of FOI and saw that many suggested that implementation could be determined by identifying whether critical components of a program are present or not (Bond et al., 2000; Huntley, 2005; Mowbray et al., 2003; Sabelli & Dede, 2001; Wang et al., 1984). With further consideration, we saw that critical components of program models are the operationalizations of developers' intervention theories and comprise the variables one must measure to determine a program's FOI. Given this, we operationalized our FOI definition by rewording it as *the extent to which the critical components of an intended program are present when that program is enacted*. This definition became the starting point for an iterative process of critical component identification and framework development that led to our conceptual framework structure.

Defining Critical Components

As we reviewed the literature and developed various iterations of a FOI framework, we began the process of critical component identification for the five mathematics and science instructional

materials programs we were focused on. Leithwood and Montgomery (1980) suggest that information about the critical components of a program should be taken from the program developers, written materials produced by the developers, and those involved in the implementation of the program. Our process included all three of these sources.

Written materials review. We began with a careful reading of all five programs' written materials, including but not limited to teacher's guides, student books, assessment guidelines, and student journals. As we reviewed them, we identified all of the possible program elements that could have been critical components, noting that because developers do not always have a clearly articulated program model, we needed to include not only the critical components that were explicit in the materials but also those that seemed to be implicit. After completing the initial lists of critical components for each program, we met to arrive at consensus about which critical components were common across the programs and which were unique and then began to organize them into the evolving framework (described below).

Program developers. Although explained here as a separate sequential step in this process, we interacted with the developers of the five programs from the beginning of the project. Our initial discussions with the developers revealed that although they held strong beliefs about some elements of their programs, they did not necessarily share an explicit single program model nor did they necessarily share understandings of what constituted a faithful implementation.

We provided the developers with a description of the FOI conceptual framework and its categories and examples of critical components in each of those categories. Then, we asked them to complete a critical component worksheet that included creating a critical component list, placing the critical components in the FOI framework, and suggesting strategies for measurement. On receiving their lists, we reconciled the developers' critical components with our own, verifying those that matched, identifying those that we agreed were not aligned with ours, and accounting for them as we revised the framework. This process was informed by additional conversations with the developers to ensure we were correctly interpreting their work.

Users. A third step in the critical component identification process was a review by users—in this case, teachers. Although the first two parts of the process focused on refining the critical component list so that it best reflected the "intended" program model, this step brought a new perspective to the critical component identification process. Developers' critical components reflected their intentions about the program elements that they believed *would* lead to desired student outcomes; the teachers' critical component lists were informed by their experiences and reflected their perceptions of the enacted elements that they believe *did* lead to the desired outcomes. We reviewed the teachers' lists, but they did not offer any additional information that warranted a revision of the framework or critical component list.

Challenges in identifying critical components. We faced several challenges during the critical component identification process that are likely to be of interest to others engaging in this process. First was the fact that in the absence of empirical data, we faced the recurring exercise of determining which elements were "critical" and which were not. Although the materials we looked at are explicit about some developer intentions, they only allude to the importance of others. For example, this suggestion to a teacher from the materials—"students are often interested in ... so make a connection to ..." implies the importance of a critical component such as "teacher builds on/stimulates student interest" but it is never explicitly articulated as an essential program element.

Second, we realized that we could not assume that a program element's explicit inclusion in the materials warranted a conclusion that program element was "critical." For example, we identified

"multi-sensory learning" as a possible critical component. However, although this term was clearly stated in the materials, there was little other evidence (e.g., few references to it, few instructions to the teacher regarding it) that the developers considered it "critical."

Third, throughout this process, we had to address "grain size" issues that grew from determining the best level of specificity for a single critical component. For example, when considering a critical component broadly named "discussion," it was clear that it encompassed too many aspects of instruction; the grain size was too large. However, a critical component named "teacher gives wait time during discussion" was too narrow because it captured only one aspect of discussion; the grain size was too small. In this case, ultimately, we decided to call the critical component, "teacher facilitation of student discussion." This grain size lent itself to a range of indicators (including the more specific, observable behavior of "teacher gives wait time during discussion") but was not so broad that it lost meaning and utility. These decisions were empirically tested through confirmatory factor analysis conducted with data from the instrument pilot and field tests.

Another grain size issue came in the challenge of finding a balance between accounting for the differences across specific program structures and the need to succinctly capture critical components in our framework that could apply to multiple programs. For example, one of the critical components is called "unit- or lesson-level background information on pedagogy." This critical component name evolved because one program had vignettes illustrating instruction; another had a section on "instructional pedagogies;" and still another had a section on "orchestrating the active class." We recognized that all three of these (and others) were designed to provide unit- or lesson-level background information on pedagogy. Thus, we created a superordinate title for the critical component with the expectation that when operationalized, each instrument would be customized for the program by naming the specific program elements that were subsumed within the critical component. This process of recognizing the common critical components across programs was key to applying the framework to multiple programs and laid a foundation for our abilities to apply this process to other interventions.

