Abstract Title Page

Title:

Rigorous measures of implementation: A methodological framework for evaluating innovative STEM programs

Author(s):

Amy Cassata-Widera, Jeanne Century, Dae Y. Kim
Center for Elementary Math and Science Education
University of Chicago

Abstract

Background / Context:

With the goal of preparing all students with the science, technology, engineering, and mathematics (STEM) skills needed to succeed in a 21st-century technological economy, improving the extent and quality of STEM education has become a national priority (National Science Board, 2009; U.S. Department of Education, 2009). Questions of effectiveness of innovative STEM interventions are increasingly visible and frequent, as a multitude of instructional materials and other programs to enhance STEM education are adopted by schools, districts, and states. Today, several popular comprehensive elementary mathematics curriculum programs exist that reflect the recommendations of the NCTM Standards (NCTM, 2000) and the National Science Education Standards (NSES, 1996). As we move into a new era, shaped by the Common Core standards, the role of instructional materials is becoming increasingly important. In addition, we are seeing increasing numbers of other kinds of STEM interventions established across states in response to receiving federal Race to the Top funding. The RttT competition, which includes STEM education as a "competitive preference priority," prioritizes reforms in which states provided a "high quality" plan for STEM education that offered rigorous content, teacher preparation, and applied learning opportunities for students (U.S. Department of Education, 2009).

There is increasing recognition that careful description and measurement of fidelity of implementation (FOI) of STEM interventions are essential if we are to know, for example, which components of programs bolster or hinder student performance, or the differential effects of incomplete or incorrect implementation (Fullan, 1983; Lynch & O'Donnell, 2005; Wang, Nojan, Strom & Walberg, 1984; Ruiz-Primo, 2005). However, because there has not been a coherent body of literature underlying FOI measurement, there has at times been a tendency to overlook the complexity of measuring implementation and rely on single data source or otherwise broad, unidimensional FOI measures (Century, Rudnick & Freeman, 2010). A second problem in determining program effectiveness involves accounting for overlap between treatment and control conditions in experimental studies carried out in educational contexts, where treatment and control groups are not completely independent from one another. Even when the intervention in name is not present, structures and interactions characteristic 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. Finally, researchers have yet to develop a shared conceptual understanding of what fidelity of implementation is and how to measure it. Thus, they have often created implementation measures based on operational definitions that are unique to each instrument's purpose and the perspective of its designers and grounded in local and limited theories about fidelity of implementation, with less generalized utility for other projects (Century et al., 2010; Levy et al., 2008).

Purpose / Objective / Research Question / Focus of Study:

The practical need for multidimensional measures of fidelity of implementation (FOI) of reform-based STEM instructional materials, combined with a theoretical need in the field for a shared conceptual framework that could support accumulating knowledge on specific enacted program

elements across evaluations of reform-based STEM innovations, has informed CEMSE's approach to fidelity of implementation measurement. Beginning in 2007, CEMSE received support from the National Science Foundation for its "Applied Research on Science Materials Implementation: Bringing Measurement of Fidelity of Implementation to Scale" project (*Award* ESI-*0628052*). Through this project, the CEMSE team developed, pilot tested, field tested, and revised a suite of eight instruments for measuring the fidelity of implementation (FOI) of reform-based K-8 science and mathematics instructional materials programs. These instruments, which provide a variety of data collection approaches, are unique among those that measure enactment of programs in that they focus on clearly and specifically describing the nature of program implementation, using constructs representing the essential elements of reform-based mathematics and science instructional materials programs organized into a conceptual framework that facilitates their application across multiple programs (see Appendix B).

Our framework for measuring FOI, based on a review of literature on implementation across a number of fields, recognizes essential program elements as ''critical components'' (Bauman, Stein, & Ireys, 1991; Ruiz-Primo, 2005; Wang et al., 1984). We distinguish between the two broad types of critical components within an intervention: *Structural Critical Components*, and *Interactional Critical Components*. In educational contexts, these broad categories may be further subdivided. Structures that are *procedural* in nature provide the specific organizing structural elements of the program that focus on what the user needs to *do* and the ways the intervention is physically organized to communicate intentions to the user (e.g., schedules, physical location, presence of people or materials). Structures that are *educative* describe those elements of the innovation that, while part of the innovation itself, support the enactment of the other critical components of the innovation (e.g., professional development structures, structured educational materials provided in curricula). Similarly, interactional critical components may be further categorized, into those that are *pedagogical*, representing the actions, behaviors, and interactions that users are expected to engage in when enacting the intervention, including the users' interactions with the recipients of the intervention (e.g., facilitating discussion, asking learners to reflect, planning activities). In addition, there are interactional *learner engagement* critical components that describe what learners do when participating in the enactment of the intervention (e.g., engaging in intellectually challenging work, taking responsibility for learning, contributing to decision making).

