Article

Hammill Institute on Disabilities

Remedial and Special Education
2019, Vol. 40(1) 16–24

DOI: 10.1177/0741932517736515
rase.sagepub.com
SAGE

Factor Validation of a Fidelity of Implementation Measure for Social Behavior Systems

Michelle M. Massar, PhD[1], Kent McIntosh, PhD[1], and Sterett H. Mercer, PhD[2]

Abstract

Assessing fidelity of implementation of school-based interventions is a critical factor in successful implementation and sustainability. The Tiered Fidelity Inventory (TFI) was developed as a comprehensive measure of all three tiers of School-Wide Positive Behavioral Interventions and Supports (SWPBIS) and is intended to measure the extent to which the core features of SWPBIS are implemented with fidelity. The purpose of this study was to assess the extent to which the TFI can be used as one measure of all three tiers, three separate measures of individual tiers, or as a more granular level of fidelity that measures implementation on 10 subscales across the tiers. A confirmatory factor analysis was conducted to validate the factor structure of the TFI. Results indicate that the TFI is a valid measure of fidelity of implementation of SWPBIS and can be used to measure implementation by subscales, tiers, and as a comprehensive assessment of all three tiers.

Keywords

positive behavior supports, School-Wide Positive Behavioral Interventions and Supports, treatment integrity, fidelity of implementation

Fidelity of implementation refers to the extent to which the core features of a program, intervention, or system are implemented as intended to maximize effectiveness (Fagan, Hanson, Hawkins, & Arthur, 2008). Fidelity serves as the bridge between the development of evidence-based practices (EBPs) and the successful adoption and use of EBPs in natural settings such as schools, clinics, and programs. Research indicates that an EBP's fidelity of implementation is strongly related to achieving its intended outcomes (e.g., Carroll et al., 2007; Dane & Schneider, 1998; Elliott & Mihalic, 2004; Flannery, Fenning, Kato, & McIntosh, 2014). Fidelity data can be used as guide for implementation and as a progress-monitoring tool to ensure quality and effectiveness over time (e.g., Bond, Becker, Drake, & Vogler, 1997; Bond, Evans, Salyers, Williams, & Kim, 2000; Mowbray, Holter, Teague, & Bybee, 2003). When evaluating whether an intervention is effective, it is important to understand the extent to which its core features are being implemented as intended.

Although measuring fidelity has long been recognized as important for quality research, providing tools for practitioners to assess fidelity is key for large-scale implementation and sustainability (McIntosh et al., 2013). The scientific community has recognized the importance of fidelity tools to ensure that fidelity of implementation is occurring (McIntyre, Gresham, DiGennaro, & Reed, 2007). Collecting self-reported fidelity of implementation data is an appealing alternative to using direct observation or permanent products to measure fidelity (Noell et al., 2005); however, creating and disseminating self-reported measures of fidelity is insufficient to ensure that the results will guide practitioners to higher levels of fidelity and intended outcomes.

The importance of using valid and reliable measures of fidelity cannot be understated. Many fidelity measures assess the delivery of a curriculum such as a social skills program or a social-emotional learning intervention. These measures assess discrete and easily observable and measurable behaviors (e.g., *were the components of the lesson delivered as intended?*). Measuring the fidelity of systems-level interventions (e.g., interventions delivered

[1]University of Oregon, Eugene, USA
[2]University of British Columbia, Vancouver, Canada

Corresponding Author:
Michelle M. Massar, Department of Special Education and Clinical Sciences, College of Education, University of Oregon, 1235 University of Oregon, Eugene, OR 97403, USA.
Email: mmassar@uoregon.edu

at the school level as opposed to individual students) requires considerable thought because the actions for implementation are not as readily apparent. When developing systems-level fidelity of implementation measures, establishing fidelity criteria and developing a tool by which to measure fidelity is critical. Once developed, it is critical to ensure that the tool is a valid measure of fidelity of implementation.

