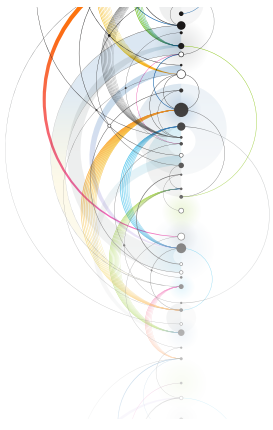


Abstract

Implementation fidelity assessment provides a means of measuring the alignment between the planned program and the implemented program. Unfortunately, the implemented program can differ from the planned program, resulting in ambiguous inferences about the planned program's effectiveness (i.e., it is uncertain if poor results are due to an ineffective program or poor implementation). We demonstrate how inclusion of implementation fidelity in the outcomes assessment process increases the validity of inferences about program effectiveness and, ultimately, student learning. Although our didactic discussion of implementation fidelity focuses on its importance to assessing student affairs programming, the concepts and process are applicable to academic programs as well.

**AUTHORS**

Jerusha J. Gerstner, M.A.
James Madison University

Sara J. Finney, Ph.D.
James Madison University

Measuring the Implementation Fidelity of Student Affairs Programs: A Critical Component of the Outcomes Assessment Cycle

What is Implementation Fidelity and Why is it Important?

Implementation fidelity has been discussed in many domains (e.g., K-12 education, health, psychology). As a result, numerous definitions of implementation fidelity exist. The general definition provided by O'Donnell (2008) is “the determination of how well an intervention is implemented in comparison with the original program design during an efficacy and/or effectiveness study” (p. 33). Specific to the student affairs context, implementation fidelity examines the extent to which the planned student affairs program matches the implemented program. That is, student affairs programs (or any educational program) should be designed thoughtfully to meet particular learning and development outcomes. However, as Berman and McLaughlin (1976) noted, “The bridge between a promising idea and the impact on students is implementation, but innovations are seldom implemented as intended” (p. 349). Importantly, research has shown that programs implemented with high fidelity have more of an impact with respect to program outcomes than those with low fidelity (e.g., Durlak & DuPre, 2008). Thus, higher education practitioners and instructors need to ask themselves, “Are students receiving the *planned* program?”

Deviations from the planned program may involve excluding critical program components or curriculum, shortening program sessions or classes, changing the mode of program delivery, or adding extraneous information or activities (Ball & Christ, 2012). Program deviation or drift may occur for many reasons, including poor training of program implementers (e.g., instructors, facilitators, interventionists), lack of motivation of implementers, or insufficient time provided for program components (Century, Cassata, Rudnick, & Freeman, 2012; Durlak & DuPre, 2008; Lane, Bocian, MacMillan, & Gresham, 2004). Drift “refers to the unplanned, gradual altering of the implementation of an intervention by the interventionist” (Hagermoser Sanetti & Kratochwill, 2009a, p. 452). We agree with

CORRESPONDENCE

Email
gerstnjj@dukes.jmu.edu

Hagermoser Sanetti and Kratochwill that some flexibility in program implementation should be tolerated, but “such flexibility does not justify, however, an acceptance of interventionist drift, which may result from a host of factors” such as “forgetting intervention components, having limited resources, [or] believing the intervention requires too much response effort” (p. 452). Implementation fidelity assessment allows for a direct evaluation of the degree of program drift. If the program is not implemented as planned, it should not be surprising when program outcomes are not achieved.

Informal conversations with student affairs professionals, numerous consultations regarding assessment of student affairs programs, and observations of professional presentations on practice at conferences suggest that student affairs professionals are not asking themselves this implementation fidelity question, which aligns with similar observations in other domains (e.g., Cochrane & Laux, 2008; Hagermoser Sanetti & Kratochwill, 2009a). If the question is being asked, implementation fidelity results appear to be neither gathered nor reported; thus, the alignment between the planned and implemented program is not known. This lack of fidelity information greatly limits interpretation of outcomes assessment results and, ultimately, evaluation of the planned program (Ball & Christ, 2012). For instance, if an outcome measure is mapped directly to an objective and students are performing poorly on this measure, it could be inferred that students are not meeting this objective as a function of the planned program. However, if the programming aligned with this objective is not implemented as planned, the outcome measure reveals nothing about the efficacy of the planned program, because the planned program was not administered. In fact, the planned program may impact student learning and development in a powerful way if implemented correctly. The combination of low implementation fidelity and the lack of its assessment can result in changing or terminating a program that would be effective if implemented as planned (Hagermoser Sanetti & Kratochwill, 2008).

Implementation fidelity examines the extent to which the planned student affairs program matches the implemented program.

Although implementation fidelity has been a topic of interest and research in health-related fields (e.g., Breitenstein et al., 2010; Garvey, Julion, Fogg, Kartovil, & Gross, 2006) and K-12 education (e.g., Ball & Christ, 2012; Cochrane & Laux, 2008; Hagermoser Sanetti & Kratochwill, 2009a), a review of several highly-esteemed books focused on assessment in higher education and student affairs uncovers no mention of implementation fidelity (American College Personnel Association [ACPA], 2006; Bresciani, Gardner, & Hickmott, 2009; Erwin, 1991; Schuh, 2009; Schuh & Upcraft, 2001; Upcraft & Schuh, 1996). Moreover, authors, such as Shutt, Garrett, Lynch, and Dean (2012), have provided recommendations regarding how best practice, with respect to student affairs programs, centers on the assessment process: “In essence, then, the intentional use of the assessment process itself is what constitutes best practice” (p. 71). We could not agree more and echo their call for empirically-based programs and curriculum. However, the importance of implementation fidelity data for making valid inferences about program effectiveness was not stressed, much less was the process of collecting and using fidelity data to evaluate program efficacy. Nonetheless, there is clearly a place for implementation fidelity assessment within all outcomes assessment processes. Despite the lack of coverage in the higher education assessment literature, the concept of implementation fidelity is analogous to “process” or “implementation evaluation” discussed in the program evaluation literature (e.g., Patton, 1997; Posavac & Carey, 1997; Weiss, 1998). In addition, we applaud Aiken-Wisniewski et al. (2010) for not only discussing the concept of implementation fidelity (termed “process/delivery outcomes”) but also noting the importance of gathering implementation data in their *Guide to Assessment in Academic Advising*. Unfortunately, practitioners in units other than advising may be unaware of this document and its recommendations regarding implementation fidelity assessment.