And finally, throughout the critical component identification process, we were confronted with the fact that there are many variables that are not program critical components but are important to measure because they affect program implementation and outcomes in various ways. In the case of an instructional intervention, these variables might include characteristics of the teacher such as years teaching, experience with the program, and amount of professional development. Although these have a significant impact on use, they are not actually part of FOI itself. Thus, we eventually made a decision that where possible, we would include items to gather information on these moderating variables as long as that data collection did not interfere with the collection of data on our primary constructs of interest—the critical components.

Organizing the FOI Framework

Using our working list of critical components and the general structure and process approach, we began to develop the FOI conceptual framework. After several iterations, we defined two broad organizational categories: (a) structural critical components and (b) instructional critical components. While aligning with Mowbray et al. (2003) and the later work by Lastica and O'Donnell (2007), we elaborated on the broad categories of structure and process by identifying specific subcategories (i.e., procedural, educative, pedagogical, and student engagement). Table 1 shows the basic FOI framework.

Structural critical components reflect the developers' intentions about the design and organization of the intervention itself, in our case, instructional materials. *Instructional* critical components, however, reflect the developers' intentions about the participants' (in our case, teachers and students)

Table 1. Basic FOI Framework

Fidelity of Implementation			
Categories of Critical Components			
Structural Critical Components		Instructional Critical Components	
Procedural	Educative	Pedagogical	Student engagement

behaviors and interactions as they enact the intervention. Then, each main category has subcategories that further categorize the critical components.

Structural Critical Components

Structural–Procedural Critical Components

These are the organizing elements of the intervention that communicate to the user in the simplest sense, *what to do*. Acknowledging that all parts of an intervention might implicitly or explicitly suggest what the user should do, the critical components in this category focus on the basic steps of the procedures and the ways the intervention is physically organized to communicate intentions to the user.

Structural–Educative Critical Components

These components represent the developers' expectations for what the user needs to *know* and reflect the developers' understandings that users need a particular body of knowledge to enact the intervention as intended. In the case of instructional materials, for example, teachers may need a basic level of content and pedagogical knowledge to enact a program and, while some teachers may come to the classroom with that knowledge, others may not. Thus, structural–educative critical components are analogous to built-in professional development or training and reflect the developers' intentions about how to structure and organize that information for the users. For some interventions, they are represented in the actual professional development or training components of the intervention.

Instructional Critical Components

Instructional–Pedagogical Critical Components

These critical components represent the actions, behaviors, and interactions that the user is expected to engage in when enacting the intervention, including the user's interactions with the participants/recipients. In the case of instructional materials, these components reflect the intentions about the teachers' behaviors, strategies, and interactions with students during the instructional transactions in the classroom.

Instructional–Student Engagement Critical Components

These critical components represent the actions, behaviors, and interactions the recipient is expected to engage in when participating in the enactment of the intervention. In the case of instructional materials, they reflect expectations about students' participation in the instructional transactions; in other cases, they represent expectations for the behaviors of any program recipient/participant such as community members, health center leaders, or people with illnesses. To see how the framework is operationalized with the critical components present in the instructional materials programs

Table 2. Fidelity of Implementation (FOI) of Instructional Materials

	Structural Critical Components		Instructional Critical Components	
	Procedural	Educative	Pedagogical	Student Engagement
Common to mathematics and science	*Time* 1. Duration of unit 2. Time spent on instruction *Order* 3. Investigation/lesson order 4. Order of segments and activities within lesson *Inclusion* 5. Inclusion of all essential segments within a lesson 6. Inclusion of all essential lessons *Pre-lesson* 7. Lesson overview 8. Lesson preparation *Essential program elements* 9. Materials presence 10. Writing structures 11. Readings 12. Assessments and assessment tools 13. Content of lesson a. Procedures b. Facts c. Concepts d. Processes 14. Use of class structures 15. Use of instructional delivery formats *Non-essential program elements* 16. Projects* 17. Extensions* a. Discipline-related b. Non-discipline-related 18. Additional resources*	1. Content background information 2. Pedagogy background information 3. National standards and benchmarks information* 4. Lesson notes	*Facilitating student engagement with others* 1. Teacher facilitation of group work 2. Teacher facilitation of student discussion *Facilitating student engagement with content* 3. Teacher facilitation of students doing potentially intellectually challenging work 4. Teacher emphasis on types of content *Facilitating student role as learner* 5. Teacher facilitation of student autonomy 6. Teacher facilitation of students taking risks 7. Teacher facilitation of student interest *Pedagogical strategies* 8. Teacher facilitation of materials, manipulatives and tools 9. Teacher use of assessment to inform instruction 10. Teacher use of differentiation	*Students engage with others* 1. Students contribute to small group work 2. Students engage in discussion *Students engage with content* 3. Students do potentially intellectually challenging work 4. Students develop role as learner 5. Students demonstrate autonomy 6. Students take risks *Students use the materials* 7. Students do/complete essential activities 8. Students do/complete optional or non-essential activities
Common to science		A. Safety	A. Teacher coordination of reading and science instruction B. Teacher facilitation of observation and data collection	A. Students observe and collect data