One of the core features of a validated measure is its construct validity. Construct validity assesses the degree to which a construct (e.g., intelligence, aptitude) and their measures correspond (Cronbach & Meehl, 1955; Peter, 1981). Evaluating the extent to which a fidelity tool assesses what it intends to measure is especially important when measuring the various core features of an intervention framework such as School-Wide Positive Behavioral Interventions and Supports (SWPBIS). SWPBIS is a three-tiered framework designed to match universal (Tier I), targeted (Tier II), and intensive (Tier III) interventions to student needs and to establish safe, predictable, consistent, and positive school environments (Sugai & Horner, 2006). SWPBIS is not a manualized intervention; however, the core features that contribute to positive outcomes have been identified at each tier. In other words, schools can adapt aspects of the practices to meet the needs and culture of each school site as long as the critical components of the tiers are being implemented with fidelity. Numerous randomized control trials have evaluated the relation between fidelity of implementation of SWPBIS and valued outcomes such as perception of increased school safety (Horner et al., 2009), improved organizational health and effectiveness (e.g., Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008; Bradshaw, Koth, Thornton, & Leaf, 2009), and reduction in level of student problem behaviors (Bradshaw, Mitchell, & Leaf, 2010). Therefore, developing a tool for assessing fidelity of implementation may increase the likelihood that schools are implementing the core features of SWPBIS as intended, thereby increasing the likelihood that valued student, staff, and school outcomes are achieved.

The tool was created to increase the ease and efficiency by which school teams can measure fidelity within and across tier. The ability to measure fidelity across tiers (or a combination of tiers) with one tool that utilizes a consistent format, scoring process, and interpretation method and provides all of the information in one online location facilitates easier and more efficient measurement of fidelity of implementation. Reducing the response effort for school teams to measure fidelity may increase the likelihood that teams consistently measure fidelity of implementation and use the data for action planning. Because measuring fidelity in a valid, reliable, and efficient manner is critical to successful implementation and sustainability of SWPBIS, a new assessment tool was developed by researchers to guide school teams in durable implementation of the SWPBIS framework.

Tiered Fidelity Inventory (TFI)

Many measures of SWPBIS fidelity of implementation have been developed over the past two decades. Tools to measure SWPBIS fidelity at Tier I include measures such as the Team Implementation Checklist (TIC; Sugai, Horner, & Lewis-Palmer, 2001), Benchmarks of Quality (BoQ; Kincaid, Childs, & George, 2005), PBIS Self-Assessment Survey (SAS; Sugai, Horner, & Todd, 2000), and School-Wide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001). Measures of Tier II and Tier III SWPBIS implementation include the Benchmarks of Advanced Tiers (BAT; Anderson et al., 2012), the Monitoring Advanced Tiers Tool (MATT; Horner, Sampson, Anderson, Todd, & Eliason, 2013), and the Individual Student Systems Evaluation Tool (ISSET; Lewis-Palmer, Todd, Horner, Sugai, & Sampson, 2003). School teams wanting to measure implementation across tiers were required to use multiple measures, often with different scoring systems, wording, and data sources. The SWPBIS Tiered Fidelity Inventory (TFI; Algozzine et al., 2014) was developed to meet the needs of schools implementing or planning to implement and measure multiple tiers.

The TFI is a comprehensive assessment that can be used to measure each tier individually or to evaluate overall implementation across all three tiers. The measure is organized into three scales: Tier I, Tier II, and Tier III, representing universal, targeted, and intensive interventions, respectively. Tier I evaluates fidelity of implementation of universal SWPBIS features by assessing 15 items organized into three subscales: (a) teams, (b) implementation, and (c) evaluation. Tier II measures the extent to which the core features of targeted SWPBIS features are in place by measuring 13 items organized into three subscales: (a) teams, (b) interventions, and (c) evaluation. Finally, Tier III evaluates the fidelity of the core features of intensive SWPBIS features by measuring 17 items organized into four subscales: (a) teams, (b) resources, (c) support plans, and (d) evaluation.