Three possible reasons why implementation fidelity is not assessed center on untested assumptions, lack of understanding of implementation fidelity, and lack of guidance on the practice of collecting and using implementation fidelity data. First, practitioners may assume the program “on paper” is implemented as planned (e.g., Hagermoser Sanetti & Kratochwill, 2009a; O'Donnell, 2008). Namely, practitioners may assume implementation fidelity is high because program implementers *should* present the program exactly as directed. However, this is an assumption that needs to be tested, as research indicates this assumption is often wrong (e.g., Ball & Christ, 2012; Durlak & DuPre, 2008; Hagermoser Sanetti & Kratochwill, 2008, 2009b; Lane et al., 2004). Second, practitioners may not understand that low fidelity can

attenuate program effectiveness. Third, even if practitioners are concerned about program implementation and understand the impact of low implementation fidelity, they may not engage in fidelity assessment because they do not understand how to assess the alignment of the planned and implemented programs. These barriers to implementation fidelity assessment align with noted barriers in other, related domains. More specifically, in the domain of school psychology, the following barriers were uncovered with respect to collecting implementation fidelity data: lack of general knowledge of implementation fidelity, lack of guidelines on procedures to collect these data, lack of resources, and lack of requirements to collect these data (Cochrane & Laux, 2008; Hagermoser Sanetti & DiGennaro Reed, 2012).

Given the push for accountability in higher education (U.S. Department of Education, 2006) and the assessment, evaluation, and research standards established for student affairs (ACPA, 2006), we propose that the measurement of implementation fidelity is past due in higher education. Moreover, high quality program assessment must incorporate implementation fidelity into the outcomes assessment process. Our goals in this article are to explicate implementation fidelity's place within the outcomes assessment cycle, to provide insight into quantifying fidelity, and to provide an example of how implementation fidelity was used to strengthen a student affairs program on our campus.

Higher education practitioners and instructors need to ask themselves, “Are students receiving the planned program?”

Implementation Fidelity in the Outcomes Assessment Cycle

The Typical Outcomes Assessment Cycle

The outcomes assessment cycle is used to evaluate how well programming functions with respect to meeting student learning and development objectives. Many of these cycles include the following six steps: establishing objectives/outcomes, mapping programming to these objectives/outcomes, selecting or designing measures of the outcomes, collecting outcomes data, analyzing and maintaining outcomes data, and using outcomes information (e.g., ACPA, 2006; Aiken-Wisniewski et al., 2010; Bresciani et al., 2009; Erwin, 1991; Suskie, 2009). First, practitioners must establish program objectives. These objectives outline intended outcomes of the program: what students should be able to know, think, or do as a result of participating in the program. Objectives provide a clear, detailed presentation of the program's purpose. Second, various programming components are developed to align with the stated objectives. These program components can be conceptualized as treatments that should result in the particular outcomes stated in the objectives. The intentional creation and mapping of programming to objectives is a critical part of the outcomes assessment cycle. Third, outcome measures are selected or designed to quantify whether students are meeting the objectives after being exposed to programming. During the fourth and fifth steps, outcomes data are collected and then analyzed. Finally, the outcomes assessment results are used to evaluate program effectiveness, with a specific focus on making informed changes to the programming components revealed to be suboptimal.

Incorporating Implementation Fidelity into the Assessment Cycle

In the standard assessment cycle, one never evaluates whether the planned program was implemented. In fact, the term “black box” has been used to characterize the disguised nature of any information regarding implementation of the program in standard outcomes-based assessment (Mowbray, Holter, Teague, & Bybee, 2003; Nelson, Cordrary, Hulleman, Darrow, & Sommer, 2012). Without implementation fidelity assessment, nothing is known about what occurred during the program, only what was planned—which could be radically different from the actual implemented program. That is, in absence of fidelity data, one is assessing the effectiveness of an unknown program (i.e., a black box). To open this black box, we advocate adding implementation fidelity into the assessment cycle (see Figure 1).

Researchers have proposed key components of implementation fidelity assessment (e.g., Dane & Schneider, 1998; Hagermoser Sanetti & Kratochwill, 2009a; Hulleman & Cordrary, 2009; Mihalic, 2002; O'Donnell, 2008). However, none of these researchers focused specifically on assessing programming in student affairs. After reviewing and integrating the literature, we outlined five implementation fidelity components, with a specific focus on aligning these components with student affairs programming: program differentiation, adherence, quality, exposure, and responsiveness. Each component is defined in Figure 2, along with a means of assessing it. An understanding of implementation fidelity and its place in the assessment cycle