Table 3. The Center for Elementary Mathematics and Science Education (CEMSE) Fidelity of Implementation (FOI) Framework Aligned With Others' Work

	FOI of Instructional Materials (1. *Adherence*)			
	Categories of Critical Components			
	Structural Critical Components (6. Structure)		Instructional Critical Components (6. Process)	
Categories of Differentiation (5. *Differentiation*)	Procedural (2. *Exposure*, *Dosage*)	Educative	Pedagogical (3. *Quality*)	Student Engagement (4. *Responsiveness*, *Participation*)
Common Across Interventions				
Unique to Interventions				

we reviewed, see Table 2. This table shows all of the critical components that were common across all five programs as well as those that were identified as unique only to science programs.

Aligning the FOI Framework With Previous Work

Now that we have described the CEMSE FOI framework, we can take a closer look at how it is aligned with and improves on the approaches to measuring FOI described earlier. In Table 3 above, our framework categories are in bold type while others' are numbered and italicized.

1. *Adherence*. Dane and Schneider (1998) define adherence as "the extent to which specified program components were delivered as prescribed in program manuals" (p. 45). Others embrace this definition and essentially equated adherence with implementation (Dusenbury, Brannigan, Falco, & Hansen, 2003; Lynch & O'Donnell, 2005; Ruiz-Primo, 2005) and sometimes use the two words—adherence and fidelity—interchangeably (Dusenbury et al., 2003; Lynch & O'Donnell, 2005). Thus, it was clear that adherence is not a *dimension* of FOI; rather, it is essentially the *same as* FOI. Indeed, the definition of Dane and Schneider is quite close to our operationalized definition of FOI (*the extent to which the critical components of an intended program are present when that program is enacted*) that itself was derived from the generally accepted definition of FOI. Our conclusion, then, was that adherence should not be considered a dimension of fidelity because it is synonymous with the generally accepted broader definition of fidelity.
2. *Exposure and dosage*. Exposure was also only loosely defined in the study of Dane and Schneider (1998). They define it as "an index that may include any of the following: (a) the number of sessions implemented; (b) the length of each session; or (c) the frequency with which program techniques were implemented" (1998, p. 45). Dusenbury et al. refer to "dose" but define it equally loosely, as "the amount of program content received by participants" (2003, p. 241). Although the authors indicate the need for a more clear definition of what exposure is, some researchers have retained the definition of Dane and Schneider as a dimension of measurement even though it is insufficiently clear and specific. We decided to include the specific elements of exposure and dosage (e.g., time spent, frequency of sessions) in our framework but measure them as separate critical components in the structural–procedural category.

3. *Quality*. Dane and Schneider (1998) define their third dimension, quality of delivery, as "a measure of *qualitative aspects* of program delivery that are not directly related to the implementation of prescribed content, such as implementer enthusiasm, leader preparedness, global estimates of session effectiveness, and leader attitudes toward program"(p. 45). On the face of it, this definition encompasses very different and broad constructs ranging from the user's affect to his or her expertise. The word "quality" was adopted by others but defined more specifically as approaching a pedagogical or theoretical ideal (Dusenbury et al., 2003; Lynch & O'Donnell, 2005; Ruiz-Primo, 2005) thus eliminating some elements found in the review of Dane and Schneider. Our approach to quality comes closer to the latter usage in that we recognize that intervention developers have a "theoretical ideal" with regard to the enactment of the program and we measure that in the instructional–pedagogical category. We do, however, recognize the importance of the contextual factors that Dane and Schneider refer to and indeed measure some of them in our instruments, but as the moderating variables described above, not as critical components of FOI.

It is worth noting, however, that the definition of what is contextual and what is part of the intervention can always shift, depending on the boundaries drawn around the intervention itself. For example, if an intervention developer defines as one of its essential elements particular leader behaviors, those then become critical components to measure within the boundaries of the FOI framework. If those behaviors are not identified as components of the intervention, however, they then become moderating variables that reside outside of the FOI critical component structure. This highlights the fact that FOI measurement can sometimes be a "moving target" as the definition of the program model shifts or evolves over time.

