Abstract Title Page

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

Authors and Affiliations:

Brandy R. Maynard, Ph.D.
Saint Louis University
Bmaynar1@slu.edu

Kristen E. Peters, MSW
Saint Louis University
Kpeter26@slu.edu

Michael G. Vaughn, Ph.D.
Saint Louis University
Mvaughn9@slu.edu

Christine M. Sarteschi, Ph.D.
Chatham University
Christyj3@gmail.com

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

Background / Context:

Over the past 2 decades, the number of after-school programs (ASP) and the number of students attending ASPs has markedly increased. Today, approximately 50,000 public elementary schools and numerous more middle and high schools offer some type of ASP (Parsad & Lewis, 2009). ASPs have thrived, at least in part, because they are viewed as important and beneficial to students, families, schools, and communities. Public recognition for the *potential* of ASPs to improve behavioral and academic outcomes has resulted in the influx of funding and growth of 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, and over $1billion dollars in federal funding of ASPs, many are left questioning how effectively ASPs achieved the perceived benefits for students, families, schools, and society. While numerous studies and several reviews have assessed the effects of ASPs, the results have provided only an ambiguous picture of the effects of these programs (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).

Although several reviews and meta-analyses have examined the outcomes of ASPs, ASP intervention study reviews have not specifically examined intervention fidelity. 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). 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 / Objective / Research Question / Focus of Study:

Given the popularity of ASPs and the growing body of intervention research resulting in ambiguous findings, it seems prudent to examine whether investigators have attended to fidelity. The purpose of this systematic review was to examine whether and to what extent researchers addressed intervention fidelity in ASP intervention research. The following research questions guided 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?

Setting:

Studies included in this review were conducted in after-school program settings.

Population / Participants / Subjects:

Fifty-five studies (15 RCT and 40 QED) were included in this review. Published and unpublished studies conducted between 1980 and 2012 were eligible for inclusion 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. 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). 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.

(Insert Figure 1 here)

Intervention / Program / Practice:

This study examined the extent to which fidelity was addressed in outcome research of after-school program interventions, 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).

Research Design:

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

Data Collection and Analysis:

Several sources were used to identify eligible published or unpublished studies conducted or published between January, 1980 and March, 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 was consulted to determine the appropriate databases to search and keyword search terms to use. Keyword 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) 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 were retrieved in full text and screened for eligibility.

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. Following data extraction and coding, data were quantitatively synthesized in SPSS 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.

Findings / 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 the seven key components described above. As seen in Table 1, 40 of the 55 studies attended to at least some aspects of intervention fidelity; however, the extent to which those studies engaged in various aspects of intervention fidelity varied. While the majority of studies attended to at least one aspect of intervention fidelity, the use of multiple fidelity components was much less frequent. About half of the studies incorporated two components, 31% incorporated three components, 18% incorporated four components, and just 15% incorporated at least five components. No studies incorporated six or more fidelity components.

(Table 1 here)

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 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. We found a paucity of studies describing implementer training, finding only 18 (33%) studies provided information about training. Similarly, we found an overall lack of information on whether or how the implementation and delivery of the intervention was supervised; only 13 (24%) provided information about the supervision of the implementers.

Measurement and Use 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. 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 three 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. Researchers used a variety of methods to collect fidelity data, with the most common methods being implementer self-administered checklists (n = 8), researcher observations (n = 9), and measurement of intervention dose (n = 9).

Of the 16 studies that reported fidelity measurement, only 2 studies used fidelity data in their analysis of outcome variables. 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.

Tables 2 and 3 here

Conclusions:

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 hampers our ability to use the extant intervention research to make evidence-based decisions about ASPs. It is important that practitioners and policymakers are aware of this deficiency in ASP intervention research and how this deficiency affects the use and interpretation of study results. Moreover, current and future researchers are encouraged 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.

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 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 education and social sciences 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.

Appendices

Appendix A. References

Apsler, R. (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence, 44*(173), 1–19.

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

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.

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

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

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*(2), 275–313. doi: 10.3102/00346543076002275

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

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

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

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*(4), 387–419. doi: 10.1177/109821400202300403

Summerfelt, W. T. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science, 7*(2), 55–61. doi: 10.1207/S1532480XADS0702_2

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*. Advance online publication. doi:10.1177/0022466911419516

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

Appendix B. Tables and Figures

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

Table 2

Reporting of Fidelity Measurement by Study Characteristic

	Yes	No
Type of report		
Journal	8 (33.0%)	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%)

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.

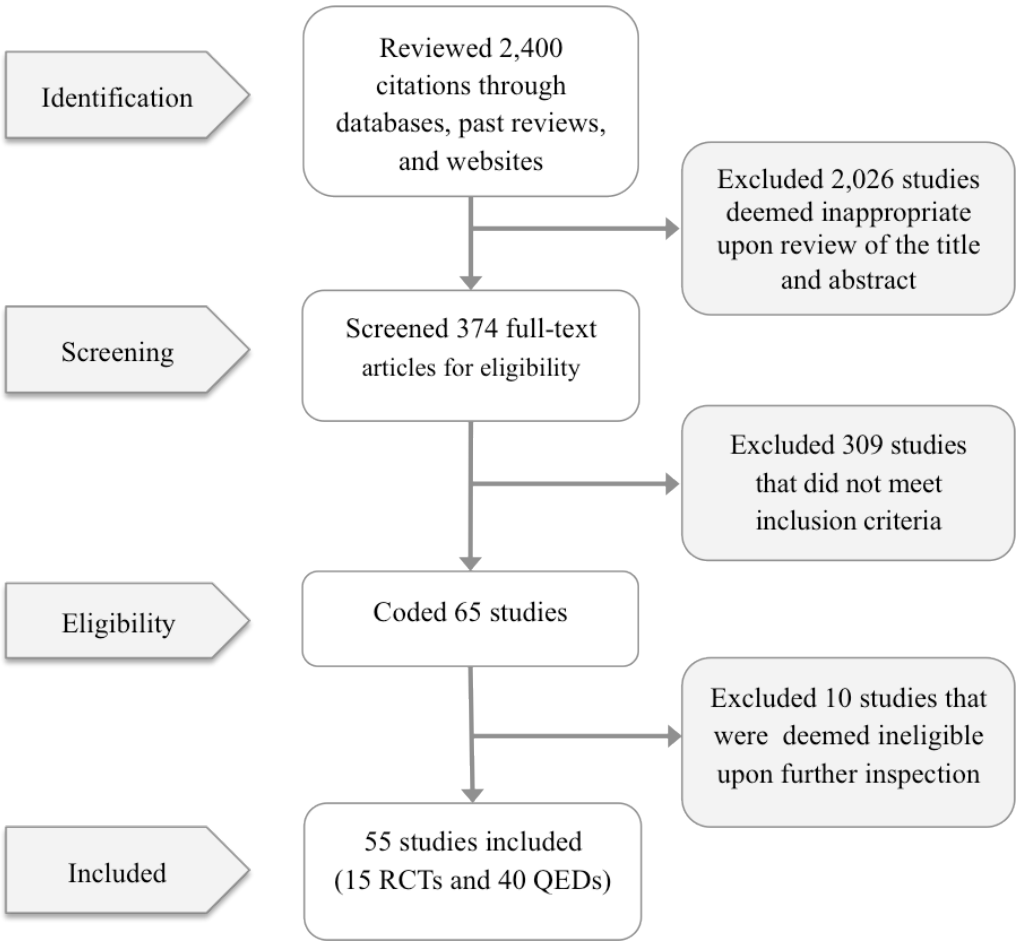

Figure 1. Study search and selection process flow chart. RCT = randomized controlled trial; QED = quasi-experimental design.