


An Investigation of Concurrent Validity of Fidelity of Implementation Measures at Initial Years of Implementation

Remedial and Special Education
2019, Vol. 40(1) 25–31
© Hammill Institute on Disabilities 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0741932518795639
rase.sagepub.com


Caitlin Rasplia Khoury, PhD, BCBA¹, Kent McIntosh, PhD¹,
and Robert Hoselton, BS¹

Abstract

Fidelity of implementation of school practices is crucial to student outcomes. Several types of tools, including self-assessments, are available for measuring fidelity, but little is known regarding the relation of self-assessments of fidelity to fidelity instruments completed with the support of external experts, specifically, during the first few years of implementation. The present study used cross-sectional data from 1,438 schools to examine relations between fidelity self-assessment and team-based fidelity measures in the first 4 years of implementation of School-Wide Positive Behavioral Interventions and Supports (SWPBIS). Results showed strong positive correlations between fidelity self-assessments and a team-based measure of fidelity at each year of implementation.

Keywords

positive behavior supports, fidelity of implementation, School-Wide Positive Behavioral Interventions and Supports, self-assessment

Educational settings host a variety of interventions to optimize academic and social success for all children. Schools are charged with the difficult task of promoting student achievement so that students reach their academic and social potential through use of evidence-based practices (EBPs). Implementation of these practices represents the connection between science and practice. Regardless of the intervention that is chosen for implementation within a school setting, the quality, or fidelity, of implementation of these practices is crucial to student outcomes (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005).

Fidelity of implementation, also known as treatment integrity, refers to the extent to which the core features of an intervention are delivered as intended (Noell, Gresham, & Gansle, 2002). The critical nature of implementation fidelity is supported by a strong research base, which shows that higher treatment integrity results in better treatment outcomes (Flannery, Fenning, Kato, & McIntosh, 2014; Noell et al., 2002; Noell, Volz, Henderson, & Williams, 2017; Sanetti & Kratochwill, 2008).

Stages of Implementation

Implementation is commonly described not as an event, but rather a process, which can take 2 to 4 years to complete in many organizations (Fixsen et al., 2005). Fixsen and colleagues (2005) identified this process as occurring in stages,

including exploration and adoption (Stage 1), program installation (Stage 2), initial implementation (Stage 3), full operation (Stage 4), innovation (Stage 5), and sustainability (Stage 6). The implementation stages represent highly integrated components that identify a trajectory toward sustainability of an intervention. The stages are representative of an order that begins first with identifying potential fit and barriers to implementation and gradually reaching initial and then full implementation. After a program has been accepted into the organization, the implementation stages shift focus to address the sustainability of that intervention.

The Current State of Fidelity of Implementation Research in Schools

Although this research base regarding fidelity of implementation exists, it has not necessarily been a priority within school settings. Previous school-based research has focused more on identifying EBP in schools (McIntosh, Martinez, Ty, & McClain, 2013), and it has sometimes been assumed that adopting a practice is evidence of adequate fidelity of

¹University of Oregon, Eugene, USA

Corresponding Author:

Caitlin Rasplia Khoury, 5208 University of Oregon, Eugene, Oregon, 97403, USA.

Email: ckhoury@childrens-clinic.com

implementation (Fixsen et al., 2005). In more recent years, however, the measurement of fidelity has become a more prominent topic in areas of school-based research (e.g., Sanetti & Kratochwill, 2008; Sheridan, Swanger-Gagne, Welch, Kwon, & Garbacz, 2009). Appropriate measurement tools are seen as highly beneficial, as they help to ensure adequate fidelity (Sanetti & Kratochwill, 2008). In addition to measuring fidelity of implementation for research purposes, fidelity measures allow for staff members to take action in planning for full and sustained implementation of the core features of interventions (Mathews, McIntosh, Frank, & May, 2013).

Measuring Fidelity of Implementation

Because fidelity of implementation is important to both researchers and practitioners, there are several tools available that are intended to assess fidelity and aid school staff in the implementation process. These tools range from self-assessments to external measures of fidelity, all of which have relative strengths and weaknesses.