Significance / Novelty of study:

By explicitly measuring the different elements of educational interventions organized according to this framework (the "FOI Framework"), our approach provides an improved way to understand and measure program implementation that reflects the complex, multidimensional reality of interventions enacted in classroom settings. Measuring the presence and level of enactment of structural and interactional critical components enables us to distinguish between the quality of the treatment as well as the degree of the treatment. In addition to providing a more detailed and nuanced method for measuring enactment of multiple aspects of interventions, using a critical components approach to conceptualize program implementation also addresses the existing problems in the field related to accounting for fidelity of implementation within experimental studies. Our tools provide a way to open up the "black box" of what happens in the treatment group to explore more nuanced relationships between interventions and outcomes.

Furthermore, by measuring the extent to which structural and interactional critical components are present or not in both the treatment and control groups, we can better quantify the difference between these groups, accounting for variability in implementation of the intervention within the treatment group, while also accounting for contamination in the control group or shared interactional elements between the treatment and control group. Finally, the FOI framework enables the description of each program as comprised of structural and interactional critical components that may be shared across programs or unique to specific programs. With the notion of "shared critical components," our framework provides a common foundation for identifying, defining, and organizing the essential elements of interventions to enable collecting data and accumulating knowledge across them.

Usefulness / Applicability of Method:

Instrument validation: In contrast to existing instruments, which only provide classroom observational measures of reform-based STEM instruction, we have created a range of instruments measuring the operationalization of identified critical components, including observation, interview, and self-report measures for teachers and school leaders. Each of the critical components is measured in more than one instrument in the suite. The suite of FOI instruments for measuring the fidelity of implementation (FOI) of reform-based K-8 science and mathematics instructional materials programs was field tested within 39 Chicago Public Schools. Item reduction and validation proceeded with the goal of retaining items identified as likely to form a highly reliable and valid scale for each critical component (i.e., items that would strongly correlate with other items measuring the same critical component, but weakly correlate with other items measuring different critical components). In addition, we examined basic descriptive statistics for each item, as well as inter-item correlations, internal consistency using Cronbach's alpha, and factorial structures at the critical component level.[1] We have received funding from the Institute of Education Sciences to further validate this suite of instruments across 50 schools in three states. As part of this study, we will examine the extent to which composite indices on the four FOI subcategories predict student achievement in science and mathematics.

Operationalizing the process of new instrument development: Although our previous work has focused on developing instruments for measuring enactment of particular mathematics and science programs, we have come to understand that the framework we created for measuring use of instructional materials could be applied to other kinds of interventions and innovations. Using a process that includes identifying the "critical components" of the intervention, organizing them in a framework and systematically measuring them, we can learn which elements work, which do not, and how to adapt the innovation over time to accomplish our intended goals for STEM education in the future. To identify the critical components of an educational intervention and develop a suite of instruments that measure the enactment of these critical components, we begin by meeting with the people most familiar with the program (e.g., program leaders, program founders, curriculum developers). Following the process we previously developed to apply our instruments to other programs, we first ask the expert(s) to articulate the theory of action for the program and in turn the program critical components that represent the operationalization of this

[1]Detailed information on critical component definitions, FOI instruments, and psychometric properties of items measuring enactment of K-8 reform based mathematics and science instructional materials is available at: http://www.researcherswithoutborders.org/foi_users_guide

theory. Specifically, we identify the program leader's expectations for what each user needs to do, the expectations for each user's behaviors and interactions with learners (interactional–pedagogical), and the expectations for the learners themselves (interactional–learner engagement). Simultaneously, we gather information about the program published in its written products and marketing materials. Our work with the program leader(s) in modifying and creating critical components and related items involves a collaborative and iterative process, requiring adjustments and revisions as the process unfolds.

