

MEASURING SELF-REPORTED FIDELITY IN RECREATION: THE FACILITATOR CHARACTERISTICS AND PROGRAM CONTRIBUTIONS SCALE

Ryan J. Gagnon, M.A.
Graduate Research Assistant
Clemson University
Department of Parks, Recreation, and Tourism Management
432 Edwards Hall
Clemson, SC 29634-0735
(509) 270-0197
rjgagno@g.clemson.edu

Barry A. Garst, Ph.D.
Associate Professor
Clemson University

Garrett A. Stone, M.S.
Graduate Research Assistant
Clemson University

Abstract

Social science researchers have called for evidenced based practices when implementing, delivering, and evaluating programs. One important component of these practices is to monitor program implementation. However, in experiential education and more broadly the social sciences program implementation is rarely assessed or evaluated; when it is assessed it is frequently resource intensive. As such, program providers have needed an implementation measurement approach more conducive to low-resource organizations. This paper introduces a self-report instrument to measure facilitator characteristics that contribute to delivering a program as designed. A Confirmatory Factor Analysis of 121 program facilitator responses to the Facilitator Characteristics and Program Contributions Scale (FCPC) indicated that the FCPC is a valid and reliable measure of facilitator characteristics and their pro-fidelity beliefs, and thus may be a promising alternative to other more resource intensive approaches to implementation assessment.

Keywords: fidelity, experiential education, program implementation, evaluation, facilitation, facilitator characteristics and program contributions scale,

The authors would like to thank Dr. Matthew Chase of Eastern Washington University, Dr. Eddie Hill of Old Dominion University, and Dr. Dewayne Moore of Clemson University for their help with this study and the anonymous reviewers for their helpful feedback on this manuscript.

Measuring Self-Reported Fidelity in Recreation: The Facilitator Characteristics and Program Contributions Scale

More than 30 years ago Ewert (1983) noted the presence of a “black box” within program implementation, a situation in which program providers “...know something works but we don’t know how or why” (p. 126). This “black box” concept conveys the primary motivation for implementation assessment (Mainieri & Anderson, 2014) and may tell a more complete story of a program, especially in situations where desired program outcomes are not achieved. In these cases, as Caldwell et al. (2008) described, “small effect sizes or findings inconsistent with well-reasoned hypotheses may not be related to the efficacy of the program as it was designed, but rather be related to failure to implement the program as intended” (p. 148). Though arguably just as important, program implementation is inherently more difficult to measure than program outcomes (Berkel, Mauricio, Schoenfelder, Sandler, 2011). Rather than simply understanding whether or not a program effect was achieved, implementation assessment seeks to determine to what degree a program was implemented as designed, the quality of program delivery, adaptations or omissions of program material, and the participant’s influence and receptiveness to the program (Berkel et al., 2011; Durlak & DuPre, 2008). Compounding this challenge is the fact that the measurement of program implementation is still very much in a developmental stage (Mainieri, & Anderson, 2014; Tucker & Rheingold, 2010) and “no widely accepted standards of measurement exist” (Berkel et al., 2011, p. 25). Thus, the purposes of this study were to examine the complexity of program implementation measurement and introduce a promising scale that measures program implementation within the context of leisure, recreation, and experiential education. Such an examination requires a closer look at two areas central to implementation science—the layers of program implementation and correspondingly program fidelity.

When considering program implementation four inter-related layers have been identified in the literature. These layers, which can influence the quality of implementation and thereby the quality of the program outcomes achieved (Gagnon et al., 2015), include characteristics associated with: (1) the organization providing the program, (2) the community being served, (3) the program itself, and (4) the primary focus of this study the facilitator delivering the program (Berkel et al., 2013; Durlak & DuPre, 2008; Gagnon, 2014; Little, Sussman, Sun, & Rohrbach, 2013). A visual representation of how these layers’ may contribute to program implementation and corresponding outcomes is presented in Figure 1 adapted from Gagnon et al. (2015).

