Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2017, Vol. 19(1) 3–13

DOI: 10.1177/1098300716637193
jpbi.sagepub.com
SAGE

Technical Adequacy of the SWPBIS Tiered Fidelity Inventory

Kent McIntosh, PhD[1], Michelle M. Massar, MEd[1], Robert F. Algozzine, PhD[2], Heather Peshak George, PhD[3], Robert H. Horner, PhD[1], Timothy J. Lewis, PhD[4], and Jessica Swain-Bradway, PhD[5]

Abstract

Full and durable implementation of school-based interventions is supported by regular evaluation of fidelity of implementation. Multiple assessments have been developed to evaluate the extent to which schools are applying the core features of school-wide positive behavioral interventions and supports (SWPBIS). The *SWPBIS Tiered Fidelity Inventory* (TFI) was developed to be used as an initial assessment to determine the extent to which a school is using (or needs) SWPBIS, a measure of SWPBIS fidelity of implementation at all three tiers of support, and a tool to guide action planning for further implementation efforts. In this research, we evaluated the psychometric properties of the TFI in three studies: a content validity study, a usability and reliability study, and a large-scale validation study. Results showed strong construct validity for assessing fidelity at all three tiers, strong interrater and 2-week test–retest reliability, high usability for action planning, and strong relations with existing SWPBIS fidelity measures. Implications for accurate evaluation planning are discussed.

Keywords

positive behavior support, school-wide positive behavioral interventions and supports, treatment integrity, fidelity of implementation, systems change

Schools across the country are facing the demand to provide rigorous educational opportunities to a highly diverse population of learners requiring various levels of academic and behavior support. The most recent reauthorization of the Individuals With Disabilities Education Act (2004) provided the impetus for an increased focus on empirically supported practices. However, simply electing to adopt evidence-based practices without attending to the implementation process is unlikely to improve outcomes (Fixsen, Blase, Duda, Naoom, & Van Dyke, 2010). Implementation abandonment, wherein schools discontinue the use of effectively implemented practices in place of new ones each year, is commonplace in schools across the country (Adelman & Taylor, 2003). This phenomenon carries costs with regard to system resources, including financial losses and reduced staff buy-in, as well as student outcomes (McIntosh et al., 2013). Empirical research shows that assessing fidelity and using those data to inform action planning can increase sustainability and decrease the likelihood of abandoning effective practices (McIntosh, Kim, Mercer, Strickland-Cohen, & Horner, 2015).

One effective and widely implemented practice is school-wide positive behavioral interventions and supports (SWPBIS; Sugai & Horner, 2009), a three-tiered framework that promotes the use of positive and preventive approaches to behavior support at a systems level. More than 21,000 schools in the United States have adopted SWPBIS in efforts to establish positive, safe, predictable, and consistent school climates (Horner, 2014). Research indicates that high fidelity of implementation of SWPBIS is associated with improved student and teacher outcomes, including an increase in student perception of school safety, a reduction in number of office discipline referrals (ODRs), a decrease in student use of school counseling services, growth in academic achievement, and an increase in teacher self-efficacy (Bradshaw, Mitchell, & Leaf, 2010; Horner et al., 2009; Kelm & McIntosh, 2012; McIntosh, Bennett, & Price, 2011; Nelson, Martella, & Marchand-Martella, 2002; Ross, Romer, &

[1]University of Oregon, Eugene, USA
[2]University of North Carolina at Charlotte, USA
[3]University of South Florida, Tampa, USA
[4]University of Missouri, Columbia, USA
[5]Midwest PBIS Network, Lombard, IL, USA

Corresponding Author:
Kent McIntosh, University of Oregon, 1235 University of Oregon, Eugene, OR 97403, USA.
Email: kentm@uoregon.edu

Action Editor: Daniel Maggin

Horner, 2012). Flannery, Fenning, Kato, and McIntosh (2014) found that SWPBIS reduced the level of problem behavior in high school students, and the level of reduction was significantly related to fidelity of implementation, as schools with higher fidelity had decreased rates of problem behavior.

Measuring Fidelity of Implementation of SWPBIS

One of the defining activities of SWPBIS is the use of data for decision making (Algozzine et al., 2010). Data are used to guide both decisions focused on improving student supports and decisions about how best to implement SWPBIS features. For schools to implement SWPBIS successfully, ongoing evaluation of fidelity of implementation and informed action planning based on data are essential. Fidelity of implementation is defined as the extent to which a program, intervention, framework, or practice, "as conceptualized in a theoretical model or manual, is implemented as intended" (Schulte, Easton, & Parker, 2009, p. 460). Although the importance of fidelity is not a new concept in educational research (O'Donnell, 2008), school-based assessment of implementation has recently become the subject of increased focus. The trend of assessing fidelity of school systems is reflected in the rapid increase in the number of assessment tools available for evaluating the core components of SWPBIS implementation. These fidelity measures include (a) the *School-Wide Evaluation Tool* (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001), (b) the *School-Wide Benchmarks of Quality* (BoQ; Kincaid, Childs, & George, 2005), (c) the *Team Implementation Checklist* (TIC; Sugai, Horner, & Lewis-Palmer, 2001), (d) the *PBIS Self-Assessment Survey* (SAS; Sugai, Horner, & Todd, 2000), (e) the *Benchmarks for Advanced Tiers* (BAT; Anderson et al., 2012), (f) the *Individual Student Systems Evaluation Tool* (ISSET; Lewis-Palmer, Todd, Horner, Sugai, & Sampson, 2003), and (g) the *Monitoring Advanced Tiers Tool* (MATT; Horner, Sampson, Anderson, Todd, & Eliason, 2013). Collectively, these measures assess implementation at each of the three tiers of SWPBIS, but there has not been a single measure that can be used to assess fidelity of implementation of all three tiers on the same scale, which has presented challenges for evaluation across schools at the district, regional, or state level.

Tiered Fidelity Inventory (TFI)

The *SWPBIS TFI* (Algozzine et al., 2014) was developed to be a comprehensive fidelity of implementation tool to be used alone or in conjunction with other SWPBIS assessments. Although the existing fidelity measures can be used to assess fidelity of implementation of SWPBIS, there was no single tool that school teams could use to measure initial implementation, develop an action plan, and monitor implementation progress across all three tiers. The TFI was designed to be a more comprehensive and efficient measure of fidelity, with a common format, scale, and language to assess each tier, for schools at any level of implementation. The TFI is intended as (a) an initial assessment to determine whether a school is using (or needs) SWPBIS, (b) a guide for implementation of Tier I, Tier II, and Tier III practices, or (c) an index of sustained SWPBIS implementation. The TFI was compiled from existing SWPBIS fidelity measures and unpublished fidelity measures used in Florida, Illinois, Maryland, Missouri, and North Carolina. Table 1 provides a description of the most commonly used existing SWPBIS fidelity measures, along with the TFI. As with these other tools, the TFI is freely available for download at http://www.pbis.org.

The TFI is organized into three scales, representing Tier I (universal), Tier II (targeted), and Tier III (intensive). Each scale can be assessed separately or together to evaluate overall implementation at all three tiers. These options allow for various intended uses: (a) as a complete index of all tiers to establish implementation status and determine focus, (b) as a quarterly progress monitoring tool to guide action planning for implementation of tiers of current focus, and (c) as an annual formative evaluation for tiers already in place. Teams use a Likert-type scale and detailed rubric to indicate whether the content of each item is *not implemented*, *partially implemented*, or *fully implemented*. Data sources are included to help teams evaluate each item objectively. Tier I (universal SWPBIS features) assesses 15 critical features of school-wide supports such as "School has five or fewer positively stated behavioral expectations and examples by setting/location for student and staff behaviors (i.e., school teaching matrix) defined and in place." Subscales in the Tier I scale include Teams (two items), Implementation (10 items), and Evaluation (three items). Tier II (targeted SWPBIS features) evaluates 13 core features of targeted interventions such as "Tier II team uses decision rules and multiple sources of data (e.g., ODRs, academic progress, screening tools, attendance, teacher/family/student nominations) to identify students who require Tier II supports." Subscales in the Tier II scale include Teams (four items), Interventions (five items), and Evaluation (four items). Tier III (intensive SWPBIS features) includes 17 items (e.g., "A written process is followed for teaching all relevant staff about basic behavioral theory, function of behavior, and function-based intervention"). Subscales in the Tier III scale include Teams (four items), Resources (three items), Plans (six items), and Evaluation (four items).

Because research has shown that self-assessment of fidelity can be artificially inflated (Noell et al., 2005; Wickstrom, Jones, LaFleur, & Witt, 1996), it is important to ensure that results from fidelity measures are accurate;

Table 1. SWPBIS Fidelity Measures.

Measure	Tiers assessed	Type	Purposes	Completers	Subscales (items)
Team Implementation Checklist	Tier I	Self-assessment	Assess fidelity at Tier I (and elements of Tier III) Guide start-up activities Progress monitoring	SWPBIS team	Commitment (2) Team (3) Self-Assessment (3) Prevention Systems (6) Classroom (2) Information Systems (3)
PBIS Self-Assessment Survey	Tier I	Self-assessment	Assess fidelity at Tier I (and elements of Tier III) Needs assessment Obtain staff input	SWPBIS team or all school staff	Function-Based Support (3) School-Wide Systems (18) Nonclassroom Setting Systems (9) Classroom Systems (11) Individual Student Systems (8)
School-Wide Benchmarks of Quality	Tier I	External or self-assessment	Assess fidelity at Tier I Guide full implementation	External coach and SWPBIS team	PBIS Team (3) Faculty Commitment (3) Effective Procedures for Dealing With Discipline (6) Data Entry and Analysis Plan Established (4) Expectations and Rules Developed (5) Reward/Recognition Program Established (7) Lesson Plans for Teaching Expectations/Rules (6) Implementation Plan (7) Classroom Systems (7) Evaluation (5)
School-Wide Evaluation Tool	Tier I	External assessment	Assess fidelity at Tier I External evaluation Obtain outside perspective	External assessor (with staff/student interviews)	Expectations Defined (2) Behavioral Expectations Taught (5) Ongoing System for Rewarding Behavioral Expectations (3) System for Responding to Behavioral Violations (4) Monitoring and Decision-Making (4) Management (8)
Monitoring Advanced Tiers Tool	Tiers II and III	Self-assessment	Assess fidelity at Tiers II and III Guide systems implementation Progress monitoring	External coach and Tier II/III team	District-Level Support (2) Tier II Tier I Critical Elements (1) Organizational Elements (7) Critical Elements (7) Tier III Tier I Critical Elements (1) Organizational Elements (7) Critical Elements (7)
Benchmarks for Advanced Tiers	Foundations,[a] Tiers II and III	External or self-assessment	Assess fidelity at Tiers II and III Guide full systems implementation	External coach and Tier II/III team	Foundations Implementation of SWPBIS (3) Faculty Commitment (3) Student Identification (4) Monitoring and Evaluation (2) Tier II Support Systems (5) Implementation (10) Monitoring and Evaluation (4) Tier III Intensive Support Systems (12) Assessment and Plan Development (10) Monitoring and Evaluation (3)
Individual Student Systems Evaluation Tool	Foundations,[a] Tiers II and III	External assessment	Assess fidelity at Tiers II and III External evaluation Obtain outside perspective	External assessor (with staff interviews)	Foundations Commitment (5) Team-Based Planning (3) Student Identification (5)

(continued)

Table 1. (continued)

Measure	Tiers assessed	Type	Purposes	Completers	Subscales (items)
					Monitoring and Evaluation (5)
					Tier II
					Implementation (4)
					Evaluation and Monitoring (2)
					Tier III
					Assessment (3)
					Implementation (6)
					Evaluation and Monitoring (2)
SWPBIS Tiered Fidelity Inventory	Tiers I, II, and III	External or self-assessment	Assess fidelity at all three tiers Guide systems implementation Progress monitoring	External coach and SWPBIS teams (I, II, and III)	Tier I
					Teams (2)
					Implementation (9)
					Evaluation (4)
					Tier II
					Teams (4)
					Interventions (5)
					Evaluation (4)
					Tier III
					Teams (4)
					Resources (3)
					Support Plans (6)
					Evaluation (4)

Note. SWPBIS = school-wide positive behavioral interventions and supports; Foundations = systems-level components needed for implementing at Tiers II and III; PBIS = Positive Behavioral Interventions and Supports.
[a]Systems-level components needed for implementing at Tiers II and III.

otherwise, decisions will be flawed. The TFI is intended for use by school teams with the support of an external SWPBIS coach, who facilitates the administration, ensures accuracy of scoring, and guides the team through interpreting the results. Due to varying team membership, the group assessing Tier I supports may be different from the assessors of Tier II and Tier III supports. The TFI can be completed online (http://www.pbisapps.org) or using pencil and paper. After a complete administration of the TFI, summary scores for each scale are provided, representing the percentage of critical features implemented at Tiers I, II and III, as well as a total score for all three tiers. Subscale and item reports are generated to guide coaching and action planning for school teams.

Purpose of the Technical Adequacy Studies

To assess the reliability and validity of the TFI to measure implementation at all three tiers and continue to refine it based on results, we evaluated the psychometric properties of the measure through three studies: (a) a content validity study, (b) a usability and reliability study, and (c) a large-scale validation study. First, an expert panel evaluated the content validity of the TFI, including evaluating the importance of each specific item, how it related to a particular aspect of fidelity, and the usefulness and appropriateness of scoring. Second, the TFI was pilot tested with a small group of school teams and coaches to evaluate the usability of the measure as well as calculate the interrater and test–retest reliability of the tool. Third, the TFI was released nationally for administration under typical conditions to assess its relation to existing SWPBIS fidelity measures. The remainder of the article describes the methods and results of these studies. Because these studies used different samples and methodologies, they are described separately.

Content Validity Study

Method

Participants. Twelve experts in SWPBIS implementation were invited to participate in the content validity study and assess how each item was related to implementation, as well as rate the measure as a whole. Participants had to be one or both of the following: (a) a researcher in SWPBIS with at least two published studies using and reporting SWPBIS fidelity of implementation data in the past 10 years ($n = 5$) or (b) an experienced SWPBIS implementer with at least 15 years of experience as a school- and district-level implementer and team trainer ($n = 7$). Individuals were not eligible to participate if they assisted in developing the TFI or shared an institutional affiliation with any developers. There was a 100% response rate, with 2% of responses with missing data.

Measure. We used a survey to assess content validity, the extent to which the specific items of the TFI adequately

represent implementation of SWPBIS, which assists in assessing whether the items should be retained, revised, or removed, as well as whether the measure as a whole is valid (Polit & Beck, 2006; Waltz, Strickland, & Lenz, 2005). The survey (based on previous content validity research; McIntosh, MacKay, et al., 2011) included three sections. For each item, we asked (a) the extent to which it addressed important aspects of fidelity of implementation (to assess item validity), (b) the extent to which it was related to the proposed subscale (to indicate factor structure), and (c) the extent to which the scoring criteria were valid (to assess validity of scoring). For each scale, we asked the extent to which the items assessed important aspects of the tier and whether any items should be added or removed. For the measure as a whole, we asked six overall questions (e.g., directions, response format, overall content validity). We used a 4-point Likert-type scale (*strongly disagree*, *disagree*, *agree*, *strongly agree*) for each question and also asked for open-ended feedback, such as suggestions for rewording items and specifying items to add or remove from the measure.

Procedure. We invited participants to complete the survey anonymously through a secure online surveying program. Two separate analyses were conducted to evaluate the data from the content validity survey. First, interrater agreement (IRA) was calculated to determine the extent to which the experts' ratings were consistent. As recommended when the number of expert panel participants is 5 or more (Davis, 1992; Lynn, 1986), the 4-point scale was dichotomized by combining *strongly disagree* and *disagree* as one rating and *agree* and *strongly agree* as one rating. The IRA was calculated for each item and for the survey as a whole. Next, a *Content Validity Index* (CVI) score was calculated for each item based on the representativeness of the assessment tool. The number of experts who rated an item as *agree* or *strongly agree* was counted for each item. This sum was divided by the total number of experts to derive the CVI for each item. The overall CVI for the instrument was determined by averaging the CVI for each item. A CVI of .80 or higher is recommended in the literature for new assessment measures, and items below .80 should be examined for revision (Davis, 1992).

Results

Overall, the expert panel reliability (i.e., the extent to which the raters agreed on their ratings) was 93% (Tier 1 = 95%, Tier II = 93%, Tier III = 91%), with 95% of items above the .80 standard. Furthermore, the reliability was 96% for item validity, 95% for factor structure, and 89% for scoring. These figures indicate a high level of agreement among the experts regarding the TFI and its items. The overall CVI was .92, with 95% of questions rated above the criterion of .80 (range = .67–1). The mean CVI for Tier I items was .95 (range = .67–1). Of the two Tier I items rated below the CVI criterion, one was rated as not aligned to the critical features of implementation, and one was rated as unclear in wording. The mean CVI for Tier II items was .93 (range = .75–1). One Tier II item was rated below the criterion. The scoring criteria for this item were noted as unclear. Finally, the mean CVI for Tier III items was .91 (range = .67–1). Three items were scored below the criterion, and feedback from the expert panel indicated the need for more universal language related to intensive interventions (e.g., person-centered planning, *R*ehabilitation for *E*mpowerment, *N*atural Supports, *E*ducation, and *W*ork [RENEW], wrap-around services). Overall, the content validity data demonstrate that the expert panel considered the items, scoring criteria, and overall structure to be a valid measure of the important aspects of fidelity of implementation of SWPBIS.

Changes to Measure

All six TFI items that were rated below the .80 content validity criterion were changed. Based on the feedback from experts, one item was removed from the measure, one item description was revised, scoring criteria for one item were changed, and three items were reworded in both the description and scoring criteria. These items were revised to reflect a common, universal language related to interventions, and scoring criteria were revised to align with the item description. Along with these changes, an item assessing meeting procedures was added to all three tiers and an item evaluating a range of Tier II interventions was included. These additions were based on the open-ended feedback. All changes were made prior to pilot testing.

Usability and Reliability Study

Method

Participants. This study included school teams and their external coaches from 15 schools in five districts across five states (Connecticut, Michigan, Missouri, North Carolina, and Oregon). School SWPBIS teams were recruited by their state SWPBIS leadership teams to provide a range of implementation (i.e., from first year of implementation of Tier I SWPBIS to strong implementation at all three tiers; mean years implementing = 5.56). Schools included elementary ($n = 6$), K-8 ($n = 2$), middle ($n = 1$), junior high/high ($n = 4$), high ($n = 1$), and K-12 schools ($n = 1$). Enrollment for schools with National Center for Education Statistics (NCES) data ($n = 14$) ranged from 33 to 1,586 ($M = 511.79$), and percentage of students eligible for free and reduced-price lunch ranged from 5% to 91% ($M = 55.79\%$). Each school team completed the TFI and a usability survey, although in some schools, separate teams completed each

scale (i.e., Tier I team completed the Tier I scale, and the Tier II/III team completed the others).

Measure. We developed a usability survey to assess the extent to which the process of administering, scoring, and interpreting the TFI was easy and straightforward. It included 14 questions with a 4-point Likert-type scale (from *strongly disagree* to *strongly agree*). For each scale, school teams reported completion time, the extent to which the items assessed important aspects of implementation, and whether items should be added or removed. We also asked them to provide open-ended feedback to improve the measure. The internal consistency of the usability survey (in terms of coefficient alpha) was .87, indicating acceptable reliability. There were no missing usability or TFI data.

Procedure. Pilot study participants completed the TFI and usability survey immediately afterward. The usability and reliability of the TFI was determined through multiple methods of evaluation: (a) the usability survey, providing both quantitative and descriptive data; (b) one TFI completed by the coach prior to using it with the team; and (c) two administrations of the TFI by the coaches facilitating the school teams, provided exactly 2 weeks apart.

Three different analyses were conducted: (a) usability interpretation, (b) calculation of interrater reliability, and (c) calculation of test–retest reliability. Usability encompasses the effectiveness, efficiency, and user satisfaction of a measure (Frøkjær, Hertzum, & Hornbæk, 2000). For consistency with the content validity analyses, we dichotomized the 4-point survey scale and calculated the percentage of responses that were coded as *disagree* or *agree*. Items with less than 80% agreement were reevaluated, with changes to the items as needed. We calculated interrater reliability, the extent to which different raters are consistent when using the measure (James, Demaree, & Wolf, 1984; Shrout & Fleiss, 1979), by comparing the score of the coach's independent TFI (i.e., before meeting with the team) with the score of the administration with the coach leading the team. To do so, we used a two-way random consistency intraclass correlation (ICC) analysis in SPSS. Finally, we calculated test–retest reliability, the extent to which scores vary when the measure (i.e., TFI) is used across time, by comparing the scores of the team's initial TFI results with those of the 2-week retest. We calculated these ICCs using a two-way random consistency analysis in SPSS.

Results

Usability. Average completion time for each scale was under 15 min (Tier I: 14.5 min, Tier II: 11 min, Tier III: 12.5 min). Responses assessing the overall TFI measure showed strong usability (easy and straightforward process: 100% agree, easy and straightforward scoring: 93% agree, validity for assessing fidelity: 100% agree). Out of 14 questions assessing usability, two had less than 80% agreement (range = .67–1). These questions evaluated the extent to which participants rated that an item should be removed from the TFI. Four participants suggested that an item should be removed from Tier II, and three participants suggested that an item should be removed from Tier III. The most common open-ended feedback from the usability survey was that the TFI was easy to use, and respondents appreciated that they could use one measure to assess fidelity at all three tiers. Respondents were split as to whether the TFI could replace existing fidelity measures. Many noted that it could replace existing Tier II and III measures, but they reported that other Tier I measures could be used for the specialized purposes noted in Table 1 (e.g., TIC for initial implementation, BoQ for deep implementation, SAS for obtaining perceptions from whole staff).

Interrater reliability. The ICCs for interrater reliability across all raters, all tiers, and all items (Tier I, Tier II, Tier III, and overall) were all .99. These scores indicate high reliability in scores between coaches (when completing the TFI about a school alone) and the teams (when assessing fidelity with the TFI with the coach as facilitator).

Test–retest reliability. The ICC for test–retest reliability was .99. These test–retest reliability scores indicate very strong agreement across administrations of the TFI over time, which indicates that the construct is being measured consistently by the measure.

Changes to Measure

Based on the information in the usability survey, TFI items were reworded for clarity in the item description, scoring, or data sources. The majority of changes were clarifying terminology (e.g., person-centered planning, wraparound) and aligning the item descriptions and scoring criteria. One item was added to the Tier I section to split the stakeholder involvement item into two separate items, one item measuring faculty involvement and another measuring student, families, and community member involvement.

Large-Scale Validation Study

Method

Participants. The pilot study included 789 schools across seven states, primarily in Florida and Illinois, in the 2013–2014 school year. Each school completed the TFI, along with at least one of four other fidelity of implementation measures (e.g., BoQ, SAS, TIC, and BAT). Scores from the usability and reliability study (the first administration with

Table 2. School Characteristics for Validation Study Sample (n = 789).

Variable	*M* or % (*SD*)
Years implementing SWPBIS	6.19 (3.40)
Enrollment	538.001 (418.170)
% of students receiving FRL	50.9% (0.269)
% of non-White students	50.1% (0.415)
Grade level	
Elementary	69.9%
Middle	19.7%
High	7.7%
Other	2.8%
Urbanicity	
Rural	13.9%
Town	10.9%
Suburb	49.5%
City	25.7%
TFI scores	
Tier I	83.9% (0.154)
Tier II	68.0% (0.321)
Tier III	32.4% (0.348)
Total	59.3% (0.22)

Note. Years implementing SWPBIS available for 96% of schools (n = 759). School demographic data obtained from National Center for Education Statistics for 91% of schools (n = 717). SWPBIS = school-wide positive behavioral interventions and supports; FRL = free and reduced-price lunch; TFI = *Tiered Fidelity Inventory*.

coach and team) were also included in analyses. Table 2 provides descriptive statistics for this sample.

Measures. Four research-validated measures were used as concurrent measures of SWPBIS implementation: (a) the School-Wide BoQ (Kincaid et al., 2005), (b) the TIC (Sugai, Horner, et al., 2001), (c) the SAS (Sugai et al., 2000), and (d) the BAT (Anderson et al., 2012). The BoQ, SAS, and TIC were used as comparisons for the Tier I scale of the TFI. The BAT Tier II and Tier III scale scores were used as comparisons with the TFI Tier II and III scales. The overall BAT score, which includes the Foundations, Tier II, and Tier III subscales, was compared with the TFI total score (i.e., Tiers I, II, and III).

BoQ. The BoQ is a 53-item Tier I SWPBIS fidelity of implementation scale. The psychometric properties of the BoQ indicate the tool is reliable and valid for measuring Tier I SWPBIS fidelity, with interrater and test–retest reliability above 90% and moderate correlations with the SET (Sugai, Lewis-Palmer, et al., 2001), another Tier I measure (R. Cohen, Kincaid, & Childs, 2007). A total of 321 schools in the sample completed both the BoQ and TFI.

SAS. The SAS is a 43-item self-assessment measure of SWPBIS implementation. For these analyses, the 18-item School-Wide Systems scale was used to assess Tier I implementation. The SAS has high internal consistency and correlations with other validated SWPBIS fidelity measures (Hagan-Burke et al., 2005; Safran, 2006). Internal consistency for all tiers is high (α = .85), and subscale scores range from moderate to high (α range = .60–.92). Concurrent validity with Tier I SET is moderately high (r = .75). A total of 559 schools in the sample completed both the SAS and TFI.

TIC. The TIC is a 17-item measure of Tier I SWPBIS implementation. It assesses the extent to which key start-up activities are implemented. The TIC is intended for use as a progress monitoring assessment measure, and a score of 80% or higher indicates implementation of SWPBIS to criterion levels. Internal consistency for the TIC is high across studies (α range = .91–.94; McIntosh, Mercer, Nese, Strickland-Cohen, & Hoselton, in press; Tobin, Vincent, Horner, Dickey, & May, 2012), and a recent confirmatory factor analysis showed a strong factor structure (McIntosh et al., in press). A total of 164 schools completed both the TIC and TFI.

BAT. The BAT is a 112-item fidelity of implementation measure that assesses implementation at Tiers II and III, as well as foundational structures for supporting systems at Tiers II and III. As with the TFI, each tier can be completed separately, if desired. No published technical adequacy data are available for the BAT. A total of 198 schools completed both the BAT and TFI.

Procedure. School teams and external SWPBIS coaches in two states (Florida and Illinois) were provided with access to the TFI as an additional fidelity of implementation measure in addition to the existing fidelity measures that they were already using. Training for TFI administration was not tightly controlled—Participants were provided access to the measure and a webinar, with no requirement of training or contact with the study authors. When completing the TFI, respondents indicated whether the measure was completed by the school team with an external coach (n = 437) or by the school team alone (n = 282).

Data analysis. Analyses in this study assessed multiple elements of reliability and validity in assessing SWPBIS fidelity. Analyses produced information regarding (a) internal consistency (through coefficient alpha),and (b) concurrent validity with existing measures of SWPBIS implementation (through Pearson correlations). There were no missing TFI data.

Results

Internal consistency. Coefficient alpha was used to evaluate the internal consistency of the measure. The overall internal consistency of the measure was .96. Alphas for Tiers I, II,

Table 3. Correlations Between TFI and Existing Measures of Fidelity of Implementation by Administration Condition.

Measures	Team without external coach	Team with external coach	z test of difference
TFI Tier I and BoQ	.416** (*n* = 106)	.643** (*n* = 215)	2.668**
TFI Tier I and SAS	.364** (*n* = 209)	.551** (*n* = 350)	2.710**
TFI Tier I and TIC	.258* (*n* = 65)	.544** (*n* = 99)	2.123*
TFI Tier II and BAT Tier II	.243* (*n* = 74)	.507** (*n* = 124)	2.200*
TFI Tier III and BAT Tier III	.639** (*n* = 74)	.723** (*n* = 124)	1.160
TFI total and BAT total	.474** (*n* = 74)	.750** (*n* = 124)	3.062**

Note. TFI = *Tiered Fidelity Inventory*; BoQ = *Benchmarks of Quality*; SAS = *Self-Assessment Survey*; TIC = *Team Implementation Checklist*; BAT = *Benchmarks for Advanced Tiers*.
*$p < .05$. **$p < .01$. ***$p < .001$.

and III were .87, .96, and .98, respectively, providing evidence of strong internal consistency.

Correlations. Pearson correlations were calculated between the TFI and other existing measures of fidelity of implementation. Correlations were calculated separately by administration condition (i.e., team without external coach and team with external coach). Results are summarized in Table 3. All correlations between the TFI and other measures were statistically significant and were stronger when the team completed the TFI with an external coach. According to criteria from J. Cohen (1988), correlations were generally moderate without a coach, and all were strong with a coach. Furthermore, teams consistently rated their implementation as higher when they completed the measure without an external coach than when they completed an administration with an external coach, indicating a small degree of self-inflation.

Discussion

Research has demonstrated that schools with higher SWPBIS fidelity scores have better student outcomes (e.g., lower rates of problem behavior, higher achievement, higher emotional regulation; Bradshaw, Waasdorp, & Leaf, 2012; Childs, Kincaid, & George, 2010; Flannery et al., 2014; Horner et al., 2009). Without reliable and valid assessment of fidelity, there is a danger of assuming that implementation is adequate when it is not. The purpose of this study was to validate and refine a new, comprehensive measure of fidelity of implementation of SWPBIS, the SWPBIS TFI. The TFI was intended to serve as a single measure for assessing SWPBIS implementation at all three tiers, which could provide advantages in terms of efficiency and ease of evaluation for districts and states. Three separate studies were conducted to assess the measure's construct validity, usability, reliability, and concurrent validity with existing, validated measures of SWPBIS fidelity of implementation. After each study, the measure was refined to continue to enhance its technical adequacy. Collectively, results showed that the measure can be used reliably and validly to assess SWPBIS fidelity of implementation. Results are described by reliability, validity, and usability.

Psychometric Properties of the TFI

Reliability. Educators and administrators need to have confidence that their selected fidelity measures will produce similar scores across conditions. Evidence for reliability comes from the usability and reliability study and the large-scale validation study. The usability and reliability study provided evidence of both IRA (between the coach alone and team facilitated with coach) and test–retest reliability (the team's ratings over time). Finally, the internal consistency of the measure (from the large-scale validation study) demonstrated high internal consistency overall and within individual tiers. These results provide evidence that the TFI can provide consistent results across raters and time.

Validity. Multiple aspects of validity were assessed. Content validity results (from the expert panel ratings) indicated that the items, scoring criteria, and perceived factor structure of the TFI are valid for assessing the construct of SWPBIS implementation. Concurrent validity analyses (comparisons between the TFI and the BoQ, TIC, SAS, and BAT) showed statistically significant correlations with the other existing SWPBIS fidelity measures, providing indications that the TFI is a valid measure of SWPBIS fidelity.

In line with previous research, relations with other measures were stronger when school teams completed the measure with the guidance of an external coach. Completing the measure without a coach produced adequately valid scores, but the scores appear to have been somewhat inflated, as seen through slightly higher mean scores and lower correlations with other measures. As a result, scores from the TFI appear to be most valid when it is completed with an external coach.

Usability. Although reliability and validity are important, a measure's utility for decision making is a key factor for applied measures. Evidence for the TFI's usability came primarily from the usability and reliability study. Users

reported that the TFI was easy and straightforward to complete and score, and that it assessed important aspects of fidelity at all three tiers. Descriptive feedback indicated that the TFI was efficient and useful for decision making and action planning to improve systems. Such results indicate that the TFI would be useful for its intended purposes.

Limitations and Future Research

Some limitations of the three studies are apparent. For example, participants in the usability and reliability study were likely to be enthusiastic. It is possible that such selection, although it may have increased the quantity and quality of descriptive feedback, may have biased the results. In addition, the authors themselves did not conduct any external evaluations of SWPBIS fidelity. As a result, the teams completing the TFI may have been the exact same groups participating in administration of the other fidelity measures. In regard to these measures, the lack of detailed technical adequacy data for the BAT makes our findings regarding the TFI Tier II and III scales more tentative than for Tier I. Furthermore, the usability and reliability study's interrater reliability assessment was conducted with a coach as part of both administrations. Although it is difficult to identify another way to evaluate interrater reliability for a team-based assessment, it is possible that the coach's presence in both administrations inflated the interrater reliability estimates. Finally, the time of year for concurrent validity was not controlled. As a result, the other measures may have been completed close or far away in time from the TFI administration.

Although these results are promising, further validation work would be useful to assess the technical adequacy of the TFI. First, it will be necessary to validate the finalized TFI measure based on the slight changes to the measure from the final round of feedback. Second, the criterion for adequate implementation (e.g., 70% of total points) has not yet been studied. It will be necessary to identify empirical criteria for adequate implementation. In absence of this research, 70% appears to be a reasonable criterion for adequate implementation at each tier, although mean implementation at Tiers II and III was considerably lower. Third, a rigorous, quantitative assessment of the TFI's factor structure is necessary (and currently underway). Fourth, it would be useful to further examine the role of coaches in facilitating accurate assessment of fidelity and what factors enhance accuracy in self-rating of fidelity.

Implications for Practice

These results provide indications that the TFI has strong technical adequacy for measuring SWPBIS fidelity at all three tiers and is an appropriate index of implementation. Coaches and coordinators at the school, district, regional, and state levels should feel confident in the measure's properties and the accuracy of its results. The measure can be used to produce valid results for total, tier, and subscale scores in typical administration (i.e., without extensive training and support in administration). However, the validation study results confirm the TFI authors' recommendations that administration be conducted with an external coach, due to the objectivity of an outside evaluator. When teams lack an external support to provide additional perspective, the phenomenon of "self-inflation" of fidelity appears to be more likely.

SWPBIS leaders at the school, district, and state can consider whether the TFI can supplement or replace current SWPBIS fidelity measures required for their evaluation plans. Respondents reported that they appreciated the TFI's comprehensive (i.e., all three tiers in one measure) nature, but that some existing Tier I measures would remain useful for school teams, depending on their specific needs at the time. All of these measures will remain available for administration, scoring, and reporting at http://www.pbisapps.org.

Acknowledgments

The authors wish to thank Stephanie Austin, Linda Bradley, Karen Childs, Bridget Drobac, Susannah Everett, Sarah Moore, Jennifer Rollenhagen, and Erin White for their assistance in data collection.

Authors' Note

The opinions expressed are those of the authors and do not represent views of the Office or U.S. Department of Education.

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: This research was supported by the Office of Special Education Programs, U.S. Department of Education (H326S130004).

References

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation, 14*, 1–25.

Algozzine, R. F., Barrett, S., Eber, L., George, H., Horner, R. H., Lewis, T. J., . . .Sugai, G. (2014). *SWPBIS Tiered Fidelity Inventory*. Eugene, OR: OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. Available from http://www.pbis.org

Algozzine, R. F., Horner, R. H., Sugai, G., Barrett, S., Dickey, C. R., Eber, L., . . .Tobin, T. (2010). *Evaluation blueprint for school-wide positive behavior support* (2nd ed.). Eugene, OR:

National Technical Assistance Center on Positive Behavior Interventions and Support. Available from www.pbis.org

Anderson, C. M., Childs, K., Kincaid, D., Horner, R. H., George, H., Todd, A. W., . . .Spaulding, S. A. (2012). *Benchmarks for Advanced Tiers*. Unpublished instrument, Educational and Community Supports, University of Oregon & University of South Florida.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, *12*, 133–148. doi:10.1177/1098300709334798

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems and adjustment. *Pediatrics*, *13*, e1136-e1145. doi:10.1542/peds.2012-0243

Childs, K. E., Kincaid, D., & George, H. P. (2010). A model for statewide evaluation of a universal positive behavior support initiative. *Journal of Positive Behavior Interventions*, *12*, 198–210. doi:10.1177/1098300709340699

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 school-wide positive behavior support implementation: Development and validation of the Benchmarks of Quality. *Journal of Positive Behavior Interventions*, *9*, 203–213.

Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, *5*, 194–197.

Fixsen, D. L., Blase, K. A., Duda, M. A., Naoom, S. F., & Van Dyke, M. (2010). Implementation of evidence-based treatments for children and adolescents. In. J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 435–450). New York, NY: Guilford Press.

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

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000, April). *Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, The Hague, Netherlands.

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. (2014, August). *Compression implementation and scaling PBIS*. Paper presented at the Wisconsin PBIS Network Conference, Wisconsin Dells, WI.

Horner, R. H., Sampson, N. K., Anderson, C. M., Todd, A. W., & Eliason, B. M. (2013). *Monitoring Advanced Tiers Tool*. Eugene: Educational and Community Supports, University of Oregon.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esparanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*, 133–144.

Individuals With Disabilities Education Improvement Act, 20 U.S.C. § 1400 P.L. 108-446 (2004).

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*, 85–98.

Kelm, J. L., & McIntosh, K. (2012). Effects of school-wide positive behavior support on teacher self-efficacy. *Psychology in the Schools*, *49*, 137–147. doi:10.1002/pits.20624

Kincaid, D., Childs, K., & George, H. P. (2005). *School-Wide Benchmarks of Quality*. Unpublished instrument, University of South Florida, Tampa.

Lewis-Palmer, T., Todd, A. W., Horner, R. H., Sugai, G., & Sampson, N. K. (2003). *Individual Student Systems Evaluation Tool (ISSET)*. Eugene, OR: Educational and Community Supports.

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, *35*, 382–385.

McIntosh, K., Bennett, J. L., & Price, K. (2011). Evaluation of social and academic effects of School-wide positive behaviour support in a Canadian school district. *Exceptionality Education International*, *21*, 46–60.

McIntosh, K., Kim, J., Mercer, S. H., Strickland-Cohen, M. K., & Horner, R. H. (2015). Variables associated with enhanced sustainability of school-wide positive behavioral interventions and supports. *Assessment for Effective Intervention*, *40*, 184–191.

McIntosh, K., MacKay, L. D., Hume, A. E., Doolittle, J., Vincent, C. G., Horner, R. H., & Ervin, R. A. (2011). Development and initial validation of a measure to assess factors related to sustainability of school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *13*, 208–218. doi:10.1177/1098300710385348

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., Strickland-Cohen, M. K., & Hoselton, R. (in press). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders*, *10*, 136–148.

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.

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.

Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*, 489–497.

Ross, S. W., Romer, N., & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, *14*, 118–128.

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.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review*, *38*, 460–475.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, *86*, 420–428.

Sugai, G., & Horner, R. H. (2009). Defining and describing schoolwide 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

Sugai, G., Lewis-Palmer, T. L., Todd, A. W., & Horner, R. H. (2001). *School-Wide Evaluation Tool (SET)*. Eugene, OR: Educational and Community Supports. Available from http://www.pbis.org

Tobin, T., Vincent, C. G., Horner, R. H., Dickey, C. R., & May, S. A. (2012). Fidelity measures to improve implementation of positive behavioural support. *International Journal of Positive Behavioural Support*, *2*, 12–19.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2005). *Measurement in nursing and health research* (3rd ed.): New York, NY: Springer.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1996). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly*, *13*, 141–154.