Research Article

Research on Social Work Practice
23(6) 613-623

DOI: 10.1177/1049731513491150
rsw.sagepub.com
SAGE

Fidelity in After-School Program Intervention Research: A Systematic Review

Brandy R. Maynard[1,2], Kristen E. Peters[1], Michael G. Vaughn[1], and Christine M. Sarteschi[3]

Abstract

Objective: This study examined whether and to what extent researchers addressed intervention fidelity in research of after-school programs serving at-risk students. **Method:** Systematic review procedures were used to search, retrieve, select, and analyze studies for this review. Fifty-five intervention studies were assessed on the following components of intervention fidelity: strategies to enhance fidelity, measurement of fidelity, and use of fidelity data in data analysis and interpretation. **Results:** Of the 55 studies examined, only 55% reported well-defined intervention procedures, 42% used an intervention manual, 33% provided training on the intervention, 24% provided supervision for the implementers, 29% measured fidelity, only 4% used fidelity data in their analysis, and no studies reported the reliability of fidelity measures. **Conclusion:** Findings indicate an overall lack of attention to and reporting of intervention fidelity in after-school intervention studies. Implications for practice, policy, and research are discussed.

Keywords

after-school programs, intervention fidelity

After-school programs (ASPs) emerged to fill the increased discretionary, "idle" time for youth, resulting from the ending of child labor force participation and the passing of compulsory education laws (Mahoney, Parente, & Zigler, 2009). Over the past century, ASPs have proliferated and continue to be shaped and influenced by sociopolitical forces. Changes in family and labor force participation, growing concerns over the neighborhood context and safety of youth, research documenting peaks in juvenile crime during after-school hours, the link between high-risk behaviors and lack of supervision, and myriad other social and political influences have contributed to the growth of ASPs (Apsler, 2009; Mahoney et al., 2009; Zief, Lauver, & Maynard, 2006). Over the past two decades, the popularity, demand, and funding for ASPs have continued to rise, resulting in a marked increase in the number of ASPs and the number of students attending ASPs. Today, approximately 50,000 public elementary schools and numerous more middle and high schools offer some type of ASP (Parsad & Lewis, 2009). ASPs are clearly well established within the public, nonprofit, and private sectors across the United States.

ASPs have thrived, at least in part, because they are viewed as important and beneficial to students, families, schools, and communities. The presumed benefits of ASPs include increasing youth safety and curbing juvenile crime and other high-risk behaviors by providing youth with adult care and supervision after school; expanding learning opportunities, improving academic achievement, and closing the achievement gap by providing academic enrichment opportunities to at-risk youth; and increasing worker productivity for parents and businesses by providing reliable child care after school (Hollister, 2003; Zief et al., 2006). Moreover, ASPs receive a high level of support from various stakeholders; school superintendents, principals, school board members, and parents believe that ASPs are necessary or important in their communities (Afterschool Alliance, 2012; Belden Russonello & Stewart, 2003; National Association of Elementary School Principals, 2001). These perceived and promising benefits have helped fuel research centers and advocacy groups' interest in ASPs, as well as funding from private and public entities across local, state, and federal levels (Mahoney et al., 2009). The federal government contributes significant resources to ASPs; between 1998 and 2012, federal funding for the 21st-Century Community Learning Center ASPs increased from $40 million to $1.152 billion. This increase in funding is due primarily to the No Child Left Behind Act of 2001, which sought to close the achievement gap through the creation of "academic enrichment opportunities during non-

[1] School of Social Work, Saint Louis University, St. Louis, MO, USA
[2] The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, Austin, TX, USA
[3] School of Social Work, Chatham University, Pittsburgh, PA, USA

Corresponding Author:
Brandy R. Maynard, School of Social Work, Saint Louis University, Tegeler Hall, 3550 Lindell Boulevard, St. Louis, MO 63103, USA.
Email: bmaynard@austin.utexas.edu

school hours for children, particularly students who attend high-poverty and low-performing schools" (U.S. Department of Education, 2011).

Public recognition for the *potential* of ASPs to improve behavioral and academic outcomes has resulted in the influx of funding for these programs; however, "the rapid growth of after-school programming resulted from lobbying and grass roots efforts and was not based on strong empirical findings" (Apsler, 2009, p. 2). A decade after No Child Left Behind went into effect, many are left questioning how effectively ASPs fulfill their goals and achieve the perceived benefits for students, families, schools, and society. Intervention research and reviews assessing the impact of ASPs have resulted in an ambiguous picture of the effects of these programs.

Although several reviews and meta-analyses have examined the outcomes of ASPs (see Apsler, 2009; Durlak, Weissberg, & Pachan, 2010; Fashola, 1998; Hollister, 2003; Lauer et al., 2006; Roth, Malone, & Brooks-Gunn, 2010; Scott-Little, Hamann, & Jurs, 2002; Zief et al., 2006), ASP intervention study reviews have not specifically examined intervention fidelity (i.e., whether and how closely the ASP intervention was implemented as intended). Establishing intervention fidelity is critically important to interpreting the effects, or lack thereof, of interventions (O'Donnell, 2008; Perepletchikova & Kazdin, 2005; Summerfelt, 2003). Moreover, fidelity data are important in interpreting negative or ambiguous findings and in determining "whether unsuccessful outcomes are due to ineffective interventions or due to a failure to implement the intervention as intended" (Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011, p. 1). Fidelity data also provide important information to guide adjustments and improvements to research and intervention design, reveal important information related to the feasibility of an intervention, and guide future investigation (Dusenbury, Brannigan, Falco, & Hansen, 2003). Examining the extent to which investigators have attended to fidelity in ASP intervention research is a critical step in understanding the conflicting findings of prior reviews and intervention studies and in providing important insights that can guide future research and development of ASP interventions.