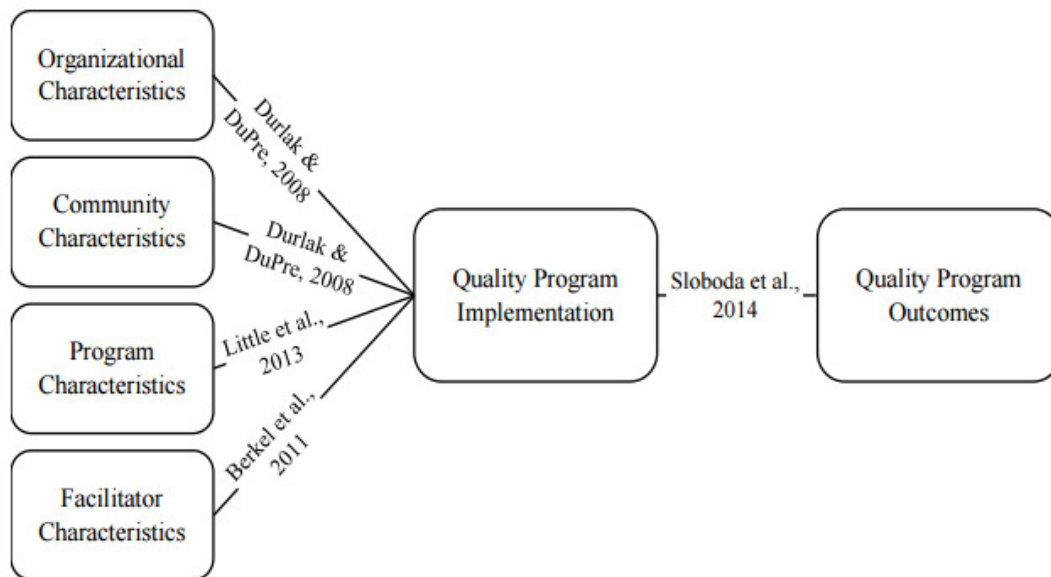


Figure 1. Conceptual Model of the Factors Contributing to Quality Program Implementation and Corresponding Outcomes (Adapted from Gagnon et al., 2015).

As highlighted in figure 1 the characteristics of the organization responsible for the program may include the administrators responsible for the selection of the program, those providing logistical and resource support, and those providing the physical space where the program takes place (Kam, Greenberg, & Walls, 2003). Clearly this layer can dramatically influence a programs implementation quality. For example, as highlighted by the meta-analysis conducted by Durlak and DuPre (2008) organizational characteristics such as the guidance provided by organizational leaders does a have a strong influence on program implementation. The characteristics of the community being served also may have an influence on the quality of program implementation. For example, in a study of programmatic implementation Castro, Barrera, and Martinez (2004), highlighted the issue of a cultural mismatch within program design and corresponding implementation. “If a program is designed for higher-resourced, English-speaking participants, but is delivered to lower-resourced, Spanish-speaking participants, it is no surprise that

the quality of implementation may suffer” (Gagnon et al., 2015, p. 72). Simply put, programs are typically designed with a participant group in mind, serious considerations for alternative groups should be made at the design phase to prevent potential issues such as cultural mismatch.

The program itself also can compromise its own implementation. For example, if a program is too complex for those outside of the original design team to deliver then it should not be surprising that concessions to implementation quality may have occurred. Perhaps conversely, if a program is too simple alterations and changes in program implementation quality may occur to ensure higher levels of participant engagement (Carroll et al., 2007). Finally, the facilitator and their characteristics clearly can have an impact on program implementation. Facilitator characteristics such as their broad experience facilitating and managing groups, their specific experience working with the program being delivered (Nobel et al., 2006), their level, frequency, and quality of training, their level of competency (Dusenbury, Brannigan, Falco, & Hansen, 2003; Stein et al., 2008), their program buy-in (Durlak & DuPre, 2008), and pro-implementation beliefs (Dane & Schneider, 1998) all may contribute to their quality of program implementation and thus may have a profound influence programmatic outcomes (Kreft, Jongbloet, & Van Have, 2014). The confluence of these characteristics and factors and their clear influence on program implementation (and thus outcomes) was the primary rationale for the development of a measure to better understand the facilitator influence on program implementation as part of a larger and more complex model.

Measuring Fidelity

Within the context of implementation, a primary component is the construct of fidelity. Indeed, the term fidelity is often used synonymously with implementation. This construct confusion has led to a high level of ambiguity about the differentiation between the constructs as evidenced below. Definitions of fidelity vary greatly in both the broader leisure and social science literature. Some scholars define fidelity as adherence to a program curriculum (Berkel et al., 2011; Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005). Other researchers treat fidelity with a higher level of complexity. For example, for Sanetti and Kratochwill (2009), fidelity includes “(a) content: what intervention steps were delivered; (b) quality: how well the intervention steps were delivered; (c) quantity: how much of the intervention was provided; and (d) process: how the intervention was delivered” (p. 446). The confusion associated with the use and assessment of fidelity within the social science literature is also amplified by how scholars have described the concept. For example, fidelity has many synonyms in current research including treatment integrity (Gresham, 1989), program integrity (Dane & Schneider, 1998), adherence, program fidelity (Dusenbury et al., 2005), fidelity of implementation (Stein et al., 2008), treatment adherence, and degree of implementation (McGrew, Bond, Dietzen, & Sayers, 1994). For this study we define *fidelity* as the degree to which a program was delivered as intended by the program designers.

The objective of measuring fidelity is to determine if a program was delivered as intended by the program developers (Hogue, Liddle, & Rowe, 1996). This measurement also helps to ensure that fidelity is even possible (Johnson, Mellard, Fuchs, & McKnight, 2006). Fidelity measurement can also illumine the impact maintaining a high level of fidelity may have on program outcomes (Dusenbury et al., 2003). If fidelity is maintained, but desired outcomes are not achieved, then this can guide program designers to possible alterations of the program aims or structure. The measurement of fidelity can also be used to determine what program aspects were and were not present and what adaptations, deviations, and omissions may have been made by the facilitator(s) (Mowbray, Holter, Teague, & Bybee, 2003). Identification of these changes may lead to programmatic design changes and thus enhance participants’ received outcomes.

Fidelity can be measured using direct and/or indirect observation (Gresham, 1989). *Direct observation* generally consists of a trained rater observing a facilitator delivering a program. For a “valid” observation to occur, raters must be highly trained in both the program (to understand its nuances) and in facilitation of the program (Bishop et al., 2013). Direct observation, however, can negatively impact “true” fidelity and therefore obscure the validity of the evaluation. For example, the presence of an observer may lead the facilitator to deliver the program at a higher level of fidelity than is typical or, conversely, cause the facilitator to be so uncomfortable that he/she delivers the program poorly (Spoth, Gyll, Trudeau, & Gold-Lillehoj, 2002). Another potential limitation of the use of direct observation is the potential higher cost associated with the training and labor expenses of the program observers.

Indirect observation methods for monitoring fidelity include interviewing, analyzing video of program delivery, and the method explored in this paper, assessment of self-report data (Gresham, 1989). Self-reports require fewer resources for the monitoring of fidelity, but often at an immaterial cost. Facilitators may be influenced by social desirability (i.e., the desire to be perceived as appropriately following the program as designed) and thus provide inaccurate, inflated reports of fidelity (Gresham, Macmillan, Beebe-Frankenberger, & Bocian, 2000). However, social desirability effects are not a given; in an assessment of teachers implementing a program targeted towards improving student mathematical performance, Hagermoser-Sanetti and Kratochwill (2009) found that teachers were accurate self-reporters of their own fidelity to a program design. With the need for an easy to use self-report measure of program fidelity in mind, this study explored the efficacy of the Facilitator Characteristics and Programmatic Contributions Scale (FCPC).

FCPC Development. The FCPC was developed after extensive research investigating facilitator characteristics that may influence program fidelity, including pro-fidelity beliefs (e.g., “It is important to deliver this program as designed.”), program buy-in (e.g., “I would recommend this program to other groups.”), facilitator experience delivering programs (e.g., “I have enough experience to deliver this program, and facilitator training.”) After developing the items and examining them for face validity, a study was conducted using the 26-item “long-form” version of the FCPC and a sample of college students participating in an experiential program (Gagnon, 2014). An Exploratory Factor Analysis (EFA) of the instrument was conducted and the results of this EFA indicated that items loaded on two primary factors: (1) buy-in and pro-fidelity beliefs and (2) belief in experience and training rather than previously theorized four factors: (1) buy-in, (2) pro-fidelity beliefs, (3) belief in experience, and (4) belief in training (Gagnon & Bumpus, In-Press). After the completion of the EFA the redesigned FCPC was delivered with the poorer performing items eliminated (those items that either cross loaded between factors or did not load unidimensionally on a single factor). This study presents the results of the redesigned and shorter FCPC.

Methods & Procedures

Facilitators ($N = 121$) from three distinct university programs were recruited to complete the FCPC at the conclusion of their various experiential education programs (See Table 1). Facilitators were primarily white (72.3%), female (61%), and had completed at least some college (72%). Facilitators from site one ($n = 81$) were located in coastal Virginia and were delivering a semester long resiliency development program to at-risk middle school youth. Facilitators from site two ($n = 28$) were located in the Pacific Northwest and were delivering a leadership development program to incoming university freshman students over a two-day period. Facilitators from site three ($n = 12$) were also located in the Pacific Northwest and delivering a leadership development program, but to college juniors and seniors. Demographic information and data reflecting the facilitator's experience and training were also collected, including: hours of experience facilitating, number of program specific trainings attended, number of challenge course trainings attended in hours, as well as gender, education level, and ethnic group.

Table 1

Sample Demographics and Descriptive Information, ($N = 121$)

Gender	Male, $n = 44$, 38.6%	Female, $n = 70$, 61.4%	
Race	White, $n = 84$, 72.3%	African American, $n = 19$, 17%	Latino, $n = 3$, 2.7%
	Multiple Race, $n = 6$, 5.4%	Pacific Islander, $n = 2$, 7.4%	No Response, $n = 9$, 7.4%
Education Level	Some College, $n = 88$, 72.7%	Associates Degree, $n = 8$, 6.6%	Bachelor's Degree, $n = 10$, 8.3%
	Master's Degree, $n = 7$, 6.1%	Ph.D., $n = 1$, .9%	
Hours of Experience Facilitating Groups	$M = 240.05$, $SD = 418.75$, Range 0 – 2,000 hours		
Program Specific Trainings Attended	$M = 13.87$, $SD = 13.19$, Range 0 - 85 Trainings		
Hours of Challenge Course Specific Training	$M = 12.79$, $SD = 36.60$, Range 0 – 250 Hours		

Scale Validation

Validation of the FCPC. Data were screened for outliers using critical values and scree plots of Cook's D, Mahalanobis Distance, and Studentized Deleted Residuals which resulted in the removal of three cases. Bartlett's Test of Sphericity was statistically significant ($p \leq .0005$), indicating that the data was likely factorizable and a large enough sample was present. Confirmatory Factor Analysis (CFA) was employed to validate the FCPC. Simply put, CFA assumes an a priori model with a fixed number of latent factors, a fixed number of items (in this case the survey questions) loading on theorized factors, and is supported by prior research and theory (Brown, 2015). A two factor model was specified consisting of Experience and Training (Factor 1) and Buy-In and Pro-Fidelity Beliefs (Factor 2) based on current fidelity and implementation research and the findings of the EFA results of the FCPC (Gagnon, 2014).

The sample variance-covariance matrix was analyzed with EQS 6.2 software and the maximum likelihood minimization function. Goodness of fit was estimated by using the Chi-Square, the Root Mean Square Error of Approximation (RMSEA), its 90% confidence interval, the Standardized Root Mean-Square Residual (SRMR), the Comparative Fit Index (CFI), and the Joreskog-Sorbom GFI Index. Used together these indices provide a conservative and reliable evaluation of the two-factor solution. Each of the overall goodness-of-fit indices suggested that the two-factor model fit the data well: $\chi^2(25) = 32.040$, $p \leq .001$, RMSEA = .066 (90%, CI .000-.125), SRMR = .068, CFI = .950, GFI = .911.

Convergent and discriminant validity were examined to assess the appropriateness of the FCPC for measuring facilitator beliefs about experience and training as well as facilitator pro-fidelity beliefs. Convergent validity is the degree of agreement between two or more measures of the same latent variable. This is typically examined by looking at statistics such as factor loadings and reliability (Byrne, 2006). Both factors reported good convergent validity as evidenced by the Average Variance Extracted (AVE) scores with Experience and Training (F1) manifesting an AVE of .591 and Pro-Fidelity Beliefs (F2) demonstrating an AVE of .531, both above the recommended threshold of .5 (Fornell & Larcker, 1981). The two factors also reported good reliability as evidenced by Joreskog's Rho (ρ) and Cronbach's Alphas (α) with F1 reporting alphas of $\rho = .923$, $\alpha = .882$ (5 items) and F2 reporting alphas of $\rho = .867$, $\alpha = .816$ (4 items). Please note we only report Cronbach's due to its heavy use in the social sciences. Joreskog's Rho is a better reliability measure than Cronbach's alpha in Structural Equation Modeling (in this case CFA), as it based on factor loadings rather than correlations between observed variables (for more information see Chin, 1998). Discriminant validity is the degree to which items are measuring distinct factors (Brown, 2015). To determine discriminant validity, the correlations between factors are examined with low correlations indicating better discriminant validity (Byrne, 2006). Discriminant validity between factors was evidenced by low correlation between the two factors (F1 and F2 $r = .167$) (see Table 3 for the complete FCPC description). In summary, there is evidence of both convergent and discriminant validity of the nine item FCPC, thus indicating it is a reliable and valid measure of beliefs about experience and training and pro-fidelity beliefs.

Table 2.

Confirmatory Factor Analysis Results of the Facilitator Characteristics and Program Contributions (FCPC) Scale ($N = 121$)

Factor	M (SD)	β	ρ	α	AVE
<i>Experience and Training</i>					
I consider myself experienced in facilitating groups	5.62 (1.21)	.66	.923	.882	.591
I had enough training to facilitate this program	5.54 (1.43)	.84			
I feel prepared for this program due to my general facilitating experience	5.79 (1.69)	.74			
I feel well trained to facilitate this program	5.74 (1.12)	.94			
I have a high level of experience facilitating groups	5.34 (1.41)	.62			
<i>Buy-in and Pro-Fidelity Beliefs</i>					
I followed the pre-designed program plans for this program	5.46 (1.29)	.57	.867	.816	.531
I am "bought-in" to this program	5.35 (1.69)	.87			
I would recommend this to other groups	5.75 (1.59)	.84			
It was important to deliver this program as it was designed	4.92 (1.40)	.58			

β : standardized coefficient (factor loading); ρ : Joreskog's Rho; α : Cronbach's Alpha; AVE: Average Variance Extracted.

Discussion, Limitations, and Future Directions

The study findings suggest that the FCPC is a promising mechanism for assessing the degree to which a program is actually delivered. The CFA results indicate that the FCPC accurately measures the constructs of buy-in, pro-fidelity beliefs, and beliefs about experience and training. This post-program information is not only valuable in terms of helping program designers understand the degree to which the program was delivered as designed, but could also serve as a useful screening tool when selecting staff to deliver a program (i.e., program heads might seek to employ staff who had higher levels of training, buy-in, and pro-fidelity beliefs as determined by the FCPC). Furthermore, the combination of the FCPC (as a process evaluation measure) and traditional outcome measures could provide a more holistic picture of what exactly is occurring during and after a program and to contribute to more accurate statements about programmatic efficacy.

Limitations associated with the study sample are notable. This study was conducted in a university setting with university employees. A sample of facilitators not working in a university setting might answer questions on the FCPC differently than facilitators in this study. Although the significant sphericity test indicated that the sample ($N=121$) was large enough, a more racially diverse and larger overall sample may yield alternative results. Moreover, the use of the FCPC outside of college-aged facilitators could highlight the alternative approaches to program implementation

The FCPC was not designed to measure specific programmatic components (e.g., "I delivered the program debrief.") rather it was designed to measure fidelity beliefs and fidelity at a more global level (e.g., "I delivered the program as designed."). This was intentional on the research team's part so the measure could be "plugged into" multiple program types, thus navigating an issue of fidelity measures often being overly specific and program dependent (Hansen, 2014).

Another strength of this study relates to staff experience and training. If facilitators score low on the FCPC items, their score can signal to program administrators that additional fidelity training is needed which, when provided, may result in the delivery of a better program. Thus the FCPC serves the dual purpose of both program and personnel assessment, improvement, and quality management.