The TFI was designed as a complete index of fidelity of implementation across the three-tiered framework to guide planning, as a progress-monitoring tool to strengthen fidelity, and as a formative measure for tiers that are already implemented. It is available for free (at https://www.pbisapps.org/Applications/Pages/PBIS-Assessment-Surveys.aspx#tfi) and can be completed online (at http://www.pbisapps.org) or by using pencil and paper. School teams use a 3-point scale with supporting data sources and a detailed rubric to determine whether the core feature addressed in each item is *not implemented, partially implemented*, or *fully implemented.* A summary of scores for each subscale, tier, and a total score for all three tiers are

generated after a complete administration of the TFI. The scores and item reports are provided to guide coaching and action planning for SWPBIS implementation.

Review of Previous Technical Adequacy Studies Regarding the TFI

To assess the technical adequacy of the TFI, three separate analyses were conducted to evaluate the validity and reliability of the measure, including a content validity study, a reliability and usability study, and a large-scale validation study (McIntosh et al., 2017). Results showed strong construct validity for assessing fidelity at all three tiers, strong interrater and 2-week test–retest reliability, and high usability for action planning.

The content validity study included 12 experts in the field of implementation and SWPBIS in school settings. To measure the content validity of the TFI, a survey was developed based on previous content validity research [name deleted to maintain the integrity of the review process]. The content validity survey included (a) a set of three questions per item, (b) a set of three questions per tier, and (c) a set of six questions evaluating the measure as a whole (e.g., directions, response format, overall content validity). A 4-point Likert-type scale (*strongly disagree, disagree, agree, strongly agree*) was used for each question. Experts were also given the opportunity to provide descriptive feedback, such as suggestions for rewording items and specifying items to add or remove from the measure. A content validity index (CVI) was generated from the data.

Next, index (CVI) score was calculated for each item, and 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 (Davis, 1992). 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). The six TFI items that had a content validity score below .80 were changed.

Based on the revised TFI, a usability and reliability study was conducted with schools in six states, including Connecticut, Florida, Michigan, Missouri, North Carolina, and Oregon. Participating school teams and coaches were asked to complete multiple measures. First, coaches were asked to complete the TFI prior to completing the measure with their school team(s). Next, school teams were asked to complete one administration of the TFI with coach support and complete an online usability survey. Finally, school teams were asked to complete a second administration of the TFI 2 weeks from the initial TFI administration date. Multiple methods of evaluation were used to evaluate usability and reliability: (a) usability interpretation, (b) calculation of interrater reliability, and (c) measurement of test–retest reliability.

Out of 14 questions assessing usability, two had less than 80% agreement (range = .67–1). Based on the results, items were reworded to clarify terminology (e.g., person-centered planning, wraparound) and align the item descriptions with scoring criteria and one item was added to the Tier I section. Interrater reliability for the TFI total score was .96, with Tier I, II, and III scales calculated at .95, .96, and .89, respectively. Test–retest reliability was .995 for the TFI overall measure and was 0.98, 0.99, and 0.99 for Tiers I, II, and III. Interrater reliability indicates strong reliability between coaches and teams. Test–retest reliability scores indicate very strong agreement across administrations of the TFI over time.

The purpose of the large-scale validation study was to assess internal consistency (through coefficient alpha) and concurrent validity with existing measures of SWPBIS implementation (through Pearson correlations). Participants included 789 schools across seven states in the 2013–2014 school year. Participating schools completed the TFI and at least one other measures of SWPBIS fidelity of implementation (e.g., BoQ, SAS, TIC, BAT). Scores from the usability and reliability study (the first administration with coach and team) were also included in analyses. Coefficient alpha was used to evaluate the internal consistency of the measure. The overall internal consistency of the measure was .96. Alpha coefficients for Tiers I, II, and III were .87, .96, and .98, respectively. These scores indicate strong internal consistency of the measure. Pearson correlations were calculated between the TFI and other existing measures of fidelity of implementation. All correlations between the TFI and other measures were statistically significant and were stronger when the team completed the TFI with an external coach.