4. *Participant responsiveness*. Dane and Schneider (1998) describe participant responsiveness as "a measure of participant response to program sessions, which may include indicators such as levels of participation and enthusiasm" (p. 45). Others refer to this as the extent of participant/student participation or engagement (Dusenbury et al., 2003; Lynch & O'Donnell, 2005; Ruiz-Primo, 2005). This dimension of FOI recognizes that some critical components essential for implementation reside not with the user or implementer of the intervention but with the participants or recipients. For example, it is insufficient to determine FOI of the programs we looked at by only measuring the instructional–pedagogical critical program component called "teacher facilitation of student discussion." One must also measure the instructional–student engagement critical component called "students engage in discussion" to determine whether the treatment is being enacted as intended. Thus, the "participant responsiveness" of Dane and Schneider and what others refer to as student participation, is captured in our critical component category of "instructional–student engagement."
5. *Differentiation*. Dane and Schneider (1998) define their dimension of differentiation as "a manipulation check that is performed to safeguard against the diffusion (unintentional spread) of treatments, that is, to ensure that the subjects in each experimental condition receive only planned intervention" (p. 45). Others refer to differentiation as determining the features of a program that distinguish one program from another (Dusenbury et al., 2003; Lynch & O'Donnell, 2005; Ruiz-Primo, 2005). We and others (e.g., Ruiz-Primo, 2005) came to understand that differentiation is not a dimension of fidelity per se but rather is an analytic process by which an evaluator determines the degree to which the critical components that distinguish one program from another are present or absent. Thus, it is a process that one undergoes before, during, or after measurement of implementation, not part of FOI measurement itself. Thus, differentiation is not reflected in our critical component categories, but rather in the fact that we have a place in our framework for identifying which critical components are common across

programs and which are common in smaller groups (e.g., common to reform-based mathematics programs) or unique to individual programs. The differentiation process is described further in the analysis section.

6. *Structure/process*. As already explained, we adopted a structure/process approach but developed it into a more specific framework that would be more useful to the field in several ways. First, we offered more specific categories within the larger structure/process approach. Given that some structural critical components were so clearly procedural and others were so clearly educative, we felt it was important to organize them into these categories so that the resulting data would yield practical information about if and how users enacted these different types of components. Likewise, we decided to keep the instructional (process) elements organized into the pedagogical and student engagement categories to facilitate more specific analyses of relationships between these categories of critical components. Without the subcategories, users of the data would not be able to know the extent to which users engaged with these very different aspects of the intervention and thus, the effects of each.

Using the Framework and Instruments

FOI Instrument Suite Overview

With the framework development and critical component identification process complete, we proceeded to develop critical component measures. The instrument development and testing process is described in project technical reports and the instruments themselves and an accompanying *User's Guide* are available from CEMSE and not discussed at length here. In brief, the suite has eight instruments for measuring science and mathematics instructional materials programs including teacher instructional questionnaire; teacher instructional log; classroom observation protocol; teacher interview protocol; school leader questionnaire; school leader interview protocol; school-wide observation protocol; and teacher attitude questionnaire. The suite includes this range of instruments to offer a variety of measurement options to users who have varying needs and resources. The *User's Guide* includes information on instrument use and adaptation, including the steps that evaluators can take to customize the instruments for use with other programs. The *User's Guide* also includes suggestions about using instruments alone or in combination with others in the suite and identifies the strengths and limitations of each instrument configuration with regard to information about an overall understanding of implementation.

Measuring FOI

Evaluators interested in understanding FOI of a single program can use the suite of instruments (or customized and retested versions) to collect data on the extent to which that program's critical components are present during enactment. The resulting data will allow evaluators to know more than FOI in a general sense. Rather, they will know which critical components were and were not enacted and be able to draw conclusions about the relationships of those components and the desired outcomes.

Differentiating Interventions From One Another

A strength of the FOI framework is that it goes beyond supporting the measurement of a single program. Because the framework provides a conceptual structure that is applicable across multiple programs, evaluators can use it as a tool to compare interventions and determine the extent to which those interventions have common and different critical components and thus actually differ from one another. As explained above, we borrow from Dane and Schneider, and call the process of

Table 4. Framework With Categories of Differentiation

	Fidelity of Implementation			
	Categories of Critical Components			
	Structural Critical Components		Instructional Critical Components	
Categories of Differentiation	Procedural	Educative	Pedagogical	Student Engagement
Common across interventions				
Unique to Intervention A				
Unique to Intervention B				

distinguishing treatments in this way "differentiation." Using this process, evaluators can determine the extent to which treatment and control conditions are in fact different when enacted and thus can make more rigorous determinations of the relationships between those conditions and desired outcomes.

By way of illustration, when comparing Intervention A and Intervention B, one first identifies the critical components present in each. Those that are common to both fall into the first category of differentiation: common across interventions (see Table 4). Those critical components that distinguish one intervention from another fall into the categories of differentiation: unique to Intervention A or unique to Intervention B. It is worth noting that the extent to which critical components would be considered common or unique depends in part on the level of specificity with which they are defined and measured. For example, in the instructional materials used in our work, all had a structural component we refer to as "lesson overview." However, those overviews look somewhat different in each of the programs. If one is interested in knowing the impact of *different approaches* to doing an overview, the data collected on a lesson overview would need to be at a sufficient level of specificity to determine the distinctions between the different approaches.

Differentiating Interventions From "Business As Usual" Conditions

Instruments built from the FOI framework can also help determine differences between the treatment and comparison groups that do not have a known intervention or what are known as "business as usual" groups. Evaluators working in schools, for example, sometimes design evaluations that compare the intervention with "typical" or business as usual instruction. However, because FOI is typically measured with instruments that are organized around the structural elements of the intervention, evaluators often have no rigorous way to measure the business as usual cases. This is problematic, because even when the intervention in name is not present, structural and instructional critical components of the intervention may be, and evaluators need a means of collecting comparable data that allow them to ascertain the extent to which the intervention and business as usual groups actually differ.