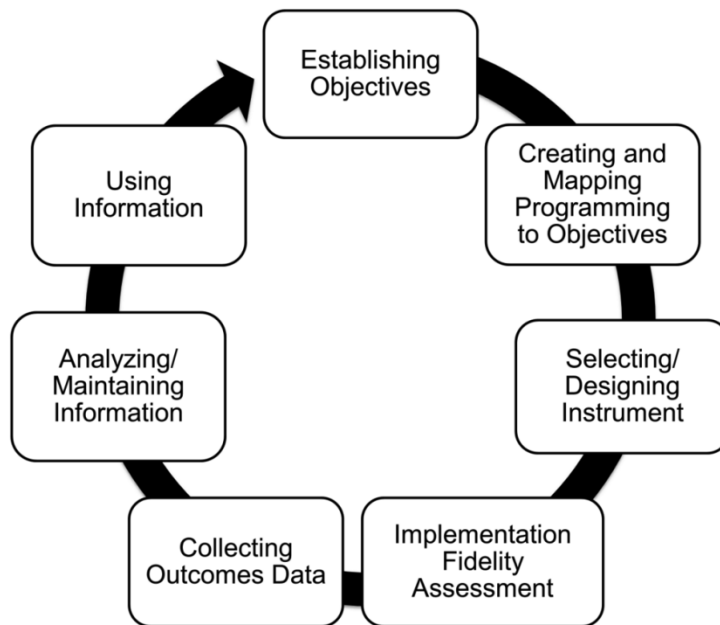


Figure 1. Outcomes assessment with implementation fidelity assessment included.

is best facilitated by a case study of implementation fidelity assessment. We therefore offer an example of implementation fidelity assessment from a large, multi-faceted student affairs program on our campus.

Gathering and Using Implementation Fidelity Data: Transfer Orientation Example

Transfer Student Orientation (TSO) is a one-day program that occurs the summer prior to the start of fall classes designed to help transfer students adjust to the campus community. Approximately 650 incoming transfer students attend one of four identical days of TSO programming. TSO programming was intentionally created to meet three objectives: increase academic requirements knowledge (ARK), increase resource knowledge (RK), and increase social acclimation (SA). Throughout the day, students attend programming aligned with these objectives. It is important to note that given the wide scope of TSO, many programming aspects are necessarily implemented by staff outside of the Orientation Office.

Outcomes Assessment Process

The three TSO objectives have outcome measures mapped to them. Data from the three measures are collected before and after TSO. A matched pre- to posttest design is used to assess growth for each objective. In summer 2011, 441 transfer students provided responses to all items on the pretest and posttest.

Although valuable information was obtained through the outcomes assessment process (i.e., which objectives were or were not met), informed program changes could not be made using only the outcomes assessment data. For instance, it was unclear why students were meeting the ARK objective better than the RK objective. Were the planned program features associated with the RK objective administered with high quality for the intended duration, implying this programming simply did not “work”? Or were we observing these findings because the program was implemented with low fidelity? Given that the administered program was a black box, we could not draw many conclusions about the effectiveness of the planned program. However, incorporating implementation fidelity into the assessment cycle enabled stakeholders to make programmatic decisions that could not be made with outcomes data alone.

Three possible reasons why implementation fidelity is not assessed center on untested assumptions, lack of understanding of implementation fidelity, and lack of guidance on the practice of collecting and using implementation fidelity data.

Implementation Fidelity Assessment Process

Implementation fidelity can be easily assessed by creating and completing a fidelity checklist (Swain, Finney, & Gerstner, 2013). Thus, an implementation fidelity checklist was developed for TSO to assess the five components of fidelity outlined in Figure 2. The checklist mapped a column of program objectives to a column of program features (i.e., program differentiation). Next to the column of program features, the planned duration of the feature was listed along with a space to record the actual duration. The next column was used to record adherence for each program feature (“yes” or “no”). The final column included a quality scale (1 = *Low* to 5 = *High*) so each implemented program feature could be rated for quality.

Data were collected using this checklist in two ways. First, three university staff affiliated with the program posed as students and audited TSO. During the day-long program, they recorded their ratings on this checklist. Second, two implementers of the various program features rated their own adherence, duration, and quality. In addition, we collected data from students regarding their responsiveness. Specifically, we added a question regarding responsiveness (*How attentive were you throughout the day?*) on the posttest, which also included outcome measures.

Without implementation fidelity assessment, nothing is known about what occurred during the program, only what was planned—which could be radically different from the actual implemented program.

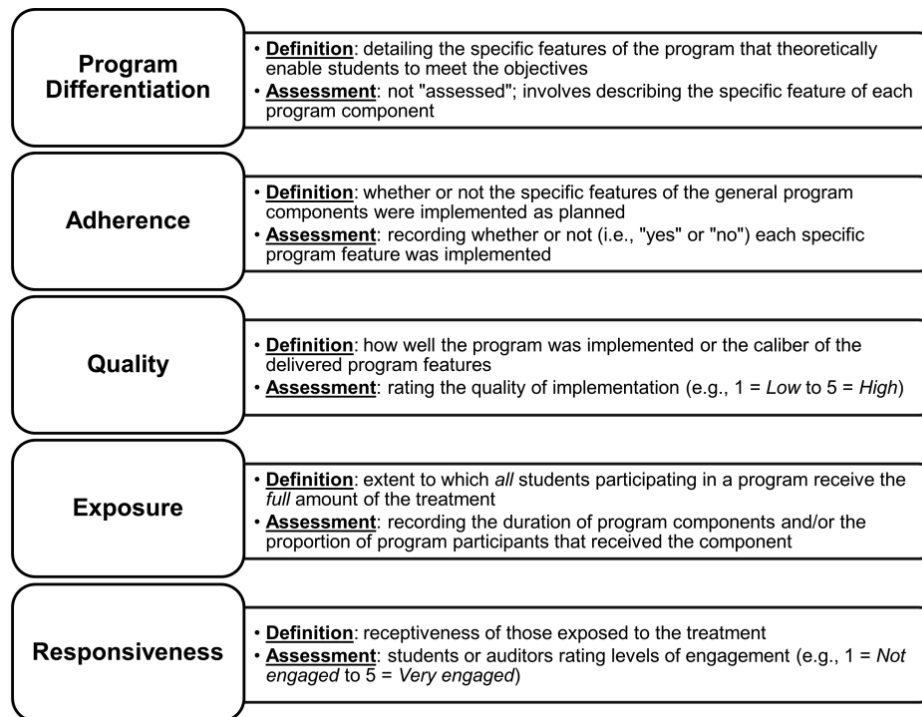


Figure 2. Implementation fidelity components: Definitions and assessment.