Research Gaps Identified in Technical Adequacy Studies

The technical adequacy studies provided evidence that the TFI is a reliable and valid measure of the extent to which the core features of SWPBIS are in place; however, there were limitations to the three studies. First, each study provided new information that was used to make slight modifications to the items, scoring, or criteria included in the measure. Because of the iterative process, the psychometric properties of the final version of the TFI were not assessed. Second, there was no factor analysis completed on the TFI to assess the model fit of the three scales (Tiers I, II, and III) or the 10 subscales to determine the most appropriate way to use the measure to assess SWPBIS fidelity of implementation.

Without conducting a factor analysis, queries related to administering the TFI and interpreting the results remained. Two distinct research questions arose that supported conducting a confirmatory factor analysis (CFA):

Table 1. School Characteristics.

Variable	*M* or % (*SD*)
Enrollment	587.241 (403.743)
% of students receiving FRL	65.4% (25.4%)
% of non-White students	64.3% (29.9%)
Grade level	
Elementary	64.6%
Middle	16.0%
High	8.7%
Other	10.8%
Urbanicity	
Rural	7.1%
Town	9.3%
Suburb	47.7%
City	35.8%
Tiered Fidelity Inventory (TFI) scores	
Tier I	67.0% (26.6%)
Tier II	47.3% (38.9%)
Tier III	23.9% (32.0%)
Total	45.0% (26.2%)

Note. n = 1,209. School demographic data obtained from National Center for Education Statistics for 71% of schools. FRL = free and/or reduced-price lunches.

1. Is implementation of each tier of SWPBIS distinct (i.e., a 3-factor model), or does it represent one unified construct?
2. Are there distinct factors of implementation across tiers (i.e., a 10-factor model)?

The research questions hold implications for both administration and interpretation of the TFI. If all three tiers of the TFI are distinct, then schools can measure the implementation of tiers independently. Furthermore, teams can use the results of each tier to guide planning and implementation independently. If the tiers represent one unified construct then administration of the TFI by tier would not be an appropriate use of the measure. Instead, completing the TFI would necessitate measuring all three tiers. Identifying whether the factors of implementation across tiers are distinct is important for guiding specific improvement in different components of SWPBIS implementation. If the 10 subscales are distinct factors, then teams can identify targeted areas for strengthening implementation. Teams and technical assistance providers could focus on specific areas for training and coaching to support durable implementation of unique SWPBIS components.

Purpose of the Study

To assess the factor structure of the revised TFI and to determine whether it can be used as one scale (i.e., all three tiers), three separate scales (i.e., Tier I, Tier II, and Tier III), or at the subscale level (i.e., all 10 subscales on the measure), we evaluated the model fit through a CFA. CFA results guide recommendations of the valid uses of the measure to assess fidelity of SWPBIS implementation.

Method

Participants

The sample included school SWPBIS teams from 1,708 schools across 25 states, primarily in California and Illinois, in the 2014–2015 school year. School characteristics data from the National Center for Education Statistics (NCES) were available for 71% of the sample. Table 1 provides the most recent NCES data and TFI scores for these schools. The average enrollment for schools participating in the study was 587 students. The majority of participating schools were elementary (64.6%), with middle (16.0%), high (8.7%), and other (10.8%) schools representing a smaller portion of the sample.

Procedure

The TFI was first made available to the public in the 2014–2015 school year through www.pbisapps.org, an online application where school teams can enter and analyze fidelity of implementation and student outcomes data (e.g., school climate surveys) at no charge. During that school year, the SWPBIS technical assistance providers from the Center on PBIS recommended to the states and school districts in its networks that it consider using the TFI to measure implementation in place of or in addition to existing fidelity of implementation tools. Each school then completed the TFI online as part of their typical processes of evaluation of SWPBIS, with no direct contact, incentives, or support from the researchers. The sample in this study included all schools that completed a TFI at any time that year—schools consent to have their data used for research purposes when using the site. Missing data were rare—One TFI was missing a response for one item, and another was missing two items, for .0038% missing data.