Our approach separates the structural critical components from the instructional critical components and allows them to be measured and analyzed independently of one another. Thus, evaluators can measure the extent to which instructional critical components are present in the business as usual classroom even when the structural elements are missing. Likewise, the instruments can help identify structural critical components that happen to be part of business as usual instruction even when they are not expected to be present. This differentiation allows evaluators to confidently determine the differences between the comparison groups and, in turn, the impact of the intervention.

Using the Instruments With Other Instructional Materials Programs

Although we developed instruments for the science and mathematics programs we identified at the beginning of our work, we also explored the extent to which the framework could be used as a basis for adapting the instruments to other programs and what that process might look like. In two cases, we adapted the instruments ourselves.

First, we decided to look at another instructional materials program[2] that was in use in Chicago (our field test site). As with our original programs, we began to identify the critical components of the program by reviewing the instructional materials (e.g., teacher's guides, student books). This time, however, our review was guided by our existing list of critical components. As we reviewed the materials, we determined whether each of our existing critical components was present in the program and then identified any critical components that were unique to the new program. We also identified the specific names of program components that would be necessary to customize the instruments for the new program. The ease of this process confirmed that we were able to identify our critical components in other reform-based science programs and that we could easily customize the instruments for use with these programs.

Second, we decided to experiment with customizing the instruments for instructional materials that are not part of the ''reform-based'' family of instructional materials. To do this, we selected one of the texts[3] used in Chicago and followed the same process described above. This process reconfirmed that our framework and critical component approach could be applied to other instructional materials, even those that are quite different than those in our original group.

In two more cases, we worked with the developers of other programs to adapt the instruments for their use. In the first, the developers wanted to customize the instruments to their program so that they could collect implementation data for evaluative purposes. Again, we began the process by reviewing the instructional materials and then engaged in conversations with the developers to identify the critical components that were and were not present in the program. We then worked with the developers to identify new critical components that were specific to their programs, including the particular components that the developers were interested in evaluating.

In the second case, we worked with the developers of a reform-based mathematics program.[4] The developers wanted FOI data so that they could compare a current and a newly revised version of their program. We worked with them in a process similar to that described above and not only customized the instruments to their program in general but also added items that focused on some of the very specific elements that differed between the two versions of the program.

These cases confirmed our belief that our approach to FOI and our measures are useful not only for evaluators looking at programs enacted in schools and other operational settings but also for use in evaluations of program efficacy during the development process.

Using the Framework With Other Interventions

Our work with the FOI framework then expanded to evaluations of other kinds of interventions. The framework, with its two broad critical component categories (structural and instructional) and subcategories is a useful tool for bringing a structured approach to working with program leaders during evaluations to articulate their program models. Leaders articulate their theories of action and in turn the program critical components that represent the operationalizations of those theories. The program critical components organized in the framework structure then become a basis for a specialized kind of process evaluation—what we refer to as fidelity evaluation.

For example, we have been evaluating a program to develop a high school biotechnology center. This program includes professional development for teachers, biotechnology classes, and online and onsite resource centers. For this evaluation, we used the broad FOI framework categories (structural–procedural, structural–educative, instructional–pedagogical, and instructional–student

engagement) to guide the articulation of program critical components in each of the intervention settings (i.e., professional development sessions, resource center, and classrooms).

During this process, we worked with the program developers to uncover their intentions for what was to occur in each intervention setting in each of the critical component categories. Specifically, we identified the expectations for what each setting leader needed to do (structural–procedural); what the setting leader needed to know (structural–educative); the expectations for the setting leader's behaviors and interactions with participants in the setting (instructional–pedagogical); and the expectations for the participants themselves (instructional–student engagement). For example, in the case of the professional development sessions, the critical component identification focused on the expectations for the professional development leader's enactment of the organizing elements of the professional development sessions; what the professional development leader needed to know to enact the sessions; the professional development leader's behaviors and interactions with participants (in this case, teachers) during the professional development session; and, the expectations for the participants behaviors and interactions during enactment of the professional development session. These critical components then guided the aspects of the evaluation that focused on model enactment and FOI.

Expectations for Analysis

Because our project was focused on instrument development, our analysis of the data collected during the pilot and field tests has been exploratory, having been gathered with instruments under development. Still, we have some expectations, informed by the literature on FOI for the utility of the data collected with instruments based on the FOI framework and the most useful approaches to analysis.