Intervention Fidelity

Intervention fidelity has been increasingly emphasized over the past 30 years and is now viewed as an essential component in intervention research across disciplines (Gearing et al., 2011). Although uniformity regarding the construct, definition, and labels of intervention fidelity is lacking (Gearing et al., 2011), fidelity in intervention research is generally described as comprising "strategies that monitor and enhance the accuracy and consistency of an intervention to ensure it is implemented as planned and that each component is delivered in a comparable manner to all study participants over time" (Smith, Daunic, & Taylor, 2007, p. 121). Elements of intervention fidelity include design and operationalization of the intervention (i.e., well-defined set of procedures, written intervention manual), implementer training, supervision, and monitoring of intervention delivery, verification of adherence to the intervention protocol (i.e., measuring fidelity), and use of fidelity data in analysis (Gearing et al., 2011; Moncher & Prinz, 1991). Intervention fidelity has significant implications for the interpretation, use, and replication of intervention research.

The "assessment of intervention fidelity in intervention studies helps researchers understand, as unequivocally as possible, how the intervention relates to child outcomes" (Smith et al., 2007, p. 130). If an intervention is not implemented as intended, internal validity is compromised (Chen & Rossi, 1983). Traditionally, researchers have made assumptions that interventions are implemented as designed; however, empirical data show that this assumption is untenable (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001). Having a detailed description of the intervention or even an intervention manual is not enough to ensure that the intervention is implemented properly in the field (Shadish, Cook, & Campbell, 2002). Without verification that the independent variable was implemented as intended, it cannot be determined whether outcomes are attributable to the intervention, influences of unknown variables, or the failure to implement the intervention as designed (Dumas et al., 2001; Waltz, Addis, Koerner, & Jacobson, 1993).

Intervention fidelity also has implications for external validity, statistical power, and magnitude of effect (Durlak & DuPre, 2008; Moncher & Prinz, 1991; Resnick et al., 2005; Summerfelt, 2003). External validity relates to intervention replication and generalizability to other settings. For an intervention to be replicable and adoptable by clinicians, sufficient information about the intervention is required (Moncher & Prinz, 1991). This information is especially important for complex programs that comprise multiple components, such as many current ASPs. Without sufficient detail of the components and fidelity of interventions, replication and comparison across studies are compromised (Smith et al., 2007). Fidelity also can affect statistical power. For example, lack of standardization in how the intervention is delivered may inflate error variance and lead to an underpowered analysis and greater chance of a Type II error (Cook & Poole, 1982). A nonsignificant outcome may reflect a lack of statistical power, rather than an ineffective intervention. Moreover, interventions that are implemented with fidelity are associated with better outcomes (Durlak & DuPre, 2008; Hulleman & Cordray, 2009).

The monitoring and collecting of fidelity data is also imperative to intervention implementation. Fidelity monitoring can identify and correct in real-time practice drift and implementation errors and reinforce successful implementation (Kaye & Osteen, 2011). Fidelity data can also provide valuable information related to implementation challenges and dosage effects and can improve efficiency and reduce costs of intervention research (Moncher & Prinz, 1991; Resnick et al., 2005).

Despite greater awareness of and agreement on the importance of intervention fidelity, the practice of promoting, monitoring, and measuring intervention fidelity has been limited. Prior reviews of treatment fidelity in social work, education, and psychotherapy have consistently revealed a lack of attention to and reporting of fidelity (see Mooney, Epstein,

Reid, & Nelson, 2003; Naleppa & Cagle, 2010; Swanson et al., 2011; Weisz, Doss, & Hawley, 2005). Reviews of published social work intervention research revealed that only 15.3% of the studies collected fidelity data, and, in another review, fewer than 14% of the 128 studies attended to fidelity in some way (Naleppa & Cagle, 2010; Tucker & Blythe, 2008). Likewise, intervention research in education lacks attention to and measurement of intervention fidelity. A review of intervention studies for children at risk of emotional or behavioral disorders found that 43% of the studies lacked any reporting of fidelity measures and only 38% reported content and process fidelity (Hester, Baltodano, Gable, Tonelson, & Hendrickson, 2003). Although establishing intervention fidelity is now seen as a critical aspect of intervention research, limited attention to intervention fidelity is pervasive across published research in social work and education.

Because the ultimate purpose of conducting ASP intervention research is to improve the well-being and trajectories of youth, it is critical that outcomes and intervention components are clearly defined and measured and that the intervention can be replicated. In short, demonstrating intervention fidelity is critical to the evaluation, comparison, dissemination, and implementation of ASP interventions. It is unclear, however, whether fidelity has been given adequate attention in ASP intervention research to be able to draw valid conclusions and adequately disseminate and replicate ASP interventions.

Purpose of the Present Study

Given the popularity of ASPs over the past two decades, the wide variability in ASP models, and the growing body of intervention research resulting in ambiguous findings, it seems prudent to examine whether investigators have attended to fidelity in ASP intervention research. This examination is a critical step in gaining a better understanding of the conflicting findings and providing important insights that can guide the future investigation, interpretation, and replication of ASP intervention outcome research and intervention development. The following research questions guide this review: (1) What proportion of after-school intervention studies report key components of fidelity (i.e., strategies to enhance fidelity, measure fidelity, and use fidelity data)? (2) Does the presence of fidelity measurement differ by study or intervention characteristics?

Method

Systematic review procedures, following the Campbell Collaboration procedures and guidelines (see www.campbellcollaboration.org), were used for all aspects of the search, retrieval, selection, and coding of published and unpublished studies meeting study inclusion criteria.