Interpreting Implementation Fidelity Data

In order to facilitate practitioners gathering and using implementation fidelity data, we expound on the definition and measurement of the five components of implementation fidelity within the context of TSO implementation fidelity data.

Program differentiation. The first component of implementation fidelity, *program differentiation*, involves detailing specific features of the program that theoretically enable students to meet program objectives (Dane & Schneider, 1998; Mihalic, 2002; Sheridan, Swanger-Gagné, Welch, Kwon, & Garbacz, 2009; Swain et al., 2013). As noted above, the TSO programming developed intentionally to enable students to meet each objective was specified in the fidelity checklist. For example, the program component of University Welcome, mapped to the SA objective, was broken down by stakeholders into the specific features intended to enable students to meet this objective (e.g., speech by the university president,

icebreakers in small groups). The act of program differentiation offered the stakeholders an opportunity to articulate their understanding of the link between the program outcomes and the program itself. That is, clarification of and commitment to the program objectives and programming was greatly facilitated by this differentiation process.

As noted in Figure 2, this component of implementation fidelity (unlike others) is not “assessed”; however, it is the *most* fundamental aspect of fidelity assessment. That is, program differentiation defines the program in the most specific way possible, which enables one to assess whether those features actually occurred (i.e., “adherence”) and evaluate their quality (i.e., “quality”). If specific program features cannot be discerned, fidelity assessment is impossible.

Adherence. The second component of implementation fidelity is *adherence*, which addresses whether or not specific program features were implemented (Dane & Schneider, 1998; Swain et al., 2013). In the education literature, adherence is often labeled “opportunity to learn” (e.g., Boscardin et al., 2005; Gee, 2003). Practitioners need to determine whether students had the opportunity to acquire skills and knowledge needed to meet the stated outcomes. Although Suskie (2009), in her book on higher education assessment, stressed the importance of presenting opportunities to learn, she never suggested one should evaluate if those opportunities were provided as planned (i.e., assessment of implementation fidelity was not a component of her outcomes assessment process).

There are four common methods of assessing adherence: auditors of the “live” program, videotapes of the program that are later examined, program implementers, and/or an evaluation of presentation materials.

As noted above, adherence can be easily assessed using a checklist (Cochrane & Laux, 2008; Swain et al., 2013). There are four common methods of assessing adherence: auditors of the “live” program, videotapes of the program that are later examined, program implementers, and/or an evaluation of presentation materials. The first, and the most objective and valid, method to assess adherence is through the use of auditors of the live program (Cochrane & Laux, 2008). This method was employed during TSO. Auditors attended programming (undercover) as participants and indicated whether specific program features were implemented as planned (i.e., recorded opportunity to learn as “yes” or “no” on the checklist). This method allowed auditors to experience the program as “students.” Readers should realize this approach could be resource-heavy, especially for long programs. The second method involves videotaping the program and having someone rate adherence by watching the videotape. This method may facilitate using a greater number of raters; however, the videotape may not allow an authentic representation of the actual program. Also, the presence of a camera could change the program’s dynamic. Another useful method of assessing opportunity to learn is by asking program implementers to indicate their adherence to specific program features (Breitenstein et al., 2010). This approach was also employed during TSO. Gathering adherence data from implementers and auditors provides inter-rater reliability data (i.e., consistency in ratings from auditors and implementers). The assessment of inter-rater reliability is important, as some research has found that self-ratings indicate higher fidelity when compared to ratings from independent observers (Hagermoser Sanetti & Kratochwill, 2009a; O’Donnell, 2008), whereas other research has found accurate ratings from implementers (Hagermoser Sanetti & Kratochwill, 2009b). If it can be shown that implementers and auditors provide the same implementation fidelity data, then auditors would not be needed. Finally, if a program involves the presentation and discussion of informational materials (e.g., handouts), an examination of these materials can serve as a crude measure of adherence (Lane et al., 2004). Although not an ideal approach, this may be the only possible method for assessing adherence when a program audit or videotaping is not possible (e.g., private setting, lack of time) or if implementers will not participate in assessing fidelity (Cochrane & Laux, 2008).

With respect to TSO, the auditor and implementer adherence ratings were identical. Importantly, both auditors and implementers noted specific program features that were not executed. This finding was extremely valuable, as it indicated that implementers were willing to report their lack of coverage of program features and did so accurately. Moreover, the implementers indicated that simply engaging in rating their adherence to specific features served as an additional reminder of the content intended to be covered in the program. As noted, programs can drift unintentionally from the intended features. Requiring program implementers to review a list of program features and then indicate whether they implemented those features communicates the importance of executing the program as planned and can

protect against program drift. In addition, the process of gathering implementer adherence ratings may reduce time needed to retrain implementers (Durlak & DuPre, 2008).