Historically, researchers have taken different approaches to analyzing FOI. Some, for example, represent fidelity as a single number, typically by totaling scores assigned to different parts of the program to yield a total implementation score that they leave as an implementation rating or convert to a degree of fidelity rating (Balfanz, Mac Iver, & Byrnes, 2006; Bruns, Burchard, Suter, Leverentz-Brady, & Force, 2004). Bond, Becker, Drake, and Vogler (1997), for example, created an instrument measuring fidelity to the Individual Placement and Support (IPS) employment program in mental health. It is a 15-item questionnaire with each item rated from 1 to 5. Overall fidelity to IPS, then, is the sum of all of the items with each total in a numerical range, given a label indicating degree of fidelity (e.g., "not IPS," "partial IPS," and "consistent with IPS").

Others use gradations of fidelity but tie those gradations to specific requirements. For example, in measuring levels of implementation of programs for ex-offenders, Rezmovic (1982) created a cumulative scale of implementation with each level referring to a different requirement that must be met. Others have used similar approaches with indicators of use ranging from the reporting of data to the numbers of problems attempted from a program to the use of student worksheets (Penuel & Means, 2004; Songer & Gotwals, 2005; Ysseldyke et al., 2003). These instances focus on the structural elements of the programs (worksheets, problems attempted, and length of reporting data) leaving many of the elements of the program that involve the instructional transactions unmeasured. Just as not measuring implementation can lead to unexplained program findings, only looking at a small part of implementation can leave questions as to whether data on a single aspect of implementation are sufficient to analyze the relationships between the intervention and outcomes in a meaningful way.

Although data collected using the FOI instruments could be used to do analyses like those described above, they are best suited to align with a third approach to analysis that seeks to identify gradations of implementation that take into account the whole range of possible critical components in particular combinations and considers their impact on student outcomes. This approach,

sometimes referred to as the use of ''innovation configurations'' or ''innovation profiles,'' allows for more specific analysis and clearer understandings of the roles that particular critical components play.

Implementation Types

In their early work on ''innovation configurations,'' Hall and Hord (1987) focused on measuring each critical component at particular levels of variation. For example, they measured a critical component they called *Use of Lesson Packets* by identifying variations of the component: ''Teachers and students use lessons with visual aids whenever appropriate,'' ''Uses lessons with visual aids infrequently,'' ''Uses lessons without visual aids,'' and ''Does not use lesson packets''. The Innovation Configuration then notes the ideal, acceptable, and unacceptable variations. Others have used innovation configuration approaches to conduct more specific implementation analyses that are based on determining degrees of use of each program element (Becker, Smith, Tanzman, Drake, & Tremblay, 2001; Huntley, 2005; Neale, Smith, & Johnson, 1990; Ross, McDougall, Hogaboam-Gray, & Sage, 2003).

Evaluators collecting data using the instruments derived from our framework can best capitalize on it with an innovation configuration approach that we refer to as ''types.'' An implementation type represents a particular combination of critical components enacted to particular degrees. A type is comprised of four composite scores, one from each of the critical component categories (i.e., structural–procedural, structural–educative, instructional–pedagogical, and instructional–student engagement). The composite score is calculated to account for the presence of the critical components in the category and their relative importance (relative weight). Then, the four composite scores together determine the type.

In our work on instructional materials, for example, one type of implementation (Type A) may include a high composite score of structural–procedural critical components but a low composite score of structural–educative components and even lower composite score of instructional–pedagogical critical components. Another type (Type B) may include a high score of instructional–pedagogical critical components and very low score of structural–procedural components. Knowing this, we can explore the extent to which teachers that have implementation Type A demonstrate stronger or weaker relationships to desired student outcomes than teachers that have implementation Type B. Thus, rather than focus only on *whether* implementation leads to desired outcomes, the analysis will help identify which combinations of critical components (types) lead to desired outcomes.

During the pilot and field test analysis, we have explored the use of more specific indices and types to capture FOI. In addition to exploring the use of scores for the four categories of FOI, we are looking at the potential of determining scores for smaller groups of critical components (e.g., instructional–pedagogical critical components related to facilitating student engagement with the content). During this exploratory work, we are also considering the role that differential weighting of critical components might play. To do this, we will review data collected with our FOI instruments and revisit our conversations with developers to determine which critical components are ''more critical'' than others and therefore should receive more weight. This will also inform questions about ''acceptable adaptation'' discussed below.

Relationships Between Critical Component Categories

Because the framework is divided into structural and instructional categories, evaluators can explore relationships between the presence of particular intervention structures and the intended practices, behaviors, and interactions expected to be associated with those structures. For example, if the data

show the presence of the critical component, ''teacher facilitation of student discussion'' (an instructional critical component), the evaluator can look at correlations to see if there is a relationship between that critical component and the presence of the structural critical components that the developers associate with it (e.g., background information on pedagogy). In a formative evaluation, this will help determine which structures of an intervention are and are not related to the intended instructional behaviors and thus which should remain and which should be revised or removed.

This type of correlational analysis can also inform understandings about relationships between intervention participants' behaviors and interactions. For example, evaluators might find that the critical component ''teacher facilitation of student discussion'' (in the instructional–pedagogical) category is highly correlated with ''students engage in discussion'' (instructional–student engagement), which would be an expected finding. Or, evaluators might find that they are not highly correlated, which would be quite unexpected and raise some key questions. The ability to do analyses at this level of specificity brings more useful information to intervention developers, decision makers, users, and of course, the evaluators about the status of implementation and enables them to further explore the relationship between implementation and the program outcomes.

Acceptable Adaptation

In addition to looking at relationships between specific critical components, the framework and an innovation configuration approach to analysis can help inform one of the long-standing questions in FOI literature: How much adaptation is acceptable when implementing a program? (Blakely et al., 1987; Hall & Loucks, 1977). Those who take a fidelity perspective contend that implementation should occur as intended by developers, whereas the adaptation perspective allows for changes to occur to fit specific contexts. Regardless of which perspective is taken, all of the research suggest that when users enact programs, the programs change as a result of operating in different settings with different contexts (Bodzin, Cates, & Price, 2003; Buston, Wight, Hart, & Scott, 2002). Given that adaptation happens, a unidimensional view of FOI that results in a single score or rating does not accommodate the dynamism of intervention enactment in the real world. Knowing which implementation types or combinations of critical components are most strongly related to desired student outcomes can inform evaluators who seek to understand the extent to which programs can or should be adapted and where that adaptation can take place without compromising effectiveness (Fullan & Pomfret, 1977; Emshoff et al., 1987; Lipsey & Cordray, 2000 in Penuel & Means, 2004; Songer & Gotwals, 2005).

The question of where and in what ways adaptation can or should take place can be answered with analyses guided by our framework and a critical component approach. It is noteworthy that while some define ''critical components'' much as we do (*the elements of the intended program model as defined by developers and users*), others define critical components as those elements of interventions that have been shown through empirical studies to have a significant relationship to desired outcomes. Thus, one point of view has to do with intent while the other has to do with proven effectiveness.

Neither point of view is absolute nor are they mutually exclusive. Rather, the definitions complement one another in an iterative process. Early in development, for example, an intervention may be conceived based on theory and exploratory work. Then, in pilot tests and efficacy trials, informed by appropriate formative evaluations, the developers get more information about the impact of the intervention and, if measured carefully using a tool such as the FOI framework, about the particular program elements that appear to be most related to student outcomes. These program elements, now shown to have an impact, are then defined as ''critical'' and retained in the revision process. This is, of course, an evolving process as interventions are adjusted based on new information.

As particular critical components of an intervention acquire increasing amounts of empirical validation, it becomes clear that acceptable adaptation of the intervention should not alter those components. At the same time, those components that do not appear to be as closely tied to desired outcomes are more reasonable candidates for adaptation. Analyzing interventions at this level of specificity and clarity allows for these kinds of specific analysis and lays the groundwork for accumulating knowledge about intervention components that emerge across multiple interventions as being closely correlated to desired outcomes.

FOI as a Dependent Variable

In addition to looking at FOI itself, some evaluators will want to focus on the factors that affect FOI such as a lack of time, lack of program specifications, lack of support, and attitudes/beliefs about the intervention being implemented (Bay, 1999; Bodilly & Keltner, 1998; Bodzin et al., 2003; Buston et al., 2002; Fagan, 1996; Flocks et al., 2001; Heaney, 1995; McDonel et al., 1997; Wafa & Yasin, 1998). These are among the variables we discussed earlier as moderating variables. As dependent variables, data from instruments grounded in the framework can contribute to understandings about the relationships between particular contextual factors and implementation. These analyses, then, help answer the question of *why* and *under what contexts* certain programs or critical components are effective.

Broadening Our Work in the Future

Although the instruments in our suite will contribute significantly to supporting rigorous studies of mathematics and science instructional materials, as discussed above, the framework and instrument adaptation process can support FOI measurement of instructional materials in other disciplines as well as other types of educational interventions and interventions in other fields. Our steps, including the explicit articulation of the program model through identification of the model's critical components, organization of those critical components into the FOI conceptual framework, and instrument development and adaptation based on the critical components create a sound starting point for rigorous FOI data collection for other interventions. Through the articulation and categorization of program critical components, the framework will open up the black box of FOI and provide a deeper understanding of the program components and characteristics that are most closely tied to the desired student outcomes.

In the short term, we hope that this process and its associated instruments will become increasingly refined through our and others' use of them in studies of other interventions. We also hope that evaluators in other fields will begin to use the framework as a foundation for FOI measurement in their respective areas and ultimately be able to share findings with us and one another about using the framework for FOI measurement as well as findings about effective intervention critical components. In the long term, we hope to work with others to further our collective understanding of FOI, our ability to accumulate knowledge about its measurement, and our ability to accumulate knowledge about how to develop interventions with the elements that make them most effective.

Notes

1. Full Option Science System (FOSS), Science and Technology for Children (STC), Science Education for Public Understanding Program (SEPUP), Science Companion (SC), and Everyday Mathematics (EM).
2. Investigating Earth Systems (IES).
3. Discovery works.
4. Mathematics trailblazers.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: The National Science Foundation, Award ESI-0628052.

References

Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). The implementation and impact of evidence-based mathematics reforms in high-poverty middle schools: A multi-site, multi-year study. *Journal for Research in Mathematics Education*, *37*, 33-64.

Bauman, L. J., Stein, R. E. K., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, *19*, 619-639.

Bay, J. M. (1999). *Middle school mathematics curriculum implementation: The dynamics of change as teachers introduce and use standards-based curricula* (Unpublished doctoral dissertation). University of Missouri, Columbia.

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.

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, *15*, 253-268.

Bodilly, S. J., & Keltner, B. (1998). *Lessons from new American schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: RAND (MR-1777-NAS).

Bodzin, A. M., Cates, W. M., & Price, B. (2003). *Formative evaluation of the exploring life curriculum: Year two implementation fidelity findings*. Paper presented at the meeting of the National Association for Research in Science Teaching (NARST), Philadelphia, PA.

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.

Bond, G., Williams, J., Evans, L., Salyers, M., Kim, H., Sharpe, H., & Leff, H. S. (2000). *Psychiatric rehabilitation fidelity toolkit.* Cambridge, MA: Human Services Research Institute.

Bruns, E. J., Burchard, J. D., Suter, J. C., Leverentz-Brady, K., & Force, M. M. (2004). Assessing fidelity to a community-based treatment for youth: The Wraparound Fidelity Index. *Journal of Emotional and Behavioral Disorders*, *12*, 79-89.

Buston, K., Wight, D., Hart, G., & Scott, S. (2002). Implementation of a teacher-delivered sex education programme: Obstacles and facilitating factors. *Health Education Research*, *17*, 59-72.

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.

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

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.

Fagan, P. J. (1996). 1995 implementation status of mathematics and science reform in Iowa: Based on teachers' concerns, professional activity, and philosophical beliefs. Iowa Mathematics and Science Coalition, Cedar Falls. Retrieved from ERIC database. (ED399191).

Flocks, J., Clarke, L., Albrecht, S., Bryant, C., Monaghan, P., & Baker, H. (2001). Implementing a community-based social marketing project to improve agricultural worker health. *Environmental Health Perspectives*, *109*, 461-468.

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

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, *47*, 335-397.

Gersten, R. M., & Carnine, D. (1980). *Measuring implementation of the direct instruction model in an urban school district: An observational approach.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Gresham, F. M., Gansle, K., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, *26*, 257-263.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process*. State University of New York Press.

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.

Heaney, C. A., & Burke, A. C. (1995). Ideologies of care in community residential services: What do caregivers believe? *Community Mental Health Journal*, *31*, 449-462.

Huntley, M. A. (2005). *Operationalizing the concept of "fidelity of implementation" for NSF-funded mathematics curricula*. Unpublished manuscript.

Lastica, J., & O'Donnell, C. (2007). *Considering the role of fidelity of implementation in science education research: Fidelity as teacher and student adherence to structure*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Leithwood, K. A., & Montgomery, D. J. (1980). Evaluating program implementation. *Evaluation Review*, *4*, 193-214.

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

Lynch, S., & O'Donnell, C. (2005). *The evolving definition, measurement, and conceptualization of fidelity of implementation in scale-up of highly rated science curriculum units in diverse middle schools*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

McDonel, E. C., Bond, G. R., Salyers, M., Fekete, D., Chen, A., McGrew, J. H., & Miller, L. (1997). Implementing assertive community treatment programs in rural settings. *Administration and Policy in Mental Health*, *25*, 153-173.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *The American Journal of Evaluation*, *24*, 315-340.

Neale, D. C., Smith, D., & Johnson, V. G. (1990). Implementing conceptual change teaching in primary science. *The Elementary School Journal*, *91*, 109-131.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33-84.

Penuel, W. 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.

Rezmovic, E. L. (1982). Program implementation and evaluation results. *Evaluation and Program Planning*, *5*, 111-118.

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education*, *34*, 344-363.

Ruiz-Primo, M. A. (2005). *A multi-method and multi-source approach for studying fidelity of implementation*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Sabelli, N., & Dede, C. (2001). *Integrating educational research and practice: reconceptualizing goals and policies: "How to make what works work for us?"* Retrieved June 26, 2007 from http://www.virtual.gmu.edu/ss_research/cdpapers/policy.pdf

Songer, N. B., & Gotwals, A. W. (2005). *Fidelity of implementation in three sequential curricular units*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Wafa, M. A., & Yasin, M. M. (1998). A conceptual framework for effective implementation of JIT: An empirical investigation. *International Journal of Operations & Production Management*, *18*, 1111-1124.

Wang, M. C., Nojan, M., Strom, C. D., & Walberg, H. J. (1984). The utility of degree of implementation measures in program implementation and evaluation research. *Curriculum Inquiry*, *14*, 249-286.

Ysseldyke, J. E., Spicuzza, R., Kosciolek, 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.