Study Inclusion Criteria

Studies were included in this review if they examined the effects of an ASP on social, emotional, behavioral, or academic outcomes with at-risk primary or secondary students using a randomized or quasi-experimental research design. ASPs were defined as "an organized program offering one or more activities that: (a) occurred during at least part of the school year; (b) happened outside of normal school hours; and (c) was supervised by adults" (Durlak et al., 2010, p. 296). Interventions that involved solely mentoring or tutoring, operated solely during the summer, or occurred during school hours were excluded from this review. For the purposes of this review, we used a broad definition of at risk adapted from Lauer et al. (2006). At-risk students were defined as (1) performing below grade level or having low scores on academic achievement tests; or (2) attending a low-performing or Title I school; or (3) having characteristics associated with risk of lower academic achievement, such as low socioeconomic status, racial or ethnic minority background, single-parent family, limited English proficiency, or a victim of abuse or neglect; or (4) engaging in high-risk behavior, such as truancy, running away, substance use, or delinquency. Due to significant differences in educational systems around the world, this review was limited to studies conducted in the United States, Canada, United Kingdom, Ireland, and Australia. Only English language articles were included in the review.

Search, Retrieval, and Selection of Studies

Searches were conducted in March 2012. Several sources were used to identify eligible studies published between January 1980 and May 2012. Eight electronic databases (i.e., Social Work Abstracts, PsychINFO, ProQuest Dissertation and Theses, Academic Search Complete, Social Service Abstracts, Sociological Abstracts, ERIC, and Social Sciences Citation Index); online searches of relevant government agencies, research centers and professional association websites; and reference lists of prior reviews were searched for relevant studies. A librarian specializing in social work was consulted to determine the appropriate databases to search and key word search terms to use. Key word searches within each database included combinations of "evaluation," "treatment," "intervention," and "outcome" in conjunction with "after-school program" to narrow the search field to evaluations of ASPs.

Titles and abstracts of the studies found through the search procedures were screened for relevance. Studies that were obviously ineligible or irrelevant were screened out—for example, some studies did not involve the target population (e.g., they involved college students or adults), did not involve an intervention, or were theoretical in nature. If there were any question as to the appropriateness of the study at this stage, the full-text document was obtained and screened. Documents that were not obviously ineligible or irrelevant, based on the abstract review, were retrieved in full text and screened for eligibility using a screening instrument, which is available from the authors of this review.

The search yielded a total of 374 studies for screening, with 55 of those studies meeting the full inclusion criteria outlined above. The 55 retrieved studies included 15 randomized

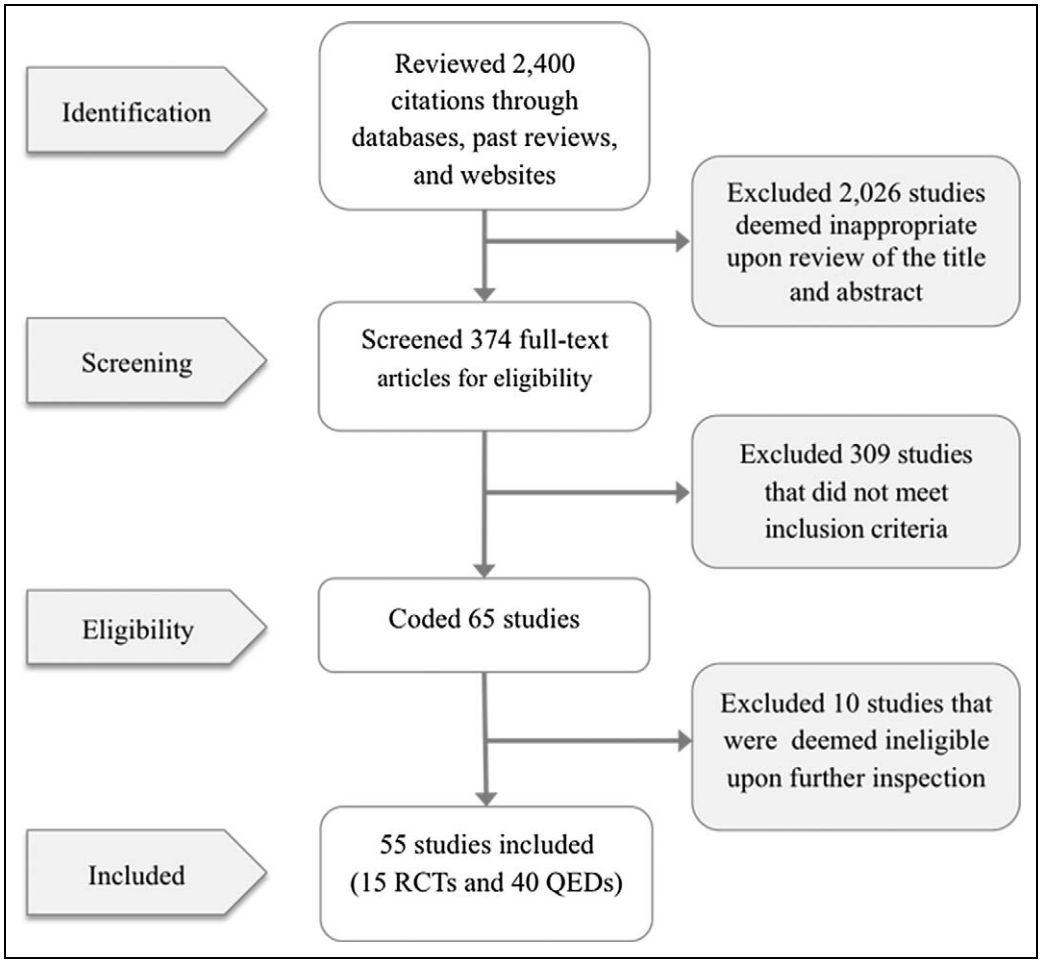

Figure 1. Study search and selection process flowchart. RCT = randomized controlled trial; QED = quasi-experimental design.

controlled trials and 40 quasi-experimental design studies. See Figure 1 for a flowchart detailing the search and selection process.