Adapting the FOI Framework to measure enactment of STEM School Models: CEMSE, in collaboration with the Battelle Center for Mathematics and Science Education Policy in the John Glenn School of Public Affairs at the Ohio State University and the Ohio STEM Learning Network (OSLN), is in the process of conducting a three-year NSF-funded study to examine the factors affecting the implementation, spread and sustainability of innovative STEM teaching and learning at the schools that are part of OSLN's Platform Schools Initiative. Data on the status of implementation of the school STEM model will be grounded in the suite of instruments developed for CEMSE's FOI project that will be adapted for this purpose. We have worked with school teams in an iterative process to describe each of the design principles and articulate how the school understands those principles and how they are represented in the structural and interactional critical components of their STEM Platform school model. We have also asked the school teams to articulate additional critical components of the school models that are not tied to specific design principles. This process has allowed us to identify critical components that are shared and unique across each of the platform schools. We are in the process of placing the identified components into one of two groups – those that apply to the whole school, and those that apply to the learning experience, and then organizing the each group into a framework using the FOI framework model. These frameworks then become the basis for creating two sets of instruments for each school (one to measure enactment at the learning experience level and the other to measure enactment at the school level), using existing validated FOI items from the existing FOI instruments where applicable.

Conclusions:

Until we are able to measure FOI accurately and with specificity, we cannot know with certainty whether a program is effective and the nature of the implementation that leads to desired outcomes. The FOI Framework focuses on providing a means to clearly and specifically describe the nature of enactment of innovative STEM programs. Developed with the goal of enabling rigorous, specific, and systematic analysis of implementation across multiple programs, the framework is unique among existing methods for measurement of fidelity of implementation. We aim to continue to refine the process of expanding the use of the framework to create instruments to measure multiple 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 (Century et al., 2010). In the long term, we hope that the use of the FOI framework may further our collective understanding of STEM interventions and our ability to accumulate knowledge about the elements that make them most effective.

Appendices

Appendix A. References

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.

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

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

Levy, A.J., Pasquale, M.M. & Marco, L. (2008, Fall). Models of providing science instruction in the elementary grades: A research agenda to inform decision makers. *The Science Educator, 17*(2), 1-18.

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. Unpublished manuscript.

National Science Board (2009). National Science Board STEM Education Recommendations for the President-Elect Obama Administration. Retrieved from http://www.nsf.gov/nsb/publications/2009/01_10_stem_rec_obama.pdf

NCTM, National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics. Retrieved from http://standards.nctm.org/

NSES, National Science Education Standards (1996). Washington, D.C., National Academies Press. Retrieved from http://www.nap.edu/html/nses/

U.S. Department of Education (2009, November). Race to the top program: Executive Summary. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Ruiz-Primo, M. (2005). A multi-method and multisource approach for studying fidelity of implementation. Paper presented at the annual meeting of the American Education Research Association, Montreal, Canada.

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*(3), 249-286.

Appendix B. Tables and Figures

	FOI of K-8 Reform-Based Mathematics and Science Instructional Materials			
	Structural Critical Components		Interactional Critical Components	
	Procedural	Educative	Pedagogical	Learner Engagement
Common to Mathematics and Science	1. Duration of unit 2. Time spent on instruction 3. Lesson order 4. Order of segments and parts within lesson 5. Inclusion of all essential segments within lesson 6. Inclusion of all essential lessons 7. Lesson overview 8. Lesson preparation 9. Materials 10. Writing structures 11. Readings 12. Assessments and assessment tools 13. Content of lesson a. Facts b. Procedures c. Concepts d. Processes 14. Class structures 15. Instructional delivery formats 16. Projects* 17. Extensions* a. Discipline-related b. Non-discipline-related 18. Additional resources* 19. Homework*	1. Content background information 2. Pedagogy background information 3. National standards and benchmarks information* 4. Lesson notes	1. Teacher facilitation of small group work 2. Teacher facilitation of student discussion 3. Teacher facilitation of students doing potentially intellectually challenging work 4. Teacher emphasis on types of content 5. Teacher facilitation of student autonomy 6. Teacher facilitation of students taking risks 7. Teacher facilitation of student interest 8. Teacher facilitation of materials, manipulatives, and tools use 9. Teacher use of assessment to inform instruction 10. Teacher use of differentiation	1. Students contribute to small group work 2. Students engage in discussion 3. Students engage in potentially intellectually challenging work 4. Students demonstrate autonomy 5. Students take risks 6. Students do/complete essential activities 7. Students do/complete optional or non-essential activities
Common to Science		A. Safety	A. Teacher facilitation of student data collection	A. Students collect data
Common to Mathematics	A. Unit Order		A. Teacher supports multiple solution strategies	A. Students use multiple solution strategies