



Issues of rigor and feasibility when observing the quality of program implementation: A case study



Paul R. Brandon^{*}, Brian E. Lawton, George M. Harrison

University of Hawaii at Manoa, Curriculum Research & Development Group, 1776 University Avenue, Honolulu, HI 96822, United States

ARTICLE INFO

Article history:

Received 20 July 2013

Received in revised form 30 January 2014

Accepted 5 February 2014

Available online 14 February 2014

Keywords:

Quality of program implementation

Observations

Tradeoffs between feasibility and rigor

ABSTRACT

Program evaluators have paid little attention in the literature to the manner in which measuring the quality of implementation with observations requires tradeoffs between rigor (reliability and validity) and program evaluation feasibility. We present a case example of how we addressed rigor in light of feasibility concerns when developing and conducting observations for measuring the quality of implementation of a small education professional development program. We discuss the results of meta-evaluative analyses of the reliability of the quality observations, and we present conclusions about conducting observations in a rigorous and feasible manner. The results show that the feasibility constraints that we faced did not notably reduce the rigor of our methods.

© 2014 Elsevier Ltd. All rights reserved.

In all components of evaluations, the more rigorous the study, the more reliable and valid the inferences that the study will produce. Rigorous evaluation methods can be expensive, however, and evaluators developing and implementing methods sometimes have to yield to financial, personnel, and time constraints (Braverman, 2013). For example, evaluators conducting observations when examining program fidelity of implementation (FOI)—particularly in small or mid-sized evaluations of modestly funded programs—might have to consider tradeoffs between the rigor of observations and the feasibility of conducting them fully and well. In the parlance of the widely used Joint Committee on Evaluation Program Evaluation Standards (Yarborough, Shulha, Hopson, & Caruthers, 2011), evaluators must attend to the *accuracy* program evaluation standards (i.e., rigor of design and methods, with particular attention given to reliability and validity) in light of the *feasibility* program evaluation standards (e.g., using resources efficiently and using practical procedures).

As far as we know, no empirical findings about the effects of tradeoffs between feasibility and accuracy have been reported in the FOI literature. To help address this deficit, we present a case example of how we addressed issues of rigor and technical adequacy within fiscal and human resource constraints when measuring the quality of implementation of a teacher professional

development (PD) project. The presentation of the case example is consistent with the widespread practice of publishing reflective case narratives describing what evaluators have learned (Cousins & Chouinard, 2012). We describe our methods for examining quality, and we present the results of analyses of validity and reliability that we conducted during the evaluation. We follow this with meta-evaluative analyses of reliability that we conducted for the purposes of this article. The follow-up analyses address the issue of how well we ensured rigor within the feasibility constraints of the evaluation.

Our purpose in this article is to inform evaluators who are measuring quality in small or mid-sized evaluations, typically at the state or local level, with particular relevance for education evaluators. For evaluators in other sectors and for the evaluation community as a whole, we intend that the article will broaden the discussion about attending to evaluation rigor in light of the exigencies of typical evaluation contexts. We hope that the details we provide serve to elucidate the kinds of issues that evaluators might have to address and the tradeoffs that they might have to accept. Fortunately, our results show that the tradeoffs between feasibility and accuracy need not necessarily result in invalid inferences.

1. Background

It is increasingly commonplace for evaluators of education or social programs to examine FOI. FOI is defined as the degree to which program personnel conduct programs in a manner reflecting

^{*} Corresponding author. Tel.: +1 808 956 4928.

E-mail addresses: brandon@hawaii.edu (P.R. Brandon), blawton@hawaii.edu (B.E. Lawton), georgeha@hawaii.edu (G.M. Harrison).

program developers' and program deliverers' intentions—that is, the extent to which programs are implemented fully and well. Several aspects of FOI are frequently addressed in the evaluation literature: (a) *exposure, dosage, or duration*—the frequency with which program activities are implemented and the length of implementation, (b) *adherence*—the extent to which the components and activities of a program are implemented, (c) *quality*—the extent to which program activities are delivered competently and meritoriously, with integrity to the prescribed manner of delivery, and (d) *participant responsiveness*—the degree of program participants' (e.g., teachers in a professional development program or students in an instructional program) involvement in, and enthusiasm for, program activities (Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003; O'Donnell, 2008). FOI findings can be used for formative evaluation purposes by helping program personnel fine-tune their endeavors and for summative evaluation purposes by showing the degree to which implementation affects outcomes, and ultimately, the success of a program. FOI findings can even be applied immediately to give feedback to program personnel such as teachers, thus serving a formative personnel evaluation purpose.