Coding Procedures and Data Analysis

Included studies were coded by two trained coders using a data-coding instrument developed by the authors to guide systematic examination and extraction of data related to aspects of fidelity and characteristics of the interventions and study designs. The first author coded all of the studies, and a second coder independently coded a random sample of 20% of the studies. Overall agreement between the two coders was assessed, with coders achieving 92% agreement overall and 94% agreement on items related to fidelity components and procedures.

Content analysis of the included studies was conducted to systematically examine the presence of seven key components of intervention fidelity: operationalization of the intervention (i.e., the independent variable), use of a treatment manual, presence of training on the intervention, supervision of the implementers, measurement of fidelity, reliability of fidelity measures, and use of fidelity data in analysis. Each component of intervention fidelity was measured with 1 item. The first component, operationalization of the intervention, was coded using a Likert-type scale question. Coders rated how clearly the author operationalized the treatment procedures on a scale from 1 (*very clear and well-defined treatment could be replicated based on description*) to 4 (*no description of the program was provided*). The item assessing the use of a treatment manual was coded as 0 if the authors did not report use of a treatment manual, 1 if authors reported the use of a manual for at least one component of the intervention, and 2 if authors reported the use of a treatment manual for the entire program. The item assessing training was coded as a 0 if training was not provided, 1 if some training was provided, and 2 if comprehensive training was provided. The item assessing supervision was coded as 0 if supervision was not provided, 1 if the supervision component was built into the program implementation, 2 if supervision was provided for purposes of the study, but not normally part of the intervention, or 3 if some oversight was provided, but it was not systematic. The items assessing components related to measuring fidelity (whether fidelity was measured and if reliable measures were used) and using fidelity in data analysis were coded as a 0 if there was no indication in the study of the measurement or use of fidelity data or 1 if the authors reported measurement of fidelity, use of reliable

Table 1. Number and Types of Fidelity Components Included in Studies.

	n	%
Number of fidelity components		
1	40	72.7
2	27	49.1
3	17	30.9
4	10	18.2
5	8	14.6
6 or more	0	0.0
Components of fidelity		
Well-defined intervention	30	54.6
Use of a treatment manual	23	41.8
Defined training for implementers	18	32.7
Defined supervision for implementers	13	23.6
Measurement of fidelity	16	29.1
Reliability of fidelity measures	0	0.0
Use of fidelity data	2	3.6

fidelity measures, or use of fidelity data in analysis for each of the respective items by totaling each of the components present in the report.

Following data extraction and coding, data were quantitatively synthesized in Statistical Package for the Social Sciences version 20 (IBM Corp., 2011). In addition to analyzing descriptive statistics to describe the characteristics of the included studies, frequencies were calculated for each of the seven fidelity components assessed in this study. In addition, we calculated the total number of fidelity components present in each study.

Results

Fifty-five intervention outcome studies assessing the effects of ASPs with at-risk students were reviewed to examine the extent to which the investigators attended to seven key components of intervention fidelity: operationalization of the intervention, use of a treatment manual, presence of training on the intervention, supervision of the implementers, measurement of fidelity, reliability of fidelity measures, and use of fidelity data in analysis. As seen in Table 1, 40 of the 55 studies demonstrated some concern for, or awareness of, intervention fidelity, as evidenced by attending to at least one component of fidelity examined in this review; however, the extent to which those studies engaged in various aspects of intervention fidelity varied. The use of multiple fidelity components was much less frequent, with about half incorporating two components, 31% incorporating three components, 18% incorporating four components, and just 15% incorporating at least five components. No studies incorporated six or more fidelity components. Following is an examination of the extent to which the reviewed studies included these seven components of intervention fidelity.

Strategies Used to Enhance Fidelity of Intervention

Specific procedures—such as clearly specifying intervention procedures, following a treatment manual, and providing training and supervision to implementers—have been identified in prior research as key factors in promoting and improving intervention fidelity (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). The extent to which investigators engaged in these strategies to enhance intervention fidelity was examined for each study (see Table 1).

The specific strategies the researchers used to enhance fidelity varied between studies. Of the 55 studies included in the review, just more than half (55%) specified well-defined intervention procedures and less than half (42%) indicated the use of a written treatment manual to guide the implementation of the intervention, two critical components to establishing internal validity. Another key aspect of implementing interventions with fidelity is providing training for the implementers. We found a paucity of studies describing implementer training. Of the 55 studies assessed, only 18 (33%) provided information about training. Of these 18 studies, 10 reported providing comprehensive training for implementers and 8 reported offering some implementer training. Similarly, we found an overall lack of information on whether or how the implementation and delivery of the intervention was supervised. Of the 55 included studies, only 13 (24%) provided information about the supervision of the implementers. Of these 13 studies, 10 described supervision components that were built into program implementation procedures, 2 described supervision that was conducted for the purposes of the study, and 1 described the provision of some oversight, but it was not systematic.

Measurement of Fidelity

Of the 55 studies included in this review, only 16 (29%) explicitly measured and collected data on at least one aspect of intervention fidelity. Several reasons for measuring fidelity were reported in the included studies, with the most frequently stated reason being to ensure that treatment was delivered as intended ($n = 13$). Other reasons the authors provided included to improve intervention delivery ($n = 3$) and to establish valid and reliable findings ($n = 3$). Of the 16 studies measuring fidelity, none reported the reliability of fidelity measures.

We examined the relation between study and intervention characteristics and whether researchers reported measurement of intervention fidelity (see Table 2). Studies that used a randomized design were nearly 3 times more likely to measure fidelity than studies that used a quasi-experimental design. Studies that evaluated the effects of interventions that were not guided by a treatment manual were substantially less likely to measure fidelity. ASPs that were local in nature (i.e., not affiliated with a national organization) were less likely to measure fidelity.