Quality. The third component of implementation fidelity, *quality*, assesses the caliber of delivered program features (Dane & Schneider, 1998; Mihalic, 2002; Swain et al., 2013). With respect to TSO and higher education programming more generally, quality is an essential component of implementation fidelity. Implementers could deliver all specific program features (i.e., high adherence), yet low quality prevents the planned program from being administered fully. Anyone who has attended a presentation where information was presented quickly or unclearly can attest to the importance of assessing the quality of implementation. Although Schuh and Upercraft (2001) and Suskie (2009) mention the importance of developing high-quality student-centered programs and note that quality of presentation skills (i.e., “presenter effectiveness”) could impact the functioning of the program, they focus only on this one narrow component of quality. Moreover, they fail to discuss how to measure presenter effectiveness or how to couple these data with the outcomes assessment to inform program changes. We believe the assessment of quality should include the concept of presenter effectiveness addressed by Schuh and Upercraft and Suskie when appropriate, but it should be defined widely enough to accommodate programs without a presenter. Whereas the concept of presenter effectiveness would be irrelevant for a student affairs program targeted at weight loss, the quality of the implemented program features (e.g., exercise regime) could be rated (e.g., exercises completed too quickly, without much effort, with poor form), thus providing useful information regarding program implementation.

Similar to adherence, quality can be rated by auditors and/or implementers (Swain et al., 2013). In the case of TSO, every specific program feature that received a “yes” for adherence was rated for quality (e.g., 1 = *Low* to 5 = *High*) by the auditors and implementers. A specific feature received a low quality rating if the feature was addressed, but not well. For example, TSO has an icebreaker activity intended to increase students’ sense of belonging to campus. If the icebreaker activity occurred (i.e., adherence) but group facilitators did not present the activity in an engaging manner, then the students received the program feature but with poor quality.

With respect to TSO, the auditors and implementers were in agreement for most of the quality ratings. Although many features were adhered to, there was a range in quality. Fortunately, many features garnered high quality ratings; however, there were also some low ratings. As discussed below, these fidelity data helped to explain some unfavorable outcome results.

Exposure. The fourth component of implementation fidelity is *exposure*, which assesses the extent to which all students participating in a program receive the full treatment (Carroll et al., 2007; Dane & Schneider, 1998; Swain et al., 2013). In addition to detailing each program feature, program differentiation specified the planned duration of the program components. With respect to student affairs programming, practitioners intend for students to receive a “full dose” of each program component, but that does not always occur. If the planned 50-minute program component receives only 20 minutes of time, students are not being exposed to the “full treatment.” Thus, students may not have the opportunity to learn to the extent intended by the program. With respect to TSO, exposure was assessed by auditors recording the actual duration of each program component. All components endured for the planned amount of time, providing confirmation that students had the opportunity to be exposed to the intended, “full dose” of programming.

In addition to assessing the duration of programming, one can also assess whether everyone was exposed to each aspect of the program. We would not expect positive outcomes assessment results if half of the participants “skipped” the programming aligned with the objective. Thus, even if the program was presented for the intended duration with high quality, the programming may appear ineffective if data from program attendees and those who skipped the program were analyzed together.

With respect to TSO, plans have already been established to further assess exposure. In the future we will ask students whether they attended various optional aspects of TSO. These attendance data will allow the outcomes data to be analyzed separately for those who did and did not attend optional programs. This type of analysis is important because if

Requiring program implementers to review a list of program features and then indicate whether they implemented those features communicates the importance of executing the program as planned and can protect against program drift.

attending optional aspects of programming has a strong impact on program outcomes, it may be beneficial to make that programming mandatory in future years. That is, exposure data can help highlight which combination of programming components are most effective in meeting outcomes, which can assist in the allocation of resources when implementing the program in the future.

Responsiveness. The final component of implementation fidelity is *responsiveness*, which addresses the receptiveness of those exposed to the treatment (Dane & Schneider, 1998; Swain et al., 2013). If students are not engaged with the TSO program, it does not matter whether the implementers deliver all the planned program features in a high quality manner for the intended duration. Students will not be impacted by a high quality program if they are disengaged. Thus, assessing responsiveness, rather than making the assumption that the program is being offered to a fully captive audience, can help illuminate why well-implemented programs may be associated with poor outcomes assessment results.

Thus, even if the program was presented for the intended duration with high quality, the programming may appear ineffective if data from program attendees and those who skipped the program were analyzed together.

Responsiveness can be assessed by asking students to self-report their level of attentiveness throughout the program. Another, more distal, measure of responsiveness would entail an auditor or implementer rating the responsiveness of the audience. Both assessments have their flaws. Students' self-reports of their responsiveness may be influenced by social desirability. Alternatively, an auditor may mistakenly perceive attendees as inattentive because they are looking down when in reality, they are taking notes. However, both measures of responsiveness can supply information otherwise lacking from the assessment process. These results can also be used to analyze the outcomes data separately for those who were or were not responsive.

With respect to TSO, as noted above, students indicated whether they were attentive throughout the day (1 = *Not at all*, 2 = *Somewhat*, 3 = *Very*). Fortunately, only 1.1% of students responded "not at all." A fairly comparable number of students responded either "somewhat" or "very" to the item. We tested for a possible differential effect of the programming across "responsiveness" groups and found no differential change in the three outcomes over time. That is, responsiveness did not moderate the change in outcomes assessment scores (i.e., there was no significant interaction).

Drawing Conclusions about Program Effectiveness by Combining Fidelity and Outcomes Assessment

The implementation fidelity results were used in numerous ways to strengthen the validity of inferences about the effectiveness of TSO. It is important to note that the TSO program director and program implementers were equal partners when interpreting implementation fidelity and outcomes assessment results and when using these results to strengthen TSO for subsequent years. This equal partnership, which had not been present in the past, was facilitated by the implementers' participation in fidelity assessment.

