Studies in Educational Evaluation 42 (2014) 109–117

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

Using logic model mapping to evaluate program fidelity

Lisa R. Holliday *

The Evaluation Group, 169 Laurelhurst Avenue, Columbia, SC 29210, USA

ARTICLE INFO

Article history:
Received 2 August 2013
Received in revised form 29 March 2014
Accepted 4 April 2014
Available online 5 May 2014

Keywords:
Fidelity
Program integrity
Focus groups
Qualitative data
Logic model mapping

ABSTRACT

School districts that implement multisite interventions often run into numerous challenges. Unique school needs and cultures, as well as staff buy-in and support for the intervention can affect the way in which programs are implemented. This study looks at four schools in a rural school district that were in their first year of implementing a program aimed at providing support to teachers. It evaluates fidelity by using logic model mapping to connect qualitative data from focus groups to the program's logic model. It seeks to address whether implementation challenges encountered by schools impacted short-term program outcomes.

Introduction

School districts that implement multisite programs often run into numerous challenges. Unique school needs and cultures, as well as staff buy-in and support for the program, can affect the way in which programs are implemented. Research has shown these factors can promote program drift and adaptation (Bickman & Peterson, 1990). While some adaptation is necessary to meet a school's individual needs, too much adaptation may result in a loss of fidelity to the original program model.

Monitoring fidelity early in the implementation stages of a program provides an opportunity to identify areas where schools are encountering implementation challenges and to make necessary modifications to keep the program on track (Gingiss, Roberts-Gray, & Boerm, 2006; Zvoch, 2009). Broadly, fidelity is the "extent to which an enacted program is consistent with the intended program model" (Century, Rudnick, & Freeman, 2010, p. 202). For multisite projects, qualitative data provide an excellent window into the process of implementation and have the ability to highlight unanticipated difficulties schools may be encountering (Leech & Onwuegbuzie, 2007). By connecting these data to the program logic model, individual school-level implementation issues can be assessed in the aggregate. This type of analysis allows evaluators to pinpoint where challenges encountered at the school level are having the greatest impact at the district level. In turn, this lets districts implementing multi-site programs make modifications to the program if necessary.

This study looks at four schools in a rural school district that were in their first year of implementing a program aimed at providing support to teachers. It evaluates fidelity by using logic model mapping to connect qualitative data from focus groups to the program's logic model. It asks the question, "Did fidelity of implementation challenges encountered by schools impact short-term program outcomes?"

Why evaluate fidelity?

Without a measure of fidelity, it is impossible to discern if a program is being applied, whether the measure of outcomes is a result of the intervention or other factors, or if there is a weak program design (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Hall & Loucks, 1977; Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008). Fidelity measures help researchers to interpret ambiguous and negative findings by adding an interpretative context, as well as ensuring internal validity (Cook & Campbell, 1979; Hohmann & Shear, 2002; O'Donnell, 2008). Additionally, measures of program integrity help to increase external validity, which improves the possibility of replicating the program (Cook & Campbell, 1979). Last, fidelity helps to deter researchers from making a Type III error (Domitrovich & Greenberg, 2002).

With multisite programs, measuring treatment fidelity provides a picture of how the program is being implemented at each

* Correspondence to: 169 Laurelhurst Avenue, Columbia SC 29210, USA. Tel.: +1 704 996 5291.
E-mail address: Lisa.R.Holliday@gmail.com

http://dx.doi.org/10.1016/j.stueduc.2014.04.001
0191-491X/

site, including how it may be adapted. Without fidelity implementation data from "each site, we cannot assume that the model and its components have been equally well implemented at different sites" (Fullan, 1983, p. 218). Mowbray et al. (2003) noted that in multi-site interventions, fidelity criteria are also needed in order to document changes to the program. By studying implementation, researchers can uncover how a program functions, how users interact with it, and what obstacles users face during implementation (Mowbray et al., 2003).

By tracking fidelity, researchers can observe adaptations to programs, which are any changes to the core components of the program model (Century et al., 2010). Changes can range from modifications to the length of a training session to alterations in program delivery (Blakely et al., 1987). However, adaptation can be a necessary step in program implementation (Blakely et al., 1987; Mowbray et al., 2003). Mowbray et al. (2003) noted that adaptation occurs in response to the needs of the target population and is influenced by budgets and available resources.

There is not a current consensus among researchers as to how much adaptation can occur before outcomes are affected (Gresham et al., 2000). Some contend that adherence to the program model is essential for a successful intervention (Dumas, Lynch, Laughlin, Philips, & Prinz, 2001). Hill, Maucione, and Hood (2007) noted that this view is increasingly prevalent among state and federal funding agencies, which emphasize the potential of interventions to serve as models or best practice programs.