The procedures used to measure and collect fidelity data were also assessed in this review. The frequency of fidelity measurement and fidelity data collection methods are summarized in Table 3. The frequency with which fidelity was measured varied across studies and ranged from daily to annually. Of the 16 studies that measured fidelity, the most commonly reported frequencies of fidelity measurement were daily ($n = 5$) and annually ($n = 4$). The remaining studies

Table 2. Reporting of Fidelity Measurement by Study Characteristics.

	Yes	No
Type of report		
Journal	8 (33.3%)	16 (66.7%)
Government report	3 (100.0%)	0 (0.0%)
Unpublished report	2 (20.0%)	8 (80.0%)
Dissertation	3 (16.7%)	15 (83.3%)
Type of project		
Efficacy	4 (26.7%)	11 (73.3%)
Effectiveness	12 (30.0%)	28 (70.0%)
Researcher role		
Delivered intervention	2 (50.0%)	2 (50.0%)
Designed intervention	1 (20.0%)	4 (80.0%)
Independent of intervention	7 (30.4%)	16 (69.6%)
Unsure	6 (26.1%)	17 (73.9%)
Study design		
Randomized controlled trial	8 (57.1%)	6 (42.9%)
Quasi experiment	8 (19.5%)	33 (80.5%)
Setting		
School	11 (36.7%)	19 (63.3%)
Community	3 (18.8%)	13 (81.3%)
Mixed	2 (50.0%)	2 (50.0%)
Unsure	0 (0.0%)	5 (100.0%)
Manualized procedures		
Full	5 (38.5%)	8 (61.5%)
Partial	5 (50.0%)	5 (50.0%)
None	6 (18.8%)	26 (81.3%)
Program affiliation		
National	7 (43.8%)	9 (56.3%)
Local	9 (23.1%)	30 (76.9%)
Number of treatment sessions		
0–50	3 (30.0%)	7 (70.0%)
51–100	3 (50.0%)	3 (50.0%)
101–150	5 (45.5%)	6 (54.5%)
151–200	2 (40.0%)	3 (60.0%)
201–250	1 (20.0%)	16 (88.9%)
Unsure	2 (11.1%)	16 (88.9%)

measured fidelity weekly ($n = 1$), monthly ($n = 1$), quarterly ($n = 1$), or biannually ($n = 1$). The frequency of fidelity measurement was unclear in three of the studies.

The researchers used a variety of methods to collect fidelity data. The most common methods were implementer self-administered checklists ($n = 8$), researcher observations ($n = 9$), and measurement of intervention dose ($n = 9$). Additional methods for collecting fidelity data included interviews with implementers ($n = 6$), participant or parent surveys ($n = 5$), audio or video recording ($n = 3$), and researcher self-administered checklists ($n = 1$). Seven of the studies used one method of fidelity data collection, and 10 of the studies used multiple methods. As part of the evaluation, two of the authors included a sample form used to monitor fidelity.

Use of Fidelity Data

Of the 16 studies that reported fidelity measurement, only 2 studies used fidelity data in their analysis of outcome variables. Specifically, Gottfredson, Cross, Wilson, Rorie, and Connell (2010) collected fidelity data to measure the quality of program implementation at five different sites and used that data to examine whether quality of implementation was associated with more positive outcomes. Schinke, Cole, and Poulin (2000) analyzed student ratings alongside self-reported outcomes to examine the association between degree of participation and student outcomes. For the 14 studies that measured fidelity but did not use fidelity data in their analysis, the most frequently reported use of the fidelity data was to provide feedback to staff members on their current implementation and to assist them in strengthening the programs. Fidelity data provided staff members with an opportunity to identify and address barriers to intervention implementation and make necessary adjustments to adhere more closely to the program model.

Table 3. Measurement of Fidelity.

	n	%
Why fidelity was measured[a]		
Deliver intervention as intended	13	81.3
Improve intervention delivery	3	18.8
Establish reliable or valid findings	3	17.7
Measure group contamination	0	0.0
Not reported	1	6.3
Explicit fidelity measures[a]		
Checklist, implementer	8	50.0
Checklist, researcher	1	6.3
Observations, researcher	9	56.3
Audio or video recording	3	18.8
Implementer interview	6	37.5
Measure of intervention dose	9	56.3
Participant or parent survey	5	31.3
Frequency of fidelity measuring		
Daily	5	31.3
Weekly	1	6.3
Monthly	1	6.3
Quarterly	1	6.3
Biannually	1	6.3
Annually	4	25.0
Unsure	3	18.8

Note. A total of 16 studies measured fidelity.
[a] Categories not mutually exclusive.

Discussion and Applications to Social Work

The popularity and proliferation of ASPs in the United States suggest they fill a need and serve important purposes. The desire to provide youth with positive, prosocial activities for a time when lack of supervision and idleness converge presents a compelling need for well-executed programming. Although the rationale justifying the proliferation of ASPs seems sound, systematic analyses of their outcome effectiveness have been plagued by ambiguous findings. One important factor that could provide insight into the discrepancies between the promise of ASPs and the findings of ASP intervention research is intervention fidelity. Intervention fidelity not only is critical to understanding whether and how interventions relate to outcomes but also has implications for interpretation of research

findings, external validity, statistical power, and the success of the intervention.