Data Analysis

Given the results of an expert content validity study that provided descriptive evidence for the TFI's existing factor structure (3 scales and 10 subscales), we tested a three-factor model (one for each scale, or tier; see Figure 1) and a 10-factor model (one for each subscale; see Figure 2) using CFA through Mplus (Muthén & Muthén, 1998–2012). Because there were only three response options for each item, we used the weighted least square mean and variance adjusted (WLSMV) estimator, which uses

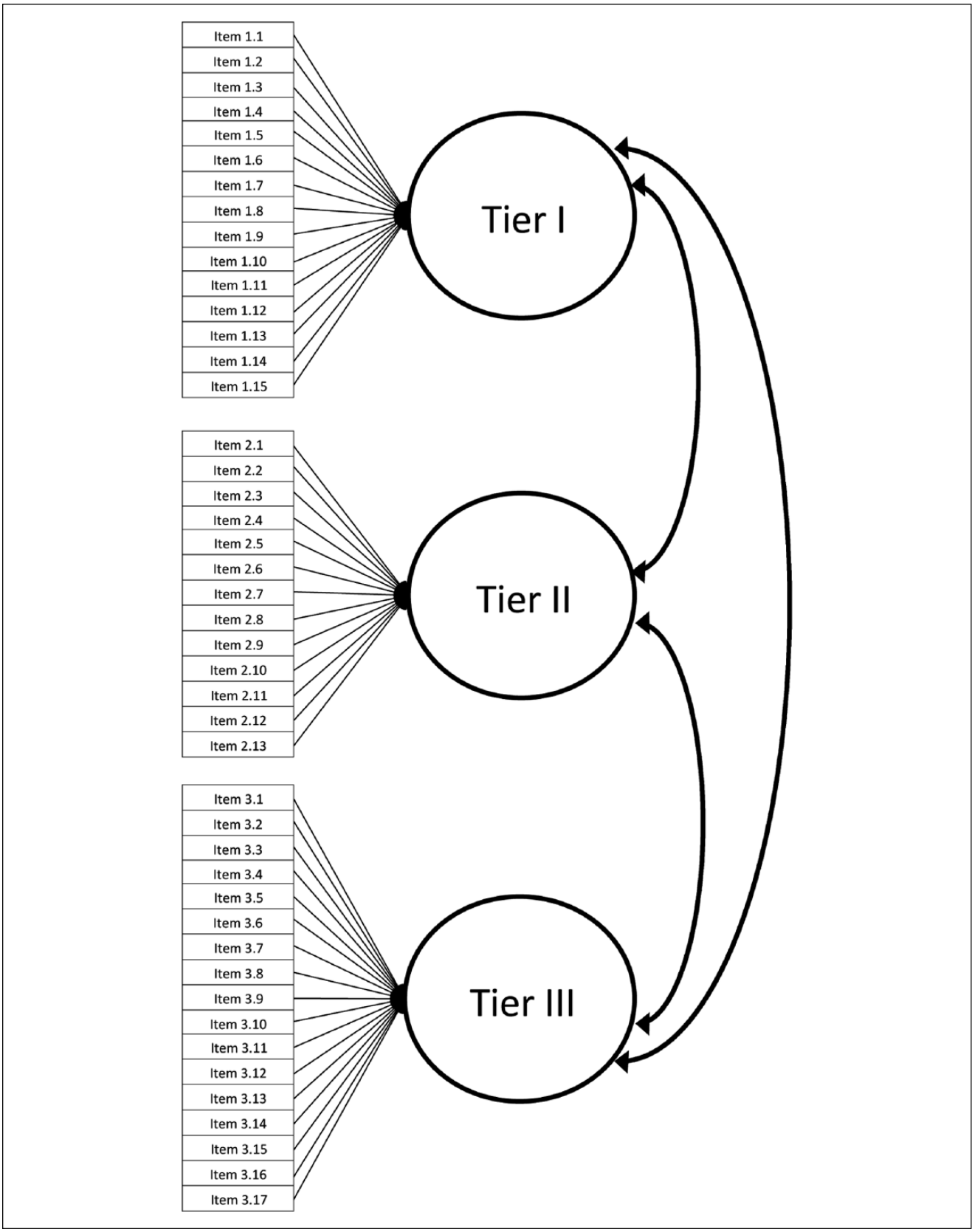

Figure 1. TFI three-factor structure.
Note. TFI = Tiered Fidelity Inventory.

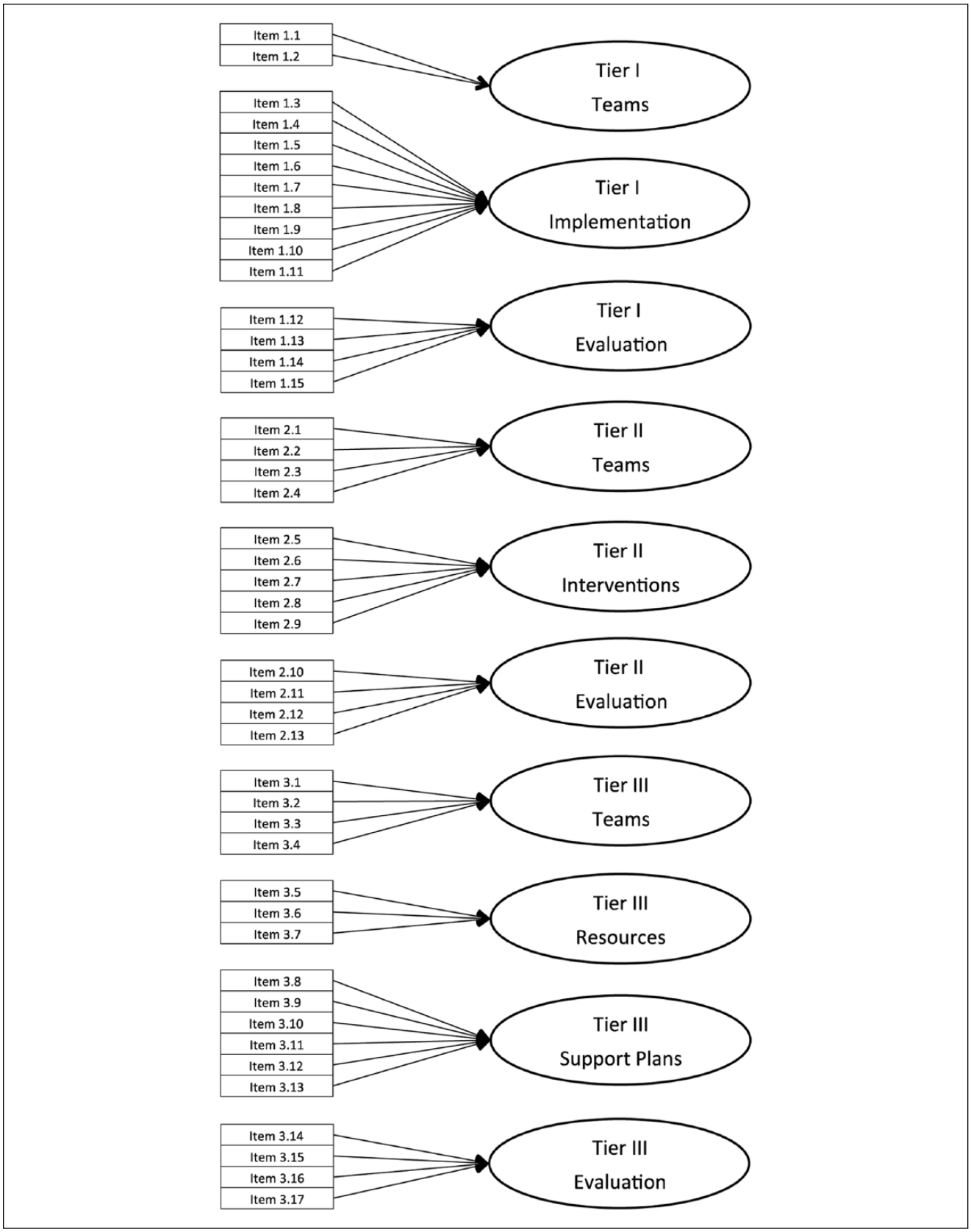

Figure 2. TFI 10-factor structure.
Note. TFI = Tiered Fidelity Inventory.

polychoric correlations and handles missing data well (Asparouhov & Muthén, 2010). We judged model fit of each model using the model fit indices and criteria recommended by Mueller and Hancock (2010). Because the models were nested, we were able to assess differences in model fit using the Mplus DIFFTEST function and its resulting chi-square difference statistic to compare the improvement in model fit with the added free parameters of the 10-factor model.

Results

For the three-factor model, apart from chi-square (χ^2 = 3272.314, $p < .001$), model fit indices were strong (comparative fit index = .993, Tucker-Lewis index = .993, root mean square error of approximation = .038 [90% CI = .036, .039]). Fit for the 10-factor model was similar but slightly stronger (χ^2 = 2376.579, $p < .001$, comparative fit index = .996, Tucker-Lewis index = .995, root mean square error of approximation = .031 [90% CI = .030, .033]). Results for all model fit indices for both the three- and 10-factor models surpassed the recommended criteria of Mueller and Hancock (2010).

The fit for the two models Results for the chi-square difference test (χ^2 = 505.454, $p < .001$) indicated statistically significantly stronger model fit for the 10-factor model over the three-factor model. Results indicated that although model fit was strong for both solutions, the 10-factor model had stronger model fit.

Discussion

Results from the CFA add to the existing results from the technical adequacy studies. These studies indicate that the TFI has demonstrated strong internal consistency and content validity for assessing SWPBIS implementation, statistically significant correlations with existing fidelity measures (i.e., concurrent validity), and high test-retest and interrater reliability. The CFA results presented here indicate that the revised TFI has a consistent factor structure with extremely strong model fit.

The findings support the recommendation that the TFI can be used to measure overall implementation, implementation by tier, and implementation by subscale. The results of the three-factor model indicate that, rather than one unified construct, each tier of SWPBIS measured on the TFI is distinct. Resultantly, teams can plan for and measure the implementation of Tiers I, II, and III independently. The results of the 10-factor model support the hypothesis that there are different components to SWPBIS implementation within each tier. Based on the CFA, teams are encouraged to use the subscales to focus on implementing all components of SWPBIS. As such, teams can identify the targeted areas for training and ongoing coaching support (e.g., evaluation).

Teams can use the action planning tool with the TFI to drive implementation and sustained use of SWPBIS. The TFI can be administered throughout various stages of the implementation cycle, from adoption and initial implementation to full implementation over multiple years. Teams looking to implement SWPBIS for the first time and teams planning to expand upon preexisting SWPBIS efforts will find the data useful for planning and strengthening implementation efforts. Because the TFI can be used at the subscale and item level, teams can also assess one core feature at a time and progress monitor in a targeted and specific manner. Because research has supported the connection between fidelity of implementation and valued outcomes (Durlak & DuPre, 2008; McIntyre et al., 2007; Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012), these results are important to supporting the ongoing effort to increase practitioners' use of fidelity of implementation measures and data-based action planning to increase the likelihood that valued student, staff, and school outcomes are achieved.

Not only do these results support the use of the TFI by school teams, the reliability and validity of the TFI indicate that it is appropriate for use in research related to implementation of the core features of SWPBIS. Fidelity of implementation is related to valued outcomes for students, teachers, and the school environment. Assessing the extent to which the core features of SWPBIS are being implemented with fidelity can help school teams achieve desired results and help researchers better understand how implementation and outcomes are related. The TFI allows teams and researchers to efficiently measure the fidelity with which the core features of SWPBIS are in place. The results of the CFA indicate researchers can also assess the relative importance of different tiers and subscales in improving student, staff, and school outcomes. This finding is important for ongoing research, training, and support related to the core features that produce student outcomes.

Limitations and Future Research

The results of the CFA are promising and hold positive implications for both school teams and researchers; however, it is important to note two limitations of this study. First, the TFI scores were self-reported by school teams. The researchers did not conduct any direct observations of the TFI administration process. Second, results from the large-scale pilot study indicated that TFI results were more reliable when an external SWPBIS coach facilitated the administration of the assessment. It is not known how many school teams used an external coach during TFI administration.

This study highlights areas for future research. First, researchers should explore the extent to which fidelity at each tier is associated with positive student outcomes. To date, no research study has linked scores on the TFI with valued social, academic, or behavioral outcomes for students. Second,

evaluating the relation between fidelity at various subscales and valued outcomes may also be considered. Third, assessing the validity of using the TFI subscales to deliver differentiated training and coaching would offer important insight into the use of the measure as a guide for durable implementation. Fourth, utilizing direct observational data to compare reported implementation to actual implementation would establish more validity for the TFI (Noell et al., 2005). Finally, exploring the value of the TFI as a formative, summative, and progress-monitoring tool is important. This information will help inform the use of fidelity measures as more than an end-of-year evaluation and help researchers understand how best to support teams in efficiently and effectively using fidelity data to guide SWPBIS implementation efforts at regular points during the school year.

Implications for Use

The TFI is a freely-available tool that school teams can use to plan for adoption, implementation, and sustained use of SWPBIS over time. The tool is designed to be efficient, simple to use, and effective in action planning for school teams. The tool is valid and reliable for measuring the extent to which the core features of SWPBIS are in place across all three tiers. The recommended practice is for teams to complete all three tiers of the TFI once per year and then use one of the tiers as a progress-monitoring tool for two or three additional times throughout the year (Algozzine et al., 2014). For example, a team may complete Tiers I, II, and III in the fall (with an external coach), and if Tier I is below 70%, they would complete Tier I in winter and spring (as a self-assessment). Our findings support the use of the TFI in this modular fashion; in addition teams can assess their implementation using the TFI's subscales, and researchers may use the subscales to identify which active ingredients are most closely related to improved outcomes.

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

Algozzine, R. F., Barrett, S., Eber, L., George, H., Horner, R. H., Lewis, T. J., . . . Sugai, G. (2014). *PBIS Tiered Fidelity Inventory*. Eugene, OR: OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. Available from http://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*. Eugene, OR: Educational and Community Supports, University of Oregon & University of South Florida.

Asparouhov, T., & Muthén, B. O. (2010). *Weighted least squares estimation with missing data*. Available from http://www.statmodel.com

Bond, G. R., Becker, D. R., Drake, R. E., & Vogler, K. M. (1997). A fidelity scale for the Individual Placement and Support model of supported employment. *Rehabilitation Counseling Bulletin*, *40*, 265–284.

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H.-W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, *2*, 75–87.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through School-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science*, *10*, 100–115.

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of School-Wide Positive Behavioral Interventions and Supports on the organizational health of elementary schools. *School Psychology Quarterly*, *23*, 462–473.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of School-Wide 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

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 40–49.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45.

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

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, *5*, 47–53.

Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the Community Youth Development Study. *American Journal of Community Psychology*, *41*, 235–249.

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

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.

Kincaid, D., Childs, K., & George, H. (2005). *School-wide benchmarks of quality*. Unpublished instrument, University of South Florida.

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.

McIntosh, K., Massar, M. M., Algozzine, R. F., George, H., Horner, R. H., Lewis, T. J., Swain-Bradway, J. (2017). Technical adequacy of the SWPBIS Tiered Fidelity Inventory. *Journal of Positive Behavior Interventions*, *19*, 3–13.

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.

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991-2005. *Journal of Applied Behavior Analysis*, *40*, 659–672.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315–340.

Mueller, R. O., & Hancock, G. R. (2010). Structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 371–384). New York, NY: Routledge.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nelson, M., Cordray, D., Hulleman, C., Darrow, C., & Sommer, E. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services & Research*, *39*, 374–396. doi:10.1007/s11414-012-9295-x

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.

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, *18*, 133–145.

Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining the implementation of school-wide positive behavior support. *School Psychology Review*, *35*, 245–259.

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