As Fagan, Hanson, Hawkins, and Arthur, (2008) stated, the quality of implementation is “more subjective than adherence and dosage and, therefore, more difficult to define and measure” (p. 258). Quality is nuanced and occurs to varying degrees; it is best recorded as a continuous variable. Thus, it is best examined through live observations or recordings of program implementation. Observations are direct measures that are likely to be sensitive to (a) the competence of delivery, (b) nuances in the project frequency, intensity, and sequencing of intervention procedures (Moncher & Prinz, 1991), and (c) the degree to which project procedures are adapted to implementation circumstances and contexts (Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005). Observations are also popular with project personnel, who find their results useful for formative evaluation purposes.

Systematically observing programs to measure quality can be anything but efficient and practical. Observations require the comprehensive training of observers, scheduling and coordinating observations or videotaping, the sometimes multifaceted and lengthy process of making quality judgments, and evidence of acceptable inter-observer reliability. Observers must be prepared to participate in training, coordinate their schedules, and devote considerable time and attention when making judgments about quality, which might require more time than expected. Program participants such as teachers must make their classrooms available, risk disrupting their schedules or instruction, and contend with the complications of having observers in the classroom. All this must be accomplished with the limited personnel and financial resources common in small or mid-sized evaluations.

2. Description of the case example

The evaluation in our case examined the Arts First Windward Research Project (AFWRP), a small arts-integration PD project for elementary-school teachers. It was conducted by a principal investigator funded at .20 of his time and a full-time project manager/researcher. AFWRP drew heavily from recent research on arts integration in elementary and secondary schooling (e.g., Bresler, 1995; Krug & Cohen-Evron, 2000; Wiggins, 2001). It was funded by a grant to the Hawai'i Arts Alliance from the United States Department of Education's Arts in Education Model Development and Dissemination (AEMDD) program. The project endeavored to train teachers how to integrate the arts into reading comprehension instruction. The project staff included a manager, who also served as the primary teacher trainer, and five teaching artists who assisted in the training and served as teacher classroom

mentors. The PD taught teachers how to have their students (a) observe details closely to gain a deeper understanding of the subject material, (b) identify patterns of elements in the subject material, and (c) represent the details and patterns through drama or dance activities. After identifying the meaningful components of a story and representing them through movement in the classroom, the students reflected with their peers and teachers on the quality of their representations and on their overall understanding of the target reading material.

Over the course of the project, the personnel, with some external consultants, developed and refined approaches to using the arts in reading comprehension instruction, resulting in several arts strategies. The strategies drew on drama and dance techniques for the purposes of improving students' reading achievement, engaging students in learning and increasing their interest in the arts, and changing participating teachers' pedagogical practices. The project manager and the participating arts educators trained and guided elementary school teachers (mostly in Grades 3–5) in how to use the strategies. They provided PD in summer institutes, conducted follow-up training days, and provided one-on-one mentoring by trained arts educators. This approach was consistent with the recommendations of PD theory and empirical findings (Desimone, 2009).

The AFWRP evaluation was intended to provide project personnel with information about project quality for formative evaluation purposes and to gather information about the quality of implementation to use in the final summative evaluation (Brandon, Lawton, & Krohn-Ching, 2007). One of the major features of the evaluation was to examine FOI. We initially collected data only on the adherence and exposure aspects of FOI with self-report questionnaires and interviews, but as the project progressed, it became apparent that we also needed to examine the quality of teachers' implementation of the arts strategies. This evolution, occurring without having budgeted for observations, was like many evaluative endeavors (Smith, 2012), including those we have conducted over the years, in which evaluation designs and methods change as programs develop. Without data on quality, we would be unable to provide important information useful for (a) improving project implementation, (b) providing feedback to teachers about their progress, or (c) reporting how well the project was implemented. Accordingly, toward the end of the project, we collaborated with project personnel to develop an observation measure that we call the Quality of Implementation Protocol (QIP), with the project (not the evaluation) covering observer expenses. We provide a copy of the instrument in Appendix A. The purpose of developing the QIP was to have a method for evaluating the extent to which the AFWRP teachers showed a sophisticated use of the arts strategies in the prescribed manner, with the results to be useful for formative project evaluation, formative teacher evaluation, and summative project evaluation.

The steps in developing and using the instrument were to identify and refine the criteria for judging quality, prepare and revise the observation form, train the observers, and conduct the observations, as follows.

3. Developing and refining the QIP criteria

Our first step in developing the instrument was to generate a list of all the criteria that it might address. The goal of this step was to have a set of criteria, with definitions, that the observers would use for rating quality, focusing on the project's primary features (Becker, Smith, Tanzman, Drake, & Tremblay, 2001; Blakely et al., 1987; Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008; Ruiz-Primo, 2005). These criteria were intended to address aspects of arts integration classrooms, based on the project's theory, that the teachers had learned and practiced in the project PD. Matching criteria to a project's features and theory is the first step in

Download English Version:

<https://daneshyari.com/en/article/319421>

Download Persian Version:

<https://daneshyari.com/article/319421>

[Daneshyari.com](https://daneshyari.com)