The implementation fidelity results coupled with the assessment results uncovered findings that neither set of results could have yielded independently. When fidelity and outcomes assessment results are combined, there are four possible scenarios that could occur (McIntyre, Gresham, DiGennaro, & Reed, 2007; Swain et al., 2013). All four scenarios presented in Figure 3 were evidenced in the assessment of TSO and we provide examples of each to model interpretation of such findings. Importantly, the combination of implementation fidelity and outcomes assessment results informed modifications to the programming components and allocation of resources.

High fidelity and favorable outcomes. Some results reflected high levels of implementation fidelity coupled with favorable outcomes assessment results (scenario 2 in Figure 3). For example, the SA objective has numerous specific program features, such as the University Welcome and Peer Discussion Groups, and the auditors observed and reported that all specific features were presented and in a fairly high-quality manner. Moreover, outcomes assessment results revealed a significant increase for the SA outcome measure from pre- to posttest. Thus, the fidelity results suggested the increase on the outcome measure might be a function of TSO programming.

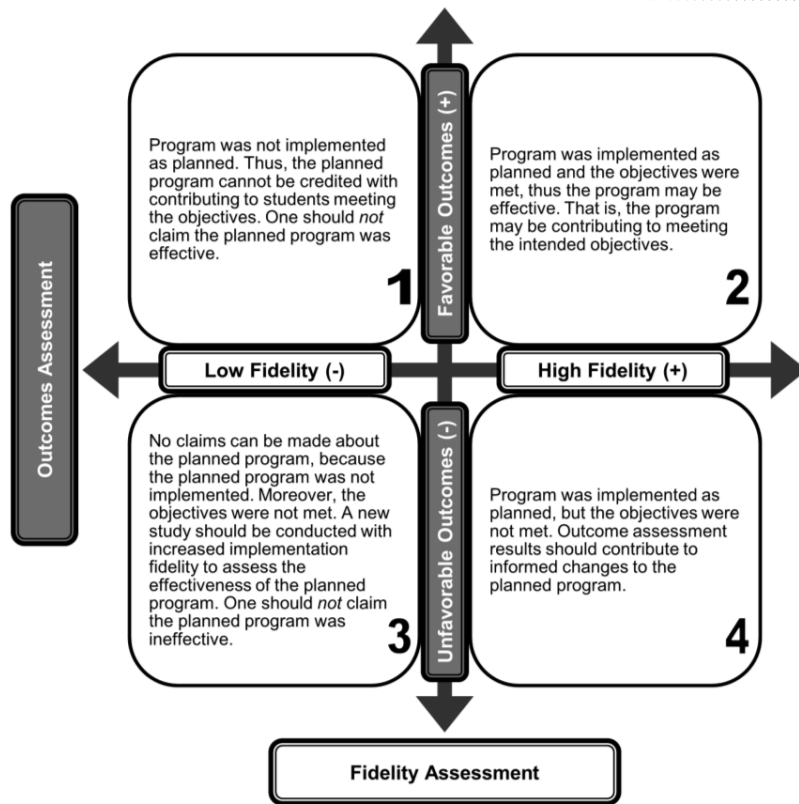


Figure 3. Four scenarios resulting from pairing implementation fidelity assessment results with outcomes assessment results. This figure is general with respect to research design; it does not assume a true experiment. Thus, positive outcomes assessment results do not imply program effectiveness; it simply reflects the objective was met. In quasi-experimental designs there could be several reasons other than program effectiveness that explain objectives being met.

Low fidelity and unfavorable outcomes. Some of the outcomes and fidelity assessment results aligned with scenario 3: the outcomes assessment results were poor and the fidelity assessment revealed the program had not been implemented as planned. Therefore, the obtained outcomes results were not reflective of the *planned* program. For instance, one specific program feature associated with the RK objective is explaining how and where to transfer credits. The fidelity assessment results indicated this specific program feature, although adhered to, had been delivered with extremely low quality. Not surprisingly, the outcomes assessment results indicated students did not understand the process of evaluating whether a course could transfer from another institution. By pairing implementation fidelity and outcomes assessment results, stakeholders discovered that the poor performance might be due to poor program implementation, which can be easily remedied before the next transfer orientation.

High fidelity and unfavorable outcomes. Some results reflected fairly high levels of implementation fidelity coupled with poor outcomes assessment results (scenario 4). One specific program feature associated with the RK objective involved explaining how and where one pays tuition. Fidelity assessment results revealed this information was presented in a high-quality manner; however, students performed poorly on the outcome measure. Because the fidelity assessment results indicated this poor performance was not due to poor program implementation, additional or different types of programming may need to be developed to help students meet this objective. In short, it appears the programming in place is not working, thus resources should be allocated to replace or modify the existing programming.

Low fidelity and favorable outcomes. Finally, there were instances of low fidelity paired with good outcomes assessment results (scenario 1). Outcomes assessment results revealed students increased significantly from pre- to posttest on measures associated with ARK. From the perspective of a standard outcomes assessment cycle, one would conclude that

Exposure data can help highlight which combination of programming components are most effective in meeting outcomes, which can assist in the allocation of resources when implementing the program in the future.

programming mapped to this objective may be effective in teaching students this information. However, fidelity data indicated programming mapped to this objective was *not* implemented. Thus, students were evidently learning this information elsewhere. Fortunately, during the fidelity assessment the auditors noted that the information was mentioned in another (albeit unplanned) programmatic component. Absent fidelity information, one would have wrongly attributed the favorable outcomes assessment results to the planned programming. Given the success in presenting this information in the unplanned, alternative programming component, the program director and implementers decided to adjust the program to reflect this change (i.e., no longer expend resources on the original programming but instead on the alternative programming).

Implications for Practice and Suggestions

In order to equip decision-makers with the necessary data to make informed decisions, implementation fidelity data must be presented and used to interpret outcomes assessment results.

The lack of implementation fidelity data challenges valid inferences and decision making regarding program impact (i.e., internal validity), as a lack of student learning and development could be due to poor implementation for which no data are available to aid administrators making program-related decisions (Ball & Christ, 2012; Durlak & DuPre, 2008). In turn, it is extremely difficult (if not impossible) to make informed, data-based decisions about resource allocation. Moreover, the lack of implementation fidelity data compromises conclusions concerning the replication and generalization of program effects (i.e., external validity; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2013). Finally, lack of fidelity data makes evaluation of the properties of outcome measures ambiguous. Understanding the properties of outcomes measures is critical, as a measure may appear more difficult if students did not have the OTL (i.e., low implementation fidelity). That is, students will have trouble answering items correctly if they were not taught the material (i.e., did not have the OTL), thus making the measure appear more difficult than it would be if the students had the OTL (as intended). This failure to assess fidelity could result in practitioners discarding a high-quality measure that would have functioned properly (i.e., appeared adequate and not overly challenging) if students had the OTL (Coleman, Kaliski, & Huff, 2012; Huff & Ferrara, 2010; Polikoff, 2010). Thus, in order to equip decision-makers with the necessary data to make informed decisions, implementation fidelity data must be presented and used to interpret outcomes assessment results.

Allowing practitioners to allocate the necessary time and resources (both financial and human) to fidelity assessment is critical. Of course, this may result in assessing fewer programs each year, but that is weighed against having more accurate assessment of these programs.

Given the importance of implementation fidelity data, how do we engage higher education practitioners (e.g., faculty, administrators, staff) in the practice of gathering these data? We offer three suggestions to increase the practice of evaluating implementation fidelity and overcoming barriers. First, practitioners must be educated about the concept of implementation fidelity and its importance for evaluating program effectiveness. Research has shown that those trained in implementation fidelity are more likely to perceive it as important and engage in its measurement (Cochrane & Laux, 2008). Articles appearing in higher education, student affairs, or assessment journals that explicate the concept and model use of these data could increase awareness and support practitioners engaging in this practice for the first time. Participation in listservs, conferences, or other activities that focus on making empirically-based decisions regarding program effectiveness is also advised (Hagermoser Sanetti & Kratochwill, 2009a).

Second, resource barriers must be minimized. That is, human and financial resources have been found to be a barrier to implementation fidelity assessment (e.g., Cochrane & Laux, 2008; Hagermoser Sanetti & DiGennaro Reed, 2012). Thus, allowing practitioners to allocate the necessary time and resources (both financial and human) to fidelity assessment is critical. Of course, this may result in assessing fewer programs each year, but that is weighed against having more accurate assessment of these programs.

The final suggestion addresses the barrier associated with a lack of requirement to gather implementation fidelity data (Hagermoser Sanetti & DiGennaro Reed, 2012). Although we agree with Shutt et al. (2012) that program assessment is part of best practice and thus should be engaged in *without* mandates, requiring the gathering and use of implementation fidelity data would spur engagement in this practice. Research has shown that lack of perceived value of implementation fidelity data by administrators or the system serves as a barrier to fidelity assessment (Cochrane & Laux, 2008). Thus, we urge administrators to request these

data and the explication of how these data were used to provide a more complete and accurate picture of program effectiveness.

Moreover, implementation data are just as critical for research on higher education programs as they are for internal program effectiveness studies, which aligns with Hagermoser Sanetti and DiGennaro Reed's (2012) call for the integration of implementation fidelity and outcomes data in intervention research: "Treatment integrity and student outcome data are not only important in school and clinical settings but they are also essential to drawing valid conclusions in treatment outcome research" (p. 196). Thus, these authors call for journal editors and reviewers to require implementation fidelity data, which further addresses the barrier associated with lack of reporting requirements.

Finally, one may ask, "Is implementation fidelity a part of the outcomes assessment process or part of the program development process?" We present the following two thoughts in response to this question. First, program development and outcomes assessment should never be two separate processes. Program development has always been considered a key *part* of the outcomes assessment process (e.g., ACPA, 2006; Bresciani et al., 2009; Suskie, 2009). More specifically, when engaging in the program development process, the goal is to create programming that aligns with the stated student learning and development outcomes. To remove program development from the assessment cycle (Step 2 in Figure 1) would be nonsensical—the two are necessarily integrated. In fact, it is the clear link between the objectives and the programming that is critical to any assessment process or quality programming.

Second, and given our first point, implementation fidelity is part of the outcomes assessment process. In fact, implementation fidelity strengthens the outcomes assessment cycle. For example, program differentiation essentially makes the mapping of programming to objectives more overt, thus strengthening and better integrating the first (i.e., establishing objectives) and second (i.e., creating and mapping programming to objectives) stages of the assessment cycle. During the final stage of the assessment cycle, "Use of Information," fidelity data make diagnostics, program changes, and resource allocation much easier for stakeholders. In sum, we view the outcomes assessment cycle as including the key components of program development (e.g., ACPA, 2006; Bresciani et al., 2009; Suskie, 2009) and implementation fidelity.

Conclusions

When employing the standard outcomes assessment cycle, we have observed two common (although not necessarily appropriate) conclusions are often made following unfavorable performance on an outcome measure: the measure is not functioning properly and thus cannot reflect program effectiveness; or the program needs revision or termination. If the outcome measure was meticulously selected/designed for the program and has adequate psychometric properties, poor measurement would not seem to be a likely cause of poor performance. Moreover, concluding that program revision/termination is necessary would be premature without any information as to whether or not the planned program was truly the implemented program.