Research indicates that optimum fidelity may vary based upon a multitude of factors, such as context, target population, quality of the intervention, and available support. While it has not been established whether high fidelity (85% or greater) consistently leads to high outcomes, research suggests that for proven interventions treatment integrity is directly related to the degree of treatment outcome (Gresham, Gansle, & Noell, 1993). In a large analysis that looked at the results from over 500 studies, Durlak and DuPre (2008) found that the degree of program implementation impacted program outcomes. Watson and McCurdy (as cited in Gresham et al., 2000, p. 203) found that 60–65% fidelity was effective in achieving desired results, and Gresham et al. (2000) found that "the median correlation between level of treatment integrity and treatment outcome was .54, suggesting that higher integrity was associated with larger effect sizes" (p. 202). However, researchers have also noted that programs have achieved desired outcomes with lower levels of fidelity (Durlak & DuPre, 2008). Wickstrom, Jones, LaFleur, and Witt (1998) found that low implementation of a behavioral consultation program did not correspond to similarly low outcomes. Schulte, Eaton, and Parker (2009) observed that some parts of an intervention may be more important to program outcomes than others, and high fidelity to the program model may not be necessary, though higher integrity tends to be associated with better outcomes. Zvoch (2009) found that high fidelity is associated with improved outcomes in some studies, but in others, no clear relationship between fidelity and outcomes was identified. He concluded that these contradictory findings suggest that neither strict program fidelity nor widespread adaptation are likely to achieve desired outcomes. Rather, programs must find the balance that works best for their particular circumstances and needs. Durlak and DuPre (2008) argued that researchers should focus on identifying the "right mix of fidelity and adaptation" which promotes optimal program results (p. 341).

The quality and appropriateness of the program being implemented have bearing on the level of fidelity needed to achieve desired outcomes. High levels of fidelity to a poorly designed intervention will not likely result in desired outcomes, whereas high levels of fidelity to well-designed interventions *may* result in desired outcomes. For programs that are new and have not been previously implemented, this means that the optimum level of fidelity is unknown until the conclusion of the program.

Monitoring both fidelity and adaptation provides additional information on how programs function in real-world settings, where adaptation may be necessary (Blakely et al., 1987; Durlak & DuPre, 2008; Mowbray et al., 2003). From this stance, adaptations should not be regarded as program failures (Durlak & DuPre, 2008). This view allows researchers to pinpoint which program components are essential for success and which may need to be altered or removed entirely (Mowbray et al., 2003).

A number of factors that influence fidelity have been identified. Providing ample resources throughout the duration of the intervention has been associated with promoting fidelity (Chi-Ming, Greenberg, & Walls, 2003; Durlak & DuPre, 2008; Gresham et al., 2000). Information monitoring and feedback have also been found to promote fidelity (Bickman et al., 2009; Fullan, 1983; Ringwalt et al., 2010). Gingiss (1992), and Chi-Ming et al. (2003) observed that support from principals and administrators helped to encourage fidelity. Buston, Wight, Hart, and Scott (2002) found that interventions that were easy to understand and deliver, provided beneficial and observable changes, and did not require additional resources were more likely to have fidelity to the program model. Also, Buston et al. (2002) argued that the program must be a good fit with the pre-existing culture and practices at the site where it is being implemented. Turnbull (1999, 2002) found that teachers who had sufficient training and resources, as well as administrative support, were more likely to buy into a program than those who did not. However, she also found that factors which predicted high buy-in during Year One did not predict buy-in during Year Two, which indicates a need for ongoing fidelity monitoring. Lastly, Noel, Witt, Gilbertson, Ranier, and Freeland (1997) found that fidelity was facilitated by providing regular feedback to teachers during implementation.

There is no standardized way to measure fidelity, and the type of data used varies based upon the program being measured (Century et al., 2010). Many researchers emphasize collecting data on the quality, the dosage, and reach of the treatment delivered (Kaufman, Perry, Hepburn, & Duran, 2012). One of the more widely used methods to measure fidelity are ratings based upon best-practices checklists applied to record reviews, surveys, observations, and interviews (Bickman et al., 2009, p. 77). This approach typically utilizes thresholds to establish acceptable levels of fidelity. However, fidelity thresholds may not be appropriate for all types of programs. Assessing fidelity using any variation of a "met/not met" standard may prompt misgivings among program staff, particularly in environments where evaluation results are tied to funding opportunities, as with some grants, and in instances where fidelity findings may be politicized, as with some educational reforms. Focusing on the relationships between program activities and outcomes is an alternative to the threshold approach. This method allows researchers to identify challenges sites encounter when implementing programs and to explore how these challenges affect the relationship between activities and outcomes. Additionally, this type of fidelity check is well-suited to the early stages of program implementation, and it provides valuable information that program directors can use to make modifications to service delivery.

Description of the intervention

This study focused on a five-year federally grant funded program designed to improve the quality and amount of support