The purpose of this study was to examine the extent to which ASP researchers attended to fidelity in ASP intervention studies to better understand this corpus of research and to inform research and practice. Our findings revealed a notable lack of attention to intervention fidelity in the included studies. This paucity of attention to fidelity corroborates findings from prior research of intervention fidelity in education, social work, and psychology (see Gresham & Gansle, 1993; Moncher & Prinz, 1991; Naleppa & Cagle, 2010; Perepletchikova, Treat, & Kazdin, 2007; Swanson et al., 2011; Tucker & Blythe, 2008). Unlike prior reviews examining intervention fidelity reported in journal articles within a specific discipline, this review examined the intervention fidelity of a corpus of studies in multiple disciplines that evaluated the effects of popular and highly regarded interventions: ASPs for at-risk students. Studies of ASP interventions are published across disciplines; thus, it is not surprising that the present study results are reflective of prior reviews of fidelity published solely in social work, education, or psychology journals. It is surprising, however, given ASPs' popularity and widespread support, the vast expense of resources, and often unquestioned claims of positive effects, despite ambiguous evidence, that intervention fidelity in ASP research has been largely ignored.

Of the fidelity strategies examined in this review, the reporting of well-specified intervention procedures occurred most frequently and the use of manuals the second most frequently. However, only 55% of the studies reported well-specified intervention procedures and only 42% reported using a treatment manual. If authors of ASP intervention studies do not provide sufficient details of the ASP components and mechanisms of change, we cannot ascertain what was actually tested. Moreover, even if positive effects on outcome variables were found, we would not be able to determine whether the planned intervention contributed to the outcomes nor would anyone be able to replicate the intervention. As Chen and Rossi (1983) so clearly explained, ''Without careful specification of the treatment as delivered, interpretation of treatment effects may become very muddy indeed'' (p. 294). Clearly specifying the independent variable (i.e., the intervention) is essential to the testing and implementation of the intervention, can contribute to our understanding of the mechanisms of the intervention, and can lead to the intervention being implemented more efficiently and successfully (Fixsen et al., 2005; Summerfelt, 2003). Treatment manuals are important because they provide clear and explicit descriptions of the components of the model, intervention activities, and equipment and material needs (Gearing et al., 2011); guide and standardize the intervention; and help reduce the variability in intervention implementation (Perepletchikova & Kazdin, 2005). To facilitate the clear operationalization of the independent variable and allow for replication of the intervention, it is important that researchers provide sufficient detail of the intervention, describe the essential components and mechanisms of change, and develop and use a treatment manual.

Although explicating a well-defined independent variable is essential, and the use of manuals is an important part of specifying the intervention, neither is sufficient to ensure intervention fidelity. Evidence suggests that due to a number of factors, implementers do not carry out interventions as designed or in the way that researchers assume (Dumas et al., 2001; Fixsen et al., 2005; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Noell et al., 2005); however, there is evidence that training and supervision improve the competence and adherence of implementers, both of which are essential to intervention fidelity (Milne, Baker, Blackburn, James, & Reichelt, 1999; Perepletchikova & Kazdin, 2005). Providing training and supervision of implementers are two frequently recommended strategies for improving intervention fidelity; however, relatively few studies in this review reported providing training or supervision of the implementers. Similar findings were reported in reviews of fidelity in social work and psychology intervention research (Moncher & Prinz, 1991; Naleppa & Cagle, 2010; Tucker & Blythe, 2008). To implement an intervention with fidelity, initial training that gives implementers a thorough knowledge of and sufficient skill level with the intervention is essential, and ongoing training and supervision is important to reduce implementer drift, deviations from the intervention, and decay of implementer skills over time (Bellg et al., 2004; Gearing et al., 2011).

Despite training and supervision, interventions are rarely implemented perfectly. Adaptations to interventions occur due to numerous factors related to implementer, organizational, and intervention characteristics (see Durlak & Dupre, 2008; Fixsen et al., 2005). As such, measuring the degree to which an intervention was implemented is essential to understanding what and how much of the intervention was delivered and the extent to which the intervention differed from the counterfactual condition (Hulleman & Cordray, 2009; Schoenwald et al., 2011; Smith et al., 2007). This knowledge is critical to interpreting and explaining the outcomes of the intervention research, establishing the internal validity of the study, and detecting and correcting poor implementation early (Dumas et al., 2001; Summerfelt, 2003; Swanson et al., 2011). Fidelity measurement can take many forms (i.e., observation, self-report; audio or video recording), and in most cases, must be designed specifically for the intervention (O'Donnell, 2008). Of the 55 intervention studies assessed in this review, only 16 (29%) measured fidelity. Of these 16 studies, none reported the reliability of the fidelity measures and only 2 used fidelity data in their analysis. The paucity of ASP intervention studies that measure intervention fidelity limits the confidence in the functional relationships between the intervention and the outcomes and calls into question the internal and external validity of the studies.

The lack of reporting of intervention fidelity in studies included in this review seriously limits the utility of ASP intervention research to inform evidence-based practice and policy. A reasonable question to ask at this point is: If intervention fidelity has such important implications for internal and external validity, power, and effect and is recognized as a methodological necessity for intervention research (Perepletchikova et al., 2007),

why is there such a paucity of published ASP intervention studies attending to intervention fidelity? Although we are not aware of studies examining the barriers to fidelity monitoring and measurement, there are several potential reasons for the lack of attention to fidelity in ASP intervention studies.

Despite increased attention to fidelity in intervention research, practitioners and researchers may have a general lack of awareness of the critical importance of intervention fidelity. Perepletchikova, Treat, and Kazdin (2007) recommend increasing awareness and training of fidelity issues through various professional outlets, such as conference presentations, symposia, and special sections of journals devoted to intervention fidelity. Training future practitioners and researchers about fidelity and fidelity measurement while they are students is also critical to improving the frequency and sophistication of intervention fidelity in research (Naleppa & Cagle, 2010; Smith et al., 2007). Even when researchers are knowledgeable and understand the importance of intervention fidelity, additional barriers may undermine their ability to attend to fidelity in their research. Building strategies to enhance, monitor, and measure fidelity in intervention research is expensive and time intensive; significant planning and resources are needed. Moreover, there is an overall lack of research regarding the most effective and efficient strategies and measurement procedures (Durlak & Dupre, 2008). The measurement of fidelity is in itself a significant challenge, as fidelity measures often need to be developed specific to the intervention. As a result, the reliability and validity of those measures is unknown and can be established only after the study is complete (Schoenwald et al., 2011), thus creating challenges to using valid and reliable fidelity measures.

In addition to a general lack of awareness and education of intervention fidelity and additional barriers, there appears to be relatively few expectations of attending to and reporting intervention fidelity in published research. Professional standards and guidelines for reporting intervention fidelity are largely missing (Naleppa & Cagle, 2010; Smith et al., 2007; Swanson et al., 2011). Indeed, page limits have been cited as a barrier to providing fidelity processes and data in journal articles; however, sacrificing fidelity data to bring articles within publisher page limits sends a message that fidelity is not important (Perepletchikova et al., 2007). One can argue that the specification and measurement of the independent variable should require the same level of detail and attention as the dependent variables and thus should be given adequate space in journal articles (McIntyre, Gresham, DiGennaro, & Reed, 2007). Durlak and Dupre (2008) and Mayo-Wilson (2007) acknowledged the importance of reporting implementation and fidelity and have made recommendations to expand journal policies and reporting guidelines to include reporting of intervention implementation and fidelity in published research reports. Indeed, published standards and guidelines for reporting randomized trials, such as the Consolidated Standards of Reporting Trials, have begun to include standards for reporting details of intervention implementation in the social and psychological sciences (Grant, Mayo-Wilson, Melendez-Torres, & Montgomery, 2013). Moreover, several federal and private funders, such as the National Institutes of Health and the Institute of Education Sciences, have increased attention to fidelity in their calls for proposals. Adding standards for reporting implementation and fidelity to intervention research reporting guidelines and as a requirement to secure research funding will encourage researchers to attend to and report this information while also encouraging peer reviewers and editors to require this information from authors (Mayo-Wilson, 2007).

Although barriers to intervention fidelity do exist and much work still needs to be done in this area, attending to fidelity in intervention research is critical and needs to be considered a central concern of intervention researchers and consumers of intervention research. To that end, several scholars have provided recommendations and guidelines to assist researchers in understanding, enhancing, monitoring, and measuring fidelity in intervention research (see Carroll et al., 2007; Durlak & Dupre, 2008; Fixsen et al., 2005; Gearing et al., 2011; Moncher & Prinz, 1991; Perepletchikova et al., 2007). It must be noted, however, that consensus regarding constructs and definitions of fidelity or what constitutes the most important or necessary components of intervention fidelity has not been reached. Indeed, fidelity remains a contentious topic in many fields; intervention fidelity and implementation are relatively new issues and thus are not well studied or understood. Perhaps some of the most critical steps we can take at this point are to be aware of the issues related to intervention fidelity, take steps to attend to aspects of fidelity that are relevant and possible for a particular intervention study, and make every effort to be transparent in the reporting of intervention research.

The findings of the present study must be interpreted in light of the study's limitations. First, this review is limited to studies that examined the effects of ASP interventions for at-risk youth and that met the other inclusion criteria. Also, we may not have captured every eligible ASP intervention study, despite our comprehensive and systematic search process. Findings from this review may not generalize to studies that examine the effects of different types of ASP interventions or studies that we excluded or did not identify in the search. Further, the use of fidelity strategies and assessment appear to be related to study quality and, as such, findings from this review may not reflect other areas of applied research with a strong set of studies. However, many nascent, programmatic areas in social work and education that employ interventions are likely to have similar fidelity deficits. This analysis also was limited to the fidelity components we identified and to the information the authors provided. It is possible that study authors reported other components of fidelity or attended to fidelity but did not provide the information in the published article. Thus, it is possible the results of this review underestimate the frequency with which ASP intervention research uses fidelity procedures.

Conclusion

Demonstrating the fidelity of an intervention is a critical component of intervention research; fidelity has important implications

for the design, delivery, testing, and validity of inferences of intervention research. Indeed, "the cost of inadequate fidelity can be rejection of powerful treatment programs or acceptance of powerless programs" (Moncher & Prinz, 1991, p. 250). Although ASP intervention research aims to determine whether ASP interventions make a positive difference in the lives of at-risk youth, it is clear from the lack of attention to fidelity found in this corpus of studies that the vast majority of ASP intervention research studies are inadequate to draw valid inferences from the results. In short, the lack of attention to intervention fidelity in ASP intervention research hampers our ability to use the extant intervention research to make evidence-based decisions about ASPs. It is important that social work practitioners and policy makers are aware of this deficiency in ASP intervention research and how this deficiency affects the use and interpretation of ASP intervention study results. Moreover, current and future social work researchers need to make greater efforts to be transparent about issues related to fidelity, use strategies to enhance and ensure intervention fidelity, measure intervention fidelity, and report fidelity data in published studies.

Authors' Note

The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting entities.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are grateful for support from the Meadows Center for Preventing Educational Risk, the Greater Texas Foundation, the Institute of Educational Sciences (Grants R324A100022 and R324B080008), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P50 HD052117).

References

Afterschool Alliance. (2012). *Afterschool essentials: Research and polling*. Retrieved from http://www.afterschoolalliance.org/Essentials_and_Polling_2013_010713.pdf

Apsler, R. (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence*, *44*, 1–19.

Belden Russonello & Stewart. (2003). School board presidents' views of after-school programs in American schools: Results of survey research conducted for the National School Boards Association. Retrieved from http://www.nsba.org/Board-Leadership/EDLO/WhatBoardsCanDo/ResearchandEvaluation/2003SurveyResults/PresidentViews.pdf

Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, *23*, 443–451. doi:10.1037/0278-6133.23.5.443

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 40. doi:10.1186/1748-5908-2-40

Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, *7*, 283–302. doi:10.1177/0193841X8300700301

Cook, T. J., & Poole, W. K. (1982). Treatment implementation and statistical power: A research note. *Evaluation Review*, *6*, 425–430. doi:10.1177/0193841X8200600305

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine*, *20*, 38–47. doi:10.1016/S0749-3797(00)00272-5

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. doi:10.1007/s10464-008-9165-0

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*, 294–309. doi:10.1007/s10464-010-9300-6

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research: Theory and Practice*, *18*, 237–256. doi:10.1093/her/18.2.237

Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk.

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

Gearing, R. E., El-Bassel, N., Gesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, *31*, 79–88. doi:10.1016/j.cpr.2010.09.007

Gottfredson, D. C., Cross, A., Wilson, D., Rorie, M., & Connell, N. (2010). An experimental evaluation of the All Stars prevention curriculum in a community after school setting. *Prevention Science*, *11*, 142–154. doi:10.1007/s11121-009-0156-7

Grant, S. P., Mayo-Wilson, E., Melendez-Torres, G. J., & Montgomery, P. (2013). Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PLOS ONE*. doi: 10.1371/journal.pone.0065442

Gresham, F. M., & Gansle, K. A. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, *22*, 254–273. doi:10.1901/jaba.1993.26-257

Hester, P. P., Baltodano, H. M., Gable, R. A., Tonelson, S. W., & Hendrickson, J. M. (2003). Early intervention with children at risk of emotional/behavioral disorders: A critical examination of

research methodology and practices. *Education and Treatment of Children*, *26*, 362–381.

Hollister, R. (2003). *The growth in after-school programs and their impact*. Washington, DC: The Brookings Institution.

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. doi:10.1080/19345740802539325

IBM Corp. (2011). IBM SPSS Statistics for Windows (Version 20.0) [Computer software]. Armonk, NY: Author.

Kaye, S., & Osteen, P. J. (2011). Developing and validating measures for child welfare agencies to self-monitor fidelity to a child safety intervention. *Child and Youth Services Review*, *33*, 2146–2151. doi:10.1016/j.childyouth.2011.06.020

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 275–313. doi:10.3102/00346543076002275

Mahoney, J. L., Parente, M. E., & Zigler, E. F. (2009). Afterschool programs in America: Origins, growth, popularity, and politics. *Journal of Youth Development: Bridging Research and Practice*, *4*, 24–43.

Mayo-Wilson, E. (2007). Reporting implementation in randomized trials: Proposed additions to the consolidated standards of reporting trials statement. *American Journal of Public Health*, *97*, 630–633.

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.

Milne, D. L., Baker, C., Blackburn, I. M., James, I., & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry*, *30*, 81–92. doi:10.1016/s0005-7916(99)00011-7

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*, 247–266. doi:10.1016/0272-7358(91)90103-2

Mooney, P., Epstein, M., Reid, R., & Nelson, J. R. (2003). Status and trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education*, *24*, 273–287. doi:10.1177/07419325030240050301

Naleppa, M. J., & Cagle, J. G. (2010). Treatment fidelity in social work intervention research: A review of published studies. *Research on Social Work Practice*, *20*, 674–681. doi:10.1177/1049731509352088

National Association of Elementary School Principals. (2001). *Principals and after-school programs: A survey of preK-8 principals fact sheet*. Retrieved from http://www.naesp.org/after-school-programs-archives-0

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, *31*, 217–234. doi:10.1901/jaba.2000.33-271

Noell, G. H., Witt, J. C., Slider, N. J., Conell, J. E., Gatti, S. L., Williams, K. L., … Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, *34*, 87–106.

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. doi:10.3102/0034654307313793

Parsad, B., & Lewis, L. (2009). *After-school programs in public elementary schools* (NCES 2009-043). Washington, DC: National Center for Education Statistics.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and recommendations. *Clinical Psychology: Science and Practice*, *12*, 365–383. doi:10.1093/clipsy/bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*, 829–841. doi:10.1037/0022-006X.75.6.829

Resnick, B., Bellg, A. J., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., … Czajkowski, S. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine*, *29*, 46–54. doi:10.1207/s15324796abm2902s_8

Roth, J. L., Malone, L. M., & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology*, *45*, 310–324. doi:10.1007/s10464-010-9303-3

Schinke, S. P., Cole, K. C., & Poulin, S. R. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science*, *1*, 51–60. doi:10.1023/A:1010076000379

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 32–43. doi:10.1007/s10488-010-0321-0

Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*, 387–419. doi:10.1177/109821400202300403

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children*, *30*, 121–134. doi:10.1353/etc.2007.0033

Summerfelt, W. T. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science*, *7*, 55–61. doi:10.1207/S1532480XADS0702_2

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*, 33–13. doi:10.1177/0022466911419516

Tucker, A. R., & Blythe, B. (2008). Attention to treatment fidelity in social work outcomes: A review of the literature from the 1990s. *Social Work Research*, *32*, 185–190. doi:10.1093/swr/32.3.185

U.S. Department of Education. (2011). 21st Century Community Learning Centers: Purpose. Retrieved from http://www2.ed.gov/programs/21stcclc/index.html

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620–630. doi:10.1037/0022-006X.61.4.620

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337–363. doi:10.1146/annurev.psych.55.090902.141449

Zief, S. G., Lauver, S., & Maynard, R. A. (2006). Impacts of after-school programs on student outcomes. *Campbell Systematic Reviews*, *3*. doi:10.4073/csr.2006.3