Obtaining implementation fidelity data ensures the correct program is being evaluated rather than one distorted, possibly substantially, due to implementers drifting from the planned program. Moreover, given the more complete understanding of the program's functioning afforded by implementation fidelity data, more accurate changes can be made to the program (McIntyre et al., 2007). Although this article focused on a one-day student affairs program, we applaud practitioners who conduct implementation fidelity assessment for programs of longer durations and more complex outcomes (e.g., K-12 education [Boscardin et al., 2005], school psychology [Hagermoser Sanetti & Kratochwill, 2009b]).

As Terenzini and Upercraft (1996) noted, "While assessing the purported outcomes of our efforts with students is probably the most important assessment we do, it is seldom done, rarely done well, and when it is done, the results are seldom used effectively" (p. 217). Implementation fidelity assessment can help address this problem. In sum, integrating implementation fidelity and outcomes assessment can assist us all in making more informed programmatic decisions, increasing communication between program directors and implementers of programs, and ultimately meeting the needs of students by offering empirically-supported, effective programming.

Obtaining implementation fidelity data ensures the correct program is being evaluated rather than one distorted, possibly substantially, due to implementers drifting from the planned program.

References

- Aiken-Wisniewski, S., Campbell, S., Nutt, C., Robbins, R., Kirk-Kuwaye, M., & Higa, L. (2010). *Guide to assessment in academic advising* (2nd ed.). (NACADA Monograph No. 23). Manhattan, KS: The National Academic Advising Association.
- American College Personnel Association. (2006). *ASK standards: Assessment skills and knowledge content standards for student affairs practitioners and scholars*. Washington, DC: Author.
- Ball, C. R., & Christ, T. J. (2012). Supporting valid decision making: Uses and misuses of assessment data within the context of RTI. *Psychology in the Schools, 49*, 231–244.
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *Educational Forum, 40*, 344–70.
- Boscardin, C. K., Aguirre-Munoz, Z., Stoker, G., Kim, J., Kim, M., & Lee, J. (2005). Relationship between opportunity to learn and student performance on English and Algebra assessments. *Educational Assessment, 10*, 307–332.
- Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research, 59*, 158–165.
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2009). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs*. Sterling, VA: Stylus.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 1–9.
- Century, J., Cassata, A., Rudnick, M., & Freeman, C. (2012). Measuring enactment of innovations and the factors that affect implementation and sustainability: Moving toward common language and shared conceptual understanding. *Journal of Behavioral Health Services & Research, 39*, 1–19.
- Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools, 45*, 499–507.
- Coleman, C., Kaliski, K., & Huff, K. (2012). *Examining the relationship between opportunity to learn and difficulty statistics on exams implementing evidence-centered design*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.
- 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.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.
- Erwin, T. D. (1991). *Assessing student learning and development*. San Francisco, CA: Jossey-Bass.
- Garvey, C., Julion, W., Fogg, L., Kartovil, A., & Gross, D. (2006). Measuring participation in a prevention trial with parents of young children. *Research in Nursing & Health, 29*, 212–222.
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education, 10*, 27–46.
- Hagermoser Sanetti, L. M., & DiGennaro Reed, F. D. (2012). Barriers to implementing treatment integrity procedures in school psychology research: Survey of treatment outcome researchers. *Assessment for Effective Intervention, 37*, 195–202.
- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4*, 95–114.

- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009a). Toward developing a science of treatment integrity: Introduction to the Special Series. *School Psychology Review*, 38, 445–459.
- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009b). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning protocol. *School Psychology Quarterly*, 24, 24–35.
- Huff, K., & Ferrara, S. (2010, June). *Frameworks for considering item response demands and item difficulty*. Paper presented at the annual meeting of the National Conference on Student Assessment, Detroit, MI.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2, 88–110.
- Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential—but often forgotten—component of school-based interventions. *Preventing School Failure*, 48(3), 36–43.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991-2005. *Journal of Applied Behavior Analysis*, 40, 659–672.
- Mihalic, S. (2002, April). *The importance of implementation fidelity*. Boulder, CO: Center for the Study and Prevention of Violence. Retrieved from www.colorado.edu/cspv/blueprints/Fidelity.pdf
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315–340.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services and Research*, 39, 374–396.
- 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, 33–84.
- Patton, M. Q. (1997). *Utilization focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Polikoff, M. S. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice*, 29(4), 3–14.
- Posavac, E. J., & Carey, R. G. (1997). *Program evaluation: Methods and case studies* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Schuh, J. H. (2009). *Assessment methods for student affairs*. San Francisco, CA: Jossey-Bass.
- Schuh, J. H., & Upcraft, M. L. (2001). *Assessment practice in student affairs: An applications manual*. San Francisco, CA: Jossey-Bass.
- Sheridan, S. M., Swanger-Gagné, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review*, 38, 476–495.
- Shutt, M. D., Garrett, J. M., Lynch, J. W., & Dean, L. A. (2012). An assessment model as best practice in student affairs. *Journal of Student Affairs Research and Practice*, 49, 65–82.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Swain, M. S., Finney, S. J., & Gerstner, J. J. (2013). A practical approach to assessing implementation fidelity. *Assessment Update*, 25(1), 5–7, 13.

- Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, 47(1), 3-13.
- Terenzini, P. T., & Upcraft, M. L. (1996). Assessing program and service outcomes. In M. L. Upcraft & J. H. Schuh (Eds.), *Assessment in student affairs: A guide for practitioners* (pp. 217–239). San Francisco, CA: Jossey-Bass.
- Upcraft, M. L., & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioners*. San Francisco, CA: Jossey-Bass.
- U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education. A report of the commission appointed by secretary of education Margaret Spellings*. Washington, DC.: Author.
- Weiss, C. H. (1998). *Evaluation* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.