The FCPC is designed to assess one layer of the four presented earlier, as such it provides a limited perspective on implementation quality. Further investigation of the organizational, community, and program characteristics will help to highlight potential challenges to implementation. The combination of the FCPC with these yet-to-be developed tools will provide a more holistic and complete perspective on the importance of implementation in the context of outcome achievement and perhaps more importantly contribute to better outcomes for participants.

References

- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*(1), 23–33.
- Bishop, D. C., Pankratz, M. M., Hansen, W. B., Albritton, J., Albritton, L., & Strack, J. (2013). Measuring fidelity and adaptation: Reliability of an instrument for school-based prevention programs. *Evaluation & the Health Professions, 37*(2) 1-27.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guilford Press.
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Psychology Press: Taylor & Francis Group.
- Caldwell, L. L., Younker, A. S., Wegner, L. E., Patrick, M., Vergnani, T., Smith, E. A., & Flisher, A. J. (2008). Understanding leisure-related program effects by using process data in the HealthWise South Africa project. *Journal of Park & Recreation Administration, 26*(2), 146-162.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, Article No. 40.
- Castro, F. G., Barrera, M., & Martinez, C. R. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*(1), 41–45.
- Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-236). London: Lawrence Erlbaum Associates.
- Dane, A., & Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 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*(3-4), 327–350.
- Dusenbury, L., Brannigan, M., Falco, R., & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256.
- Dusenbury, L., Brannigan, R., Hansen, W., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Journal of Alcohol and Drug Education, 20*(3), 308-313.
- Ewert, A. W. (1983). *Outdoor adventure and self-concept: A research analysis*. Eugene, OR: Center of Leisure Studies, University of Oregon.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.
- Gagnon, R. J. (2014). Exploring the relationship between the facilitator and fidelity. *Journal of Outdoor Recreation, Education, and Leadership, 6*(2), 183-186.
- Gagnon, R. J., & Bumpus, M. F. (In Press). Fidelity and its importance to experiential and outdoor education. *Journal of Outdoor Recreation, Education, and Leadership*.
- Gagnon, R. J., Franz, N., Garst, B. A., & Bumpus M. F. (2015). Factors impacting program delivery: The importance of implementation research in Extension. *Journal of Human Sciences and Extension, 3*(2), 68-82.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and pre-referral intervention. *School Psychology Review, 18*(1), 37–50.
- Gresham, F. M., MacMillan, D., Beebe-Frankenberger, M. E. & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice, 15*(4), 198-205.
- Hansen, W. B. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 335–359). New York, NY: Springer.
- Sanetti, L. M., & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol. *School Psychology Quarterly, 24*(1), 24–35.
- Hogue, A., Liddle, H. A., Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy 33*(2), 332–345.
- Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). Responsiveness to intervention (RTI): How to do it. Lawrence, KS: National Research Center on Learning Disabilities.
- Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*(1), 55– 63.
- Kreeft, P., Jongbloet, J., & Van Havare, T. (2014). Factors affecting implementation: Cultural adaptation and training. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 315–334). New York, NY: Springer.

- Little, M. A., Sussman, S., Sun, P., & Rohrbach, L. A. (2013). The effects of implementation fidelity in the Towards No Drug Abuse dissemination trial. *Health Education, 113*(4), 281–296.
- McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62*(4), 670-678.
- Mainieri, T. L., & Anderson, D. M. (2014). Exploring the “black box” of programming: Applying systematic implementation evaluation to a structured camp curriculum. *Journal of Experiential Education, 38*(2), 1–18.
- Mowbray, C., Holter, M., Teague, G., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315-340.
- Nobel, O., Zbylut, M., Fuchs, D., Campbell, K., Brazil, D., & Morrison, E. (2006). Leader experience and the identification of challenges in a stability and support operation. *United States Army Research Institute for the Behavioral and Social Sciences, 1-38*.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445–459.
- Spoth, R., Gyll, M., Trudeau, L., & Gold-Lillehoj, C. (2002). Two studies of proximal outcomes and implementation quality of universal preventive interventions in a community–university collaboration context. *Journal of community psychology, 30*(5), 499-518.
- Stein, M., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., Fuchs, L., & Compton, D. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis, 30*(4), 368-388.
- Tucker, A. R., & Rheingold, A. (2010). Enhancing fidelity in adventure education and adventure therapy. *Journal of Experiential Education, 33*(3), 258-273.