External Measures of Fidelity

External measures of fidelity, such as direct observation and surveys completed by an outside assessor, are often considered to be the “gold standard,” as they are considered to be less susceptible to bias than self-assessments. External measures may include interviews (i.e., guided self-assessments), direct observation, or review of permanent products. Direct observation often results in objective, accurate measurement of fidelity, but it can be resource-intensive to implement (Gresham, 1989). Permanent products represent another means for measuring fidelity of implementation; however, not all interventions result in permanent products (Noell et al., 2017).

Fidelity Self-Assessments

Self-report measures are another widely used tool for assessing fidelity of implementation. Fidelity self-assessments are highly efficient, as they require that the individual implementing the intervention (e.g., teacher or other school staff member) indicate through an interview or questionnaire the extent to which they implemented the intervention according to a set criterion. However, they hold limitations, and some research has indicated that fidelity self-assessments are a less reliable method of treatment integrity (Fiske, 2008; Noell, 2007). For example, Noell et al. (2005) noted that teachers’ self-report of treatment integrity was weakly correlated with direct observation of treatment integrity; therefore, fidelity self-assessments may be inaccurate, with a possibility of inflated scores related to actual implementation. Other data suggest that self-assessment measures can be, in fact, reliable over time (Kozial & Burns,

1986). These mixed findings call for a need to examine under which conditions the results of self-assessment can be accurate. One conceptualization of accuracy is concurrent validity, or the extent to which they are comparable to results from similar measures.

Comparing the results of various methods of fidelity of implementation can provide valuable information in regard to the level of implementation and school personnel knowledge of core intervention components (McIntosh, Mercer, et al., 2013). A few studies have examined correlations between external measures and fidelity self-assessments validity of fidelity measures for school-wide positive behavioral interventions and supports (SWPBIS; Sugai & Horner, 2009), a systems-level framework for implementing and assessing effects of behavioral interventions in schools. In contrast to the classroom fidelity studies described above, these studies have found strong concurrent validity estimates (i.e., $r_s > .50$) between external and self-assessments (Horner et al., 2004; McIntosh, Mercer, Nese, Strickland-Cohen, & Hoselton, in press; Vincent, Spaulding, & Tobin, 2010). However, these studies did not assess the extent to which each school’s year of implementation (a common proxy for stage of implementation; Fixsen et al., 2005) may affect concurrent validity.

There are important reasons why stage of implementation could influence accuracy of self-assessment of fidelity of implementation, and this accuracy could affect fidelity of implementation itself. For example, schools in their first year of implementation (i.e., initial implementation) may be less aware of and knowledgeable about the critical features, leading to overrating of their implementation. Thus, self-assessments may have lower correlations with other measures compared with schools that have been implementing for a longer amount of time. Such a pattern would limit their utility (i.e., yielding accuracy so low as to be invalid for decision making) at a critically important point when schools are most at risk for abandonment (Nese et al., 2016). As a result, school teams might overestimate their implementation, leading to assuming that they are implementing SWPBIS when in fact they are not. In a different scenario that could lead to similar patterns of correlations, a team implementing for over 5 years (i.e., full implementation or sustainability phases) might overestimate their implementation because they are accustomed to rating themselves highly and have not noticed that their implementation has drifted away from tight adherence to critical features (McIntosh, Mercer, Nese, & Ghemraoui, 2016). Hence, examining these correlations by year can be useful to detect any systematic differences across stage of implementation.

SWPBIS

SWPBIS is a systematic approach to proactively improve school climate and prevent student problem behaviors

across all school settings. Based on applied behavior analysis and implementation science, it includes implementation of instructional practices and strategies (e.g., defining, teaching, and acknowledging desired social behaviors) based on student need and teaming and data systems to enhance implementation and student outcomes (Lewis, McIntosh, Simonsen, Mitchell, & Hatton, 2017). SWPBIS was chosen as a focus of the present study because it is a widely used EBP (Sugai & Horner, 2009), and specifically aligns with Fixsen et al. (2005) implementation stages, as it is a systems-level framework that can take years to fully implement.

Several measures are available to assess fidelity of implementation of SWPBIS, including those that are intended to be completed by a school team with the assistance of an external coach (e.g., *School-Wide Benchmarks of Quality* [BoQ]), tools intended to be completed by school teams independent of a coach (e.g., *Team Implementation Checklist* [TIC], *PBIS Self-Assessment Survey* [SAS]), and external evaluations (e.g., *School-Wide Evaluation Tool* [SET]). The specific measures used within the present study are described in detail.

Purpose of the Study

Although it is critical to measure fidelity of implementation, little is known regarding the utility of self-assessments intended to be completed independent of an external coach and their relations with other measures of fidelity intended to be completed with the assistance of an external coach or expert. Implementation is dynamic and can rapidly change over the first few years of implementation. Therefore, it is important that self-assessment data, independent of an external coach, allow teams to improve their efforts, and such information is useful only to the extent that it is related consistently to other measures of fidelity. Existing SWPBIS research indicates that (a) self-assessments, particularly from novice users, are prone to inaccurate responding, and (b) self-assessments independent of an external coach of SWPBIS fidelity of implementation are moderately related to other fidelity measures involving an external coach or external evaluation team, but the years of implementation of these schools were not reported. To date, we are unaware of research examining the relation of fidelity of implementation data by year of implementation.

The purpose of the present study was to assess the concurrent validity between fidelity measures independent of an external coach and a measure of fidelity with the aid of an external coach and determine the relation of such ratings within the first 4 years of implementation of a school practice, SWPBIS. The following research question was examined:

Research Question 1: To what extent are fidelity ratings completed with the assistance of an external coach and

those without assistance of an external coach or expert associated, and does the association systematically vary by year of implementation of SWPBIS?

Method

Participants and Settings

Participants were school and external personnel (e.g., district coaches) in 1,438 schools who rated their school's implementation of SWPBIS during their first 4 years of implementation (i.e., the transition from initial to full implementation and sustainability stages). District coaches typically support multiple schools in implementation of SWPBIS and may not meet regularly with the school-based team.

Due to varying use of fidelity measures (with many schools completing only one measure per year), about half ($n = 736$) of these schools' data were represented at only one time point across 4 years, and the other schools had data across 2 years ($n = 549$), 3 years ($n = 150$), or 4 years ($n = 1$). Schools were located within Illinois, Indiana, Maryland, Minnesota, Missouri, and Wisconsin. School demographic data were available from the National Center on Educational Statistics (NCES) for 1,355 schools (94.2% of the sample). Of these, 68.9% were elementary schools, 19.2% were middle schools, 9.2% were high schools, and 2.6% were other (e.g., K–8). Table 1 provides additional school characteristics.

Measures

Three different measures of fidelity of implementation of SWPBIS were used in the present study, including the PBIS SAS (Sugai, Horner, & Todd, 2000), TIC (Sugai, Horner, & Lewis-Palmer, 2001), and BoQ (Kincaid, Childs, & George, 2005). The SAS and TIC are fidelity self-assessments of SWPBIS implementation completed without the aid of an external coach, and the BoQ is intended to be a collaborative measure of fidelity involving the external coach's perspective as well as internal team member perspectives.

SWPBIS SAS. The SAS (Sugai et al., 2000) is a 43-item survey completed by school staff to self-report fidelity of implementation of SWPBIS in four different settings within their school. It contains four components: (a) School-Wide Systems, which are across all settings in the school; (b) Non-Classroom Systems, such as playgrounds and hallways; (c) Classroom Systems, which are in classrooms; and (d) Individual Systems, which examine support for individual students. For the purposes of this study, only the School-Wide Systems scale was used in analyses. Each respondent individually self-reports the implementation status of each critical feature by rating whether the critical feature is in place, partially in place, or not in place. Each

Table 1. Descriptive Statistics for Sample.

Characteristic	M or % (SD)
Student enrollment	517.3 (403.57)
% ethnic minority students	41.6% (30.84)
% FRL	46.2% (27.08)
Grade levels	
% elementary schools	68.9%
% middle schools	19.2%
% high schools	9.2%
% other (e.g., K–8)	2.6%

Note. $N = 1,355$. School demographic data obtained from National Center for Education Statistics for 94.2% of schools. FRL = free and/or reduced lunches.

item on the SAS is scored on a 0 to 2 scale (0 = *not in place*, 1 = *partially in place*, and 2 = *in place*). The total score is calculated by summing the number of answers based on their current status (0–2) and dividing by the total number of respondents. The SAS has strong internal consistency for the subscale and total scores of the measure, with α ranging from .85 to .94 (Hagan-Burke et al., 2005; Safran, 2006). The SAS also has strong correlations with external measures of implementation fidelity, including the SET ($r = .75$, $p < .001$; Horner et al., 2004).

TIC. The TIC (Sugai et al., 2001) is a 22-item fidelity self-assessment that is completed by the SWPBIS team via a consensus rating of each item. The TIC is used to monitor progress on implementation of specific items of the SWPBIS framework, and can be used on a monthly, every other month, or quarterly basis. Schools are encouraged to complete the TIC during the initial years of SWPBIS implementation (Horner, Sugai, & Lewis-Palmer, 2005). Each item on the TIC is scored on a 3-point scale, with 2 = *achieved*, 1 = *in progress*, and 0 = *not yet started*, for a total maximum score of 44. The TIC shows moderate correlation with the SET, an external measure of fidelity, within high schools (correlations range from $r = .32$ to $.75$; Vincent et al., 2010).

BoQ. The BoQ (Kincaid et al., 2005) is a 53-item measure of fidelity of SWPBIS implementation. The BoQ consists of 10 different subscales: (a) SWPBIS Team, (b) Faculty Commitment, (c) Effective Discipline Procedures, (d) Data Entry, (e) Expectations and Rules, (f) Reward Systems, (g) Lesson Plans, (h) Implementation Plans, (i) Crisis Plans, and (j) Evaluation. Scoring varies based on individual items, with a total possible score of 107. The BoQ is intended to be completed by an external SWPBIS coach based on their knowledge of the school, with use of the BoQ scoring guide and input on scoring from the school team. More specifically, the external coach will complete the BoQ and request that each individual team member rate each

component as either “In Place,” “Needs Improvement,” or “Not in Place.” The coach will then review these team member ratings to generate a team summary, as well as identify any discrepancies between the coach’s evaluation and the team evaluation for discussion. The coach then calculates a final score. The BoQ has strong 1-week test–retest reliability ($r = .94$), 2-week interrater reliability ($r = .87$), and concurrent validity (correlation with the SET, $r = .51$), as reported by R. Cohen, Kincaid, and Childs (2007). Internal consistency for the sample was $\alpha = .91$, indicating high internal consistency.

Procedure

To build the extant dataset used in analyses, the authors used training records from six state SWPBIS initiatives. The records included information about the school (e.g., the school’s NCES number) and the year in which initial SWPBIS training took place. Each school’s first year of implementation (provided by state SWPBIS training databases) ranged from the 2001–2002 academic year to the 2012–2013 academic year. From these records, the authors obtained school demographic data from NCES and fidelity of implementation data from an extant database from the Educational and Community Supports research unit at the University of Oregon (2014), PBIS Assessment (<https://www.pbisapps.org>). PBIS Assessment is a free website into which school personnel and external coaches or coordinators can enter scores from the fidelity measures described above. School and district teams can then access results for action planning purposes. Data extracted for the study included all available SAS, TIC, and BoQ scores for each school’s first 4 years of implementation of SWPBIS. When schools reported multiple scores during the year, only the last score reported was used for analyses.

Analyses

Separate cross-sectional Pearson correlations of the BoQ and TIC, as well as the BoQ and SAS, were completed for each year to evaluate reliability and the extent to which the relation between self-assessment fidelity data, the SAS or TIC, and a team-based fidelity measure, the BoQ, systematically varied by year of implementation. Strength of correlations was determined through Cohen’s (1988) criteria of .1 = small, .3 = medium, and .5 = large. These analyses were selected as to provide a general indicator of relations between measures, a preliminary step in assessing characteristics related to fidelity in schools.

Results

As seen in the descriptive statistics in Table 2, the average implementation fidelity scores for all measures increased from the first ($M = .61$ to $.73$) to fourth ($M = .78$ to $.83$)

Table 2. Means, Standard Deviations, and Sample Sizes for Fidelity Measures by Year of Implementation.

Year	BoQ			SAS			TIC		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
1	78	.73	.18	78	.73	.14	47	.61	.21
2	412	.74	.17	412	.75	.14	317	.72	.17
3	907	.80	.13	907	.78	.11	555	.72	.17
4	896	.83	.13	896	.80	.11	366	.78	.15

Note. BoQ = School-Wide Benchmarks of Quality; SAS = Self-Assessment Survey; TIC = Team Implementation Checklist.

year of implementation, indicating increases in fidelity scores related to year of implementation. The correlations between each fidelity score independent of an external coach and those with the assistance of an external coach were consistent and above Cohen's criteria for large effects across all 4 years, ranging from .49 to .77 (see Table 3). Overall, there were strong, statistically significant ($p < .01$) positive correlations between the fidelity self-assessments independent of an external coach and the measures with the assistance of an external coach across all 4 years of implementation. Differences in correlations from year to year were minor and did not fit any consistent pattern of increasing or decreasing relation.

Discussion

The present study examined the extent of agreement between fidelity self-report and a measure of fidelity completed with the assistance of an external coach or expert over the first 4 years of implementation of SWPBIS. In a cross-sectional sample, three measures of fidelity of implementation were compared in the first 4 years of implementation. Results indicated that there was a moderately strong positive correlation between fidelity self-assessments and when an external coach or expert provides assistance in obtaining measures of fidelity across all 4 years of implementation examined.

Although external measures of fidelity of implementation are considered to be more reliable than self-report ratings (Lillejoj, Griffin, & Spoth, 2004), it is important for schools to measure implementation reliably without depending upon an external source (Domitrovich et al., 2008). These data provide initial, tentative support to the hypothesis that scores from these specific self-assessment measures are consistent with other fidelity measures where assistance from an external coach is provided. However, this preliminary work relied on correlations from an extant, cross-sectional sample that lacked external evaluations, and additional research is needed and encouraged in this area. For example, too few schools in our sample used the SET, a

Table 3. Pearson Correlations Between BoQ and SAS or TIC Across Implementation Year.

Year	Measure	
	SAS and BoQ	TIC and BoQ
1	.77 ($n = 78$)	.63 ($n = 47$)
2	.68 ($n = 412$)	.56 ($n = 317$)
3	.61 ($n = 907$)	.49 ($n = 555$)
4	.62 ($n = 896$)	.58 ($n = 366$)

Note. For each year, the n appears in parentheses. SAS = Self-Assessment Survey; BoQ = School-Wide Benchmarks of Quality; TIC = Team Implementation Checklist.

All values statistically significant at $p < .01$.

research tool completed by a trained evaluator, which is a more external measure of fidelity than the BoQ. The SET can be used in future research specifically addressing concurrent validity of self-assessment and external measures, rather than validity across measures over time.

Results also support work in previous studies of implementation that fidelity of implementation itself increases over time when it is monitored (Bradshaw, Reinke, Bevans, Brown, & Leaf, 2009), as all measures of fidelity represented showed an increase in the average implementation fidelity score from Year 1 to Year 4 of implementation. Results did not show that accuracy in self-assessment data improves over time, as the correlation among measures showed little change. Instead, correlations remained strong throughout. In addition, on average, practitioners self-rated their own fidelity of implementation slightly lower than the collaborative measure completed by internal team members and an external coach, in contrast to results seen in classroom management fidelity studies (Noell, 2007).

Based on these preliminary data, one potential reason for these findings is that the specificity of items on fidelity measures are critical (Schoenwald et al., 2011) and can greatly affect the fidelity ratings. It may be that certain qualities of fidelity measures make for more accurate measurement. For example, a tool that assesses clear, operationally defined behaviors is more likely to yield accurate results than tools that contain more subjective areas of measurement that raters are less likely to complete in the same manner. The measures included in the present study include objective, observable items, which are easier to rate more accurately and may explain the high correlations found across all 4 years. For example, one item on the SAS states, "A behavior support team responds promptly (within 2 working days) to students who present chronic problem behaviors." Items in other fidelity self-assessment instruments, such as those that assess the quality of student and teacher interactions or frequency of common instructional behaviors (which are easy to overestimate), may be more subjective.

Limitations

There are several limitations to the current study. First, the data are cross-sectional rather than longitudinal. The same schools were not represented across all 4 years of implementation. Similarly, because data were extracted from a larger database, the types of fidelity monitoring tools that were used from year to year were not controlled. This aspect of the dataset explains the variability in the number of schools that were included across all 4 years, as many schools used only one fidelity measure or alternated between measures across years of implementation. In addition, because data were extracted from an extant database, the intent for use of each measure can only be identified, as data were not collected on actual use of each measure. As a result, the BoQ may not have been completed with an external coach and may have been completed more as a different fidelity self-assessment. The extent to which the same team completed both measures in the same way is unknown. Finally, the number of years implementing an intervention is not necessarily an adequate proxy for stage of implementation (Fixsen et al., 2005). It also describes the need for additional research using measures more closely aligned with these stages, such as the Implementation Phases Inventory (Bradshaw, Debnam, Koth, & Leaf, 2009).

Implications for Research and Practice

There are many implications for both research and practice. Results of the present study suggest that even in initial implementation, self-assessment measures of fidelity are a potentially viable tool. Previous research indicates that schools show stronger fidelity of implementation when formal training and coaching occurs rather than when schools implement SWPBIS without training (Bradshaw, Reinke, et al., 2009; Sprague, Biglan, Rusby, Gau, & Vincent, 2017). However, with less objective measures or in schools that do not receive additional support, it is important to consider how self-assessment scores can be used effectively. For example, it is possible that fidelity self-assessment scores are better used for team action planning, such as identifying next steps in the implementation process, than as a tool to provide an accurate benchmark of level of implementation.

Fidelity of implementation has been an interest of researchers for a significantly longer amount of time than it has for practitioner use. However, additional research is needed on the specific variables that affect implementation such as coaching or initial training. It is critical to specifically identify the possible effect that coaching support has on accurately interpreting fidelity of implementation, as it is considered to be an essential component of implementation (Fixsen et al., 2005). Researchers have documented ways in which coaching can be used within school settings (e.g., Stormont & Reinke, 2012) as a universal support and as a

Tier 2 support. Research has also suggested that coaching involves several components, such as teaching, modeling, and performance feedback (Fixsen et al., 2005; Stormont & Reinke, 2012), all of which may have varying effects on fidelity of implementation.

Although use of SWPBIS fidelity measures are prescribed, the level at which measures were used as prescribed could not be controlled within the present study and, therefore, represents a limitation of this study. Additional research is needed to understand these variations in use of instruments and the impact of such decisions on implementation.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A120278 to University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Bradshaw, C. P., Debnam, K., Koth, C. W., & Leaf, P. J. (2009). Preliminary validation of the Implementation Phases Inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions, 11*, 145–160.
- Bradshaw, C. P., Reinke, W. M., Bevens, K. B., Brown, L. D., & Leaf, P. J. (2009). Implementation of Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*, 1–26.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring schoolwide positive behavior support implementation: Development and validation of the Benchmarks of Quality. *Journal of Positive Behavior Interventions, 9*, 203–213.
- Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., . . . Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion, 1*, 6–28.
- Fiske, K. (2008). Treatment integrity of school-based behavior analytic interventions: A review of the research. *Behavior Analysis in Practice, 1*, 19–25.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: Synthesis of the literature* (FMHI Publication #231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Flannery, K. B., Fenning, P., Kato, M. M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and

- supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, 29, 111–124. doi:10.1037/spq0000039
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37–50.
- Hagan-Burke, S., Burke, M. D., Martin, E., Boon, R. T., Fore, C., III., & Kirkendoll, D. (2005). The internal consistency of the school-wide subscales of the effective behavioral support survey. *Education & Treatment of Children*, 28, 400–413.
- Horner, R. H., Sugai, G., & Lewis-Palmer, T. L. (2005). *School-wide Positive Behavior Support evaluation template*. Eugene, OR: Educational and Community Supports.
- Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, 6, 3–12.
- Kincaid, D., Childs, K., & George, H. P. (2005). *Schoolwide benchmarks of quality*. Unpublished instrument, University of South Florida, Tampa.
- Kozial, S., & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research*, 79, 205–209.
- Lewis, T. J., McIntosh, K., Simonsen, B., Mitchell, B. S., & Hatton, H. L. (2017). School-wide systems of positive behavior support: Implications for students at-risk and with emotional/behavioral disorders. *AERA Open*, 3(2), 1–11.
- Lillejoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior*, 31, 242–257.
- Mathews, S., McIntosh, K., Frank, J. L., & May, S. L. (2013). Critical features predicting sustained implementation of school-wide positive behavior support. *Journal of Positive Behavior Interventions*, 16, 168–178.
- McIntosh, K., Martinez, R. S., Ty, S. V., & McClain, M. B. (2013). Scientific research in school psychology: Leading researchers weigh in on its past, present, and future. *Journal of School Psychology*, 51, 267–318. doi:10.1016/j.jsp.2013.04.003
- McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to sustained implementation of schoolwide positive behavior support. *Exceptional Children*, 79, 293–311.
- McIntosh, K., Mercer, S. H., Nese, R. N. T., & Ghemraoui, A. (2016). Patterns of implementation of a scaled-up school-based prevention model over a five-year period. *Prevention Science*, 17, 992–1001.
- McIntosh, K., Mercer, S., Nese, R., Strickland-Cohen, K., Kittelman, A., Hoselton, R., & Horner, R., (2018). Factors predicting sustained implementation of a universal behavior support framework. *Educational Researcher*, 47, 307–316.
- Nese, R. N. T., McIntosh, K., Nese, J. F. T., Bloom, J., Johnson, N. W., Phillips, D., . . . Hoselton, R. (2016). Predicting abandonment of school-wide behavior interventions. *Behavioral Disorders*, 42, 261–270.
- Noell, G. H. (2007). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). Mahwah, NJ: Lawrence Erlbaum.
- Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education*, 11, 51–67.
- Noell, G. H., Volz, J. R., Henderson, M. Y., & Williams, K. L. (2017). Evaluating an integrated support model for increasing treatment plan implementation following consultation in schools. *School Psychology Quarterly*, 32, 525–538.
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, 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.
- Safran, S. P. (2006). Using the Effective Behavior Supports Survey to guide development of schoolwide positive behavior support. *Journal of Positive Behavior Interventions*, 8, 3–9.
- Sanetti, L. M., & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy*, 4, 95–114.
- Schoenwald, S., Garland, A., Chapman, J., Frazier, S., SHEidow, A., & Southam-Gerow, M. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health*, 38, 32–43.
- Sheridan, S. M., Swanger-Gagne, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review*, 38, 476–495.
- Sprague, J. R., Biglan, A., Rusby, J. C., Gau, J. M., & Vincent, C. G. (2017). Implementing school wide PBIS in middle schools: Results of a randomized trial. *Journal of Health Science & Education*, 1(2), 1–10.
- Stormont, M., & Reinke, W. M. (2012). Using coaching to support behavioral interventions and support systems. *Beyond Behavior*, 21(2), 11–19.
- Sugai, G., & Horner, R. H. (2009). Defining and describing school-wide positive behavior support. In W. Sailor, G. Dunlap, G. Sugai, & R. H. Horner (Eds.), *Handbook of positive behavior support* (pp. 307–326). New York, NY: Springer.
- Sugai, G., Horner, R. H., & Lewis-Palmer, T. L. (2001). *Team Implementation Checklist (TIC)*. Eugene, OR: Educational and Community Supports. Available from <http://www.pbis.org>
- Sugai, G., Horner, R. H., & Todd, A. W. (2000). *PBIS Self-Assessment Survey 2.0*. Eugene, OR: Educational and Community Supports. Available from <http://www.pbisapps.org>
- University of Oregon. (2014). *Dataset 0055* [Database and codebook]. Eugene: Author.
- Vincent, C. G., Spaulding, S. A., & Tobin, T. J. (2010). A reexamination of the psychometric properties of the School-Wide Evaluation Tool (SET). *Journal of Positive Behavior Interventions*, 12, 161–179.