Development and Validation of a Measure Assessing Sustainability of Tier 2 and 3 Behavior Support Systems

Angus Kittelman
University of Oregon

Sterett H. Mercer
University of British Columbia

Kent McIntosh
University of Oregon

Rhonda N. T. Nese
University of Oregon

Kittelman, A., Mercer, S. M., McIntosh, K., & Nese, R. N. T. (2021). Development and Validation of a Measure Assessing Sustainability of Tier 2 and 3 Behavior Support Systems. *Journal of School Psychology, 85,* 140-154.

Author Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A180027 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.
Corresponding Author: Kent McIntosh, 1235 University of Oregon, Eugene, OR 97403, kentm@uoregon.edu

Abstract

To identify the most effective strategies for implementing and sustaining Tier 2 and 3 behavior support systems, a measure of general and tier-specific factors hypothesized to predict sustained implementation is needed. To address this need, we conducted two studies examining the construct validity of the *Advanced Level Tier Interventions Treatment Utilization and Durability Evaluation* (ALTITUDE) measure: one assessing the content aspects of construct validity (Study 1) and one assessing the structural and external aspects of construct validity (Study 2). In Study 1, participants included an expert panel of 26 members who provided iterative feedback during measure development. The results showed strong content representativeness (content validity index = .93) for assessing elements indicating sustainability. In Study 2, participants were school personnel from 646 schools who completed the measure. The results showed model fit was good for both a three-factor correlated model (CFI = .98, RMSEA = .06, SRMR = .05) and Bifactor S-1 model with correlated Tier 2 and Tier 3 specific residual factors (CFI = .98, