Symposium Proposal SREE 2010

Title: Progressing Toward a Shared Set of Methods and Standards for Developing and Using Measures of Implementation Fidelity

Contact: Chris S. Hulleman, James Madison University, hullemcs@jmu.edu

Symposium Justification:

One consistent purpose of education research—and one that has been increasingly stressed in recent years with the enactment of the Education Science Reform Act of 2002 and the establishment of the Institute of Education Sciences (IES)—is to develop and rigorously evaluate programs that are effective in supporting students' learning and achievement. Included in this recent agenda is an emphasis on measuring implementation fidelity and linking those measures to program impacts. Without defining and assessing the degree to which programs are implemented as intended, claims of treatment effectiveness may be unjustified and invalid. However, despite this emphasis on measuring implementation fidelity, recent reviews of studies in school settings have illustrated that many inconsistencies and omissions in measuring fidelity exist (Dusenbury, 2003; O'Donnell, 2008).

With this symposium we seek to draw upon current work in creating and using measures of implementation fidelity to further conceptualize implementation fidelity and progress the field toward a shared set of methods and standards for both ensuring that programs of interest are implemented as intended and measuring the extent to which the degree of implementation fidelity contributes to program outcomes. Specifically, the original empirical work described in the accompanying abstracts employs a five-step method proposed by Cordray (2007), providing cases and elaborations of that method.

The first paper provides an analysis of fidelity measures used in recent evaluations of preschool curriculum interventions to illustrate problems in how such studies measure fidelity implementation specifically, and to provide an additional argument for the urgent need for a standard in the broader field of educational program evaluation. The second paper reviews work that has employed one or more steps of Cordray's (2007) method, illustrating how the method can progress evaluation research beyond the shortcomings of intent-to-treat analyses. The third paper documents the use of this method in the case of a highly unscripted intervention, examining specifically the challenges of creating reliable and valid instruments for measuring implementation fidelity in evaluations of programs built on cognitive models of progressions of student thinking and the relationship between fidelity studies and program development. The fourth and final paper further elaborates the method by accounting for the extent to which core intervention components are present in both the treatment and control condition, providing a rationale and method for measuring and using indices of the "achieved relative strength" of an intervention.

The symposium will be organized as follows. The chair/discussant of the session will give a five-minute introduction in which he will outline the objectives of the symposium. The authors of the four papers will each have 20 minutes to present their papers. The chair/discussant will have 20 minutes to offer a synthesis and critique of the four papers and to pose questions to the authors. He will then moderate a 15-minute discussion with the audience.

Abstract Title Page
Not included in page count.

Title: Measuring Fidelity in Preschool Interventions: A Microanalysis of Fidelity Instruments Used in Curriculum Interventions

Author(s): Catherine L. Darrow, Vanderbilt University

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Without defining and assessing the degree to which an intervention was implemented as intended (i.e., implementation fidelity), claims of treatment effectiveness may be unjustified. Despite the value of measuring implementation fidelity, recent reviews of studies in K-12 school settings have illustrated that many inconsistencies and omissions in measuring fidelity exist (Dusenbury, Brannigan, Falco, & Hansen, 2003; O'Donnell, 2008).

No review of preschool studies and the ways in which they assess fidelity of implementation has yet been published. Therefore, it is difficult to determine if the problems inherent in K-12 studies are also evident in preschool interventions. The movement to establish effective preschool programs has gained momentum. The fact that some children are not prepared to enter kindergarten is a source of great concern for educational researchers, practitioners, and policy makers. This concern prompted the U.S. Department of Education to invest millions of dollars in grants for evaluations of preschool curricula (Barnett, 2008). With increased funding for such studies and heightened attention to the importance of implementation fidelity comes the emergence of a variety of fidelity instruments. However, the increase in number of fidelity instruments does not guarantee that these measures effectively assess implementation.

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

This paper examines measures used by studies associated with Preschool Curriculum Evaluation Research Initiative (PCER) funded by the Institute of Education Sciences. Analysis of 17 measures of fidelity used by 13 curriculum interventions highlights how research in preschools is and is not measuring fidelity of implementation. The following research questions are pursued:

1. In what ways do the measures used in these studies represent the three primary elements of implementation: (a) adherence, (b) exposure, and (c) participant responsiveness?
2. Do studies effectively use measures to differentiate between teachers' fidelity to crucial elements of the curriculum and teachers' general instructional quality (i.e., high fidelity as compared to good teaching)?

Population / Participants / Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

Fidelity measures were gathered from 13 research studies funded by the Institute of Education Sciences' Preschool Curriculum Evaluation Research Initiative (PCER) and located in Tennessee, North Carolina/Georgia, New Hampshire, Florida/Kansas/New Jersey, Texas, Virginia, California/New York, Wisconsin, Missouri, New Jersey, and Florida. This paper

evaluates instruments representing 12 separate curricula implemented within these 13 interventions.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Seventeen fidelity instruments were coded on the instrument and item level (see Table 1 in Appendix B for a list of fidelity measures used by each study). At the instrument level, each measure was coded for the inclusion of teacher interview, guidance on optimal length of observation, and opportunities for observers to document the length of observed activities. At the item level, the coding schema identified individual items that included direct reference to the targeted curriculum and items that provided some means to assess the quality of a particular feature of the environment or instruction. Additionally, each item was then designated as relating to either structural elements of the classroom setting or instructional features associated with teacher or child behaviors.

Findings / Results:
Description of main findings with specific details.

Results show that few studies use measures that comprehensively represent the three primary elements of implementation. Only 5.7% of all items across instruments (N = 1,113 items) and only 5 of the 16 measures contained any assessment of adherence. Likewise, less than half of the studies provided measures representing children's exposure to the curriculum. Directions included in the measures often lacked guidance on the ideal length of the observation and frequently omitted space for observers to indicate the total length of observation. Only 9 of the 15 measures analyzed at the item level included scheduling items. Thus, few measures evaluated how well teachers followed guidance on the intended length and sequence of curricular activities. Additionally, only 3 of the 16 fidelity measures included representation of teacher responsiveness in the form of a teacher interview. The included fidelity measures virtually ignored children's response to structural and instructional elements of the interventions.

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

Overall, analysis of these 16 fidelity measures reveals that the corresponding preschool studies insufficiently measured the primary components of fidelity. In many cases, the measures used in these studies were more effective at assessing general quality of instruction and less successful at evaluating fidelity of implementation. Despite increased attention, measurement of implementation fidelity in preschool research continues to be challenging and development of measures that more comprehensively represent implementation fidelity more is essential.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Barnett, W. S. (2008). Federal Pre-K curriculum study shows few stand-outs. *Preschool Matters, 6,* 4-5.

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*, *28*, 237-256.

Hulleman, C. S., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Intervention Effectiveness*, *2*(1), 88-110.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

Appendix B. Tables and Figures
Not included in page count.

Table 1
Fidelity Measures Used by Study

Study Location	Curriculum	Fidelity Measure
NH	Creative Curriculum with Ladders to Literacy	1. Ladder to Literacy Implementation Checklist (Classroom Activities) 2. Ladder to Literacy Implementation Checklist (Scaffolding)
NY & CA	Pre-K Mathematics + DLM Express	1. Early Mathematics Classroom Observation 2. Fidelity of Implementation Record Sheet 3. Classroom Obs of Early Math Environment and Teaching (COEMET)
NC, GA & NH	Creative Curriculum	1. Creative Curriculum Checklist
TX	Doors to Discovery	1. Doors to Discovery: Curriculum Fidelity Checklist 2. PCER Teacher Checklist*
TX	Let's Begin with Letter People	1. Let's Begin: Curriculum Fidelity Checklist 2. PCER Teacher Checklist*
WI	Project Approach	1. The Project Approach Fidelity Scale
MO	Project Construct	1. Project Construct: Early Childhood Observation Su (PC-ECCOS)
VA	Language Focused Curriculum (LFC)	1. LFC Fidelity Checklist
FL	DLM Early Childhood Express with Open Court Pre-K	1. DLM/Open Court Classroom Observation
FL	Literacy Express	1. Literacy Express Classroom Observation
NJ	Ready, Set, Leap!	1. Fidelity Observation

		2. Classroom Lang Arts Systematic Sampling and Instructional Coding (CLASSIC)
TN	Bright Beginnings	1. Combined CC and BB checklist
TN	Creative Curriculum	1. Combined CC and BB checklist

* PCER Checklist is the same for both TX studies (Doors to Discovery and Let's Begin with Letter People

Abstract Title Page
Not included in page count.

Title: A Procedure for Assessing Fidelity of Implementation in Experiments Testing Educational Interventions

Author(s): Michael C. Nelson, David S. Cordray, Vanderbilt University; Chris S. Hulleman, James Madison University; Catherine L. Darrow, & Evan C. Sommer, Vanderbilt University

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

An educational intervention's effectiveness is judged by whether it produces positive outcomes for students, with the randomized controlled trial (RCT) as a valuable tool for determining intervention effects. However, the intervention-as-implemented in an experiment frequently differs from the intervention-as-designed, making it unclear whether unfavorable results are due to an ineffective intervention model or the failure to implement the model fully. It is therefore vital to assess fidelity of implementation and, where possible, to incorporate fidelity data in the analysis of outcomes.

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

Our objectives are to 1) elaborate a five-step procedure for fidelity assessment, 2) describe the advantages of assessing fidelity with this approach when conducting experiments testing educational interventions, and 3) use examples based on published studies to illustrate how this procedure can be applied.

Setting:

Description of where the research took place.

N/A

Population / Participants / Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

We examine five examples of experiments that tested educational interventions in a range of subject areas (e.g., mathematics, reading, science, behavior) and grade levels (e.g., preschool, elementary, high school).

Intervention / Program / Practice:

Description of the intervention, program or practice, including details of administration and duration.

Fidelity is the extent to which an intervention is implemented as intended. Fidelity assessment describes the cause-and-effect sequences inside the experimental "black box" that collectively make up the implementation of the intervention. By assessing fidelity, we determine the extent to which this sequence occurred as originally envisioned, thus assessing construct validity and providing a basis for generalization. While not always possible, the ideal fidelity assessment includes five steps (Cordray, 2007): 1) describe the intervention change and logic models; 2) create indices to measure the fidelity of the implemented intervention to constructs identified in the models; 3) determine the reliability and validity of indices; 4) combine indices in the analysis; and 5) link fidelity measures to outcomes.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

As a foundation for our paper, we discuss the limitations of the intent-to-treat experimental model for explaining effects, and the advantages of using fidelity measures to look inside the black box. We describe the five-step assessment process and the rationale for each step. We then demonstrate the procedure's application using several published studies that followed one or more of these steps, illustrating how this approach can be applied practically and the benefits it yields.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Studies examined are drawn from a pool of 526 articles across subjects and grade-levels, identified through a review based on O'Donnell's (2008) methodology. The search employed 14 databases, including ERIC, PsycINFO, JSTOR, Dissertation Abstracts, and Google Scholar. We narrow the scope of articles to peer-reviewed studies involving experiments.

Findings / Results:
Description of main findings with specific details.

We find that, while most analyses of fidelity in educational studies include one or more steps in the five-step fidelity assessment process, the typical study does not include all steps and/or executes steps incompletely. As a result, the usefulness of fidelity data for interpreting experimental results is diminished. We demonstrate that a more complete assessment of fidelity in such cases would lead to a more meaningful analysis that interprets experimental results in light of actual implementation.

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

We conclude that the fidelity assessment procedure can be followed for a range of intervention types and subjects. Furthermore, this approach leads to more *complete* fidelity assessment and thus to better understanding of construct validity and generalizability. By advancing these arguments and providing examples, we encourage broader understanding of the importance of fidelity assessment, more complete assessment, and improved methods for assessment.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Cordray, D.S. (2007) Assessing Intervention Fidelity in Randomized Field Experiments. Funded Goal 5 proposal to Institute of Education Sciences.

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K-12 Curriculum Intervention Research. *Review of Educational Research*, *78(1),* 33-84.

Appendix B. Tables and Figures

Not included in page count.

Abstract Title Page
Not included in page count.

Title: Evaluating Math Recovery: A Case of Measuring Implementation Fidelity of an Unscripted, Cognitively-Based Intervention

Author(s): Charles Munter and Anne Garrison, Vanderbilt University

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Included in the recent push for rigorous evaluations of programs that are effective in supporting students' learning and achievement is an emphasis on measuring implementation fidelity and linking those measures to program impacts. Claims of treatment effectiveness may be unjustified and invalid unless the degree to which programs are implemented as intended is defined and assessed. However, despite this emphasis on measuring implementation fidelity, recent reviews of studies in school settings have illustrated that many inconsistencies and omissions in measuring fidelity exist (Dusenbury, 2003; O'Donnell, 2008; Ruiz-Primo, 2005). Furthermore, little is known regarding the feasibility of conducting studies of implementation fidelity of unscripted interventions, where measuring fidelity first requires the identification and operationalization of complex, subtle facets of the intervention (Cordray & Pion, 2006). The field is in need of progress in this area to capitalize on the potential of fidelity studies to identify the most problematic areas of implementing a program and to provide feedback to developers for refining a program (Ruiz-Primo, 2005).

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

In this paper, we describe a case of measuring implementation fidelity within an evaluation of Math Recovery (MR), a pullout tutoring program for low-achieving first-graders. We use this case to address two aspects of implementation fidelity studies: 1) their feasibility with respect to unscripted interventions, and 2) their relationship to ongoing program development. The aim of the MR program is to use children's current understandings of number as bases for providing instruction that will support them in constructing increasingly sophisticated strategies. Therefore, assessing fidelity in this case is not as simple as monitoring adherence to a script, but requires assessing the complex practice of delivering mathematics instruction attuned to a child's current understanding and needs.

Our intentions were to both measure the extent to which the program was implemented as intended, and link the measures to student outcomes. Determining the extent to which the tutoring is enacted as intended requires an explication of 'good' tutoring as defined by the developers and systematically evaluating tutors' practices against that ideal. However, we also go beyond MR's notion of 'good' tutoring by looking for instances of "positive infidelity" (Cordray & Hulleman, 2009) within tutoring sessions. Thus, we view studies of implementation fidelity as potential sources for refining theory and program design.

Setting:

Description of where the research took place.

The two-year evaluation of Math Recovery was conducted in 20 elementary schools (five urban, ten suburban and five rural), representing five districts in two states. Each was a 'fresh site' in that the program was implemented for the first time for the purposes of the study.

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

Students were selected for participation at the start of first grade based on their performance on MR's screening interview and follow-up assessment interview. Eighteen teachers were recruited to receive training and participate as MR tutors from the participating districts—all of whom had at least two years of classroom teaching experience. Sixteen of the tutors received half-time teaching releases to serve one school each; two of the tutors served two schools each. All tutoring positions were underwritten by their respective school districts.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

To begin, the MR tutor conducts an extensive video-recorded assessment interview with each child identified as eligible for the program. The tutor analyzes these video-recordings to develop a detailed profile of each child's knowledge of the central aspects of arithmetic using the MR Learning Framework, which provides information about student responses in terms of levels of sophistication. The one-to-one tutoring that follows is diagnostic in nature and focuses instruction at the current limits of each child's arithmetical reasoning. Each selected child receives 4-5 one-to-one tutoring sessions of 30 minutes each week for approximately 11 weeks. Every lesson is video-recorded for purposes of daily reflection and planning. The tutor's selection of tasks for sessions with a particular child is initially informed by the assessment interview and then by ongoing assessments based on the student's responses to prior instructional tasks. The Learning Framework that the tutor uses to analyze student performance is linked to the MR Instructional Framework that describes a range of instructional tasks organized by the level of sophistication of the students' reasoning together with detailed guidance for the tutor.

Guiding the fidelity assessment were what we, in collaboration with program developers, determined to be the unique aspects of Math Recovery tutoring as compared to typical tutoring: (a) the tutor's ongoing assessment of the child's thinking and strategies (both reflective assessment between tutoring sessions and in-the-moment assessment); and (b) the tutor's efforts to provide instruction within the child's zone of proximal development.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

The larger evaluation study was a randomized field trial. In each year (2007-08 and 2008-09 academic years), 17 to 36 students deemed eligible (based on an initial MR screening) from each of 20 schools were randomly assigned to one of three tutoring cohorts or to the "wait list" for MR. The cohorts, consisting of three students each, were staggered across different start dates (i.e., Cohort A—September, B—December, C—March). In both years students on the randomly ordered waiting list were selected to join an MR tutoring cohort if an assigned participant left the school or were deemed "ineligible" due to a special education placement. The number of study participants totaled 517 in Year 1 and 510 in Year 2, of which 172 received tutoring in Year 1 and 171 received tutoring in Year 2. Consistent with typical MR practice, all sessions were video-recorded.

In this paper, we focus our attention on the implementation fidelity component of the project, for which we followed the five-step method outlined by Cordray (2007): 1) describe the intervention change and logic models; 2) create indices to measure the fidelity of the implemented intervention to constructs identified in the models; 3) determine the reliability and validity of indices; 4) combine indices in the analysis; and 5) link fidelity measures to outcomes. Here, we report primarily on step three, the process of determining the reliability and validity of fidelity indices.

At the outset, we included program developers in conversations to identify key implementation components (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005) and initial schemes for measuring those constructs. The research team finalized the instruments through an iterative refinement process, grounded in MR's guiding principles. A team of coders, trained in both MR tutoring (by MR expert trainers) and video coding (by the research team), were assigned a random selection of twelve tutoring sessions from one student per tutor per cycle (a total of 108 students). For purposes of external validation, a subset of tutoring sessions spanning a range of indices of implementation fidelity as determined by our coding scheme were sent to 20 MR experts, who rated the tutoring practices based on their own notions of high-quality MR practice.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

Guided by the unique aspects of Math Recovery tutoring listed above (i.e., the tutor's ongoing assessment of the child's thinking and efforts to provide instruction within the child's zone of proximal development), our goal in assessing implementation fidelity was to answer a set of key questions regarding tutors' assessment and instruction: (a) Was the initial assessment done? If so, was it done correctly? (b) In instructional lessons, did the tutor choose procedures (i.e., sets of related tasks) that were in the child's zone of proximal development (according to the MR Frameworks)? (c) Did the tutor utilize/implement the procedures/tasks well?

Regarding the first question, we identified two possibilities for breakdown: the tutor might have 1) presented the incorrect assessment tasks (or tasks that were misaligned with those printed in the assessment), or 2) used poor judgment in interpreting the results (i.e., assigned a profile to the student that conflicted with our external assessment of the child's current understanding). For each of these we defined what constituted a *minor error*, a *major error*, or *no error*.

To answer the second question, regarding tutors' *choice* of procedures, coders first viewed up to three previous tutoring sessions to locate the child's thinking at that point on the MR Learning Framework, and then determined whether the tutor's choice of procedures matched the child's placement on the MR Learning Framework. That is, did the tutor's choice of procedures align with what the MR Instructional Framework suggested? Often tutors utilized procedures as described in the MR handbook, but when they incorporated procedures from other sources, coders located those procedures on the Instructional Framework based on the procedure's focus (e.g., arithmetical strategies, number word sequences, etc.), and the level of difficulty of the tasks within the procedure, including number range and the extent to which the tasks were scaffolded.

Lastly, to answer the question pertaining to tutors' *implementation* of tasks (within procedures), coders examined the extent to which tutors followed established "rules" within the MR program (e.g., things a tutor is supposed to do, or prohibitions). For example, tutors are

expected to consistently solicit students' strategies for solving problems (if the strategy is not already visible), and are expected to avoid merely eliciting particular behaviors.

After four weeks of refinement work (described above), agreement percentages plateaued at an inadequate level—largely due to differences in how coders 'chunked' the lessons they were coding (e.g., Was it one big task, or two?) Therefore, the evaluation team identified a representative aspect of the MR Instructional Framework about which coders' structural decisions had consistently agreed and for which all codes would remain relevant. Of the six aspects included in the MR Learning Framework, two of them (Stages of Early Arithmetical Learning, and Tens and Ones) represent the heart of the theory underlying the MR program. Although lessons typically include practice on other aspects such as number word sequences or numeral identification, it is these two aspects that pertain directly to MR's unique aspects listed above. Therefore, video coding focused on instances of activities aimed at supporting students in developing more sophisticated *strategies*, rendering the fidelity assessment process more tractable without sacrificing any attention to core implementation components.

Findings / Results:

Description of main findings with specific details.

Throughout the coding process (after the initial refinement phase), coders maintained an average percent agreement of 0.80. Furthermore, MR experts' ratings validated our coding schemes, with sufficiently high correlations between their ratings and those based on fidelity indices ($\alpha = .75$, $p < .05$).

Conclusions:

Description of conclusions and recommendations based on findings and overall study.

Our findings suggest it is possible to create a reliable instrument to measure implementation fidelity for differentiated interventions—an endeavor that has, heretofore, been largely avoided in evaluations of educational interventions. Many potentially high-quality interventions are un-scripted, instead relying on teacher knowledge and professional development, requiring considerable differentiation by implementers. As we work to rigorously evaluate such programs, we need to develop reliable fidelity measures that are both feasible and true to program components, so that evaluators can adequately link measures of treatment integrity to outcomes, to more accurately determine the relative strength of interventions (Cordray & Pion, 2006).

Secondly, in coding for instances of "positive infidelity" we have identified "local additions" (Blakely et al., 1987) that could possibly strengthen the design of the program. Members of the research team have already provided feedback (e.g., at Math Recovery practitioner conferences) to challenge developers' current conceptions and support them in improving the program.

This paper outlines the development and use of a fidelity measure as a case of how such instruments might be developed and used in the future. Critical aspects of the process included 1) the identification of the core implementation components of the intervention; 2) close work with program developers to operationalize those components; 3) training of coders in both the program itself and the coding schemes/process; and 4) collaborating with the coding team to further refine operationalizations and coding decisions, to strike a balance of feasibility and adherence to program components.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Blakely, C., Mayer, J. P., Gottschalk, R. G., Schmitt, N., & Davidson, W. S. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social program. *American Journal of Community Psychology, 15*, 253-269.

Cordray, D.S. (2007) Assessing Intervention Fidelity in Randomized Field Experiments. Funded Goal 5 proposal to Institute of Education Sciences.

Cordray, D. S. & Hulleman, C. (2009, June). Assessing intervention fidelity: Models, methods and modes of analysis. Presentation at the Institute for Education Sciences 2009 Research Conference, Washington, D.C.

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

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*, *28*, 237-256.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K-12 Curriculum Intervention Research. *Review of Educational Research*, *78(1),* 33-84.

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.

Appendix B. Tables and Figures

Not included in page count.

Abstract Title Page
Not included in page count.

Title: Achieved Relative Intervention Strength: Models and Methods

Author(s): Chris S. Hulleman, James Madison University; David S. Cordray, Vanderbilt University

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Treatment fidelity (i.e., integrity) can be defined as the extent to which an intervention is delivered as intended. Inevitable slippage between the specified and implemented intervention means that any treatment effects must be attributed not to the theoretical intervention model but to the model as implemented. Ideally, fidelity assessments would follow a specific course, beginning with a full characterization of the intervention "in theory" by outlining: the constructs and processes being manipulated (change model), the essential components of the intervention and required support services (logic model), and the context of implementation within the school year (operational model). Explication of these models leads to the development of reliable and valid measures of *achieved* intervention fidelity (Cordray & Pion, 2006), and the integration of fidelity measures into outcomes analysis. Without measuring achieved fidelity in both treatment and control conditions we are unable to determine whether differences in outcomes (or lack thereof) are due to achieved fidelity in the treatment conditions or the presence of core intervention components in the control condition. An intervention may have a high degree of achieved fidelity, but fail to have effects because it does not differ from the control conditions on core intervention components (i.e., small achieved relative strength).

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

This paper outlines a method of determining the achieved relative strength of an intervention by calculating the extent to which core intervention components are present in both the treatment and control condition (Hulleman & Cordray, 2009). Using data from several examples we demonstrate how to calculate ARS indices, and subsequently how to incorporate them into analyses of intervention effectiveness.

Setting:

Description of where the research took place.

The results from three different samples are reported in this paper. Sample 1 was collected in an experimental laboratory on a college campus. Sample 2 was collected within an undergraduate class. Sample 3 was obtained from secondary data analysis of Reading First data.

Population / Participants / Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

Sample 1: This sample consists of introductory psychology students at a major mid-Western university who participated in the experiment for course credit. Participants volunteered to visit our research laboratory and participate in a randomized experiment.

Sample 2: This sample consists of introductory psychology students at a major mid-Western university who participated in the experiment for course credit. Students completed the experimental intervention as part of an introductory psychology course assignment.

Sample 3: This sample consists of secondary data analysis from the Reading First Implementation report (U.S. Department of Education, 2008).

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

The randomized intervention in the first two samples (the psychology laboratory and the college classroom) was a curricular intervention designed to encourage participants to make connections between the material they were studying and their own lives (for a more complete description of the intervention see Hulleman & Cordray, 2009; Hulleman, Godes, Hendricks, & Harackiewicz, in press; and Hulleman & Harckiewicz, in press). In both studies students were randomly assigned to the treatment group or the control group. Treatment participants selected a topic they were studying (either mental math in the laboratory or a topic from their introductory psychology class in college classroom) and wrote about how it connected to their lives, or to the lives of someone they knew. In the control condition, participants selected a topic and wrote a summary of what they had been learning in the laboratory or the classroom.

Reading First is a policy intervention intended to ensure all children can read at grade level or above. Reading First provides funds to Title I schools to increase the amount of teacher professional development to K-3 teachers regarding scientifically-based reading instruction. The exact nature of the additional resources is up to each individual school to determine. For more details see U.S. Department of Education (2008).

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

All three samples utilized randomized control trials. Sample 1 and 2 were randomized at the student level, whereas the Reading First sample a nationally-representative sample of Reading First schools and non-Reading First Title I schools (U.S. Department of Education, 2008).

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Data was collected individually from participants in Samples 1 and 2. Data were analyzed using multiple regression. In the Reading First sample, a nationally-representative sample of Reading First and non-Reading First Title I schools were collected (Moss et al., 2008). We used the raw data from the implementation report (e.g., means and SD's, frequencies, etc.) to calculate our indices of achieved relative strength. Achieved relative strength (ARS) indices are calculated as the degree of fidelity within the treatment group minus the degree of fidelity within the control group divided by the pooled standard deviation (see Hulleman & Cordray, 2009, for more details on calculations). ARS indices represent the standardized difference in treatment implementation between the treatment and control group, and as such are indicative of relative treatment strength (Cordray & Pion, 2006).

Findings / Results:
Description of main findings with specific details.

In two randomized control trials of a motivational intervention, one in the laboratory and the other in the classroom, we demonstrate that the ARS indices are positively correlated with student motivational outcomes. In secondary analyses of Reading First data we show how ARS indices can be created and combined in more complex, multi-component interventions. One of the principle findings in this sample is that implementation fidelity of Reading First core components is already high in the control condition, thus diminishing the impact of randomization to the Reading First program.

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

Across three randomized experiments, we demonstrated how to compute the ARS and how it can be utilized to understand the results of experiments. The motivation intervention examples demonstrate consistency across three different methods of calculating achieved relative strength, and how treatment fidelity diffuses when moving from more controlled to less controlled settings. The Reading First example demonstrates the importance of assessing core intervention components in control conditions as the average level of implementation of Reading First components in control conditions was quite high, thus diminishing the achieved relative strength of the intervention.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

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

Cordray, D. S. & Hulleman, C. (2009, June). Assessing intervention fidelity: Models, methods and modes of analysis. Presentation at the Institute for Education Sciences 2009 Research Conference, Washington, D.C.

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

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*, *28*, 237-256.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Hulleman, C. S., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Intervention Effectiveness*, *2*(1), 88-110.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *Reading First Implementation Evaluation Final Report,* Washington, D.C., 2008.

Appendix B. Tables and Figures

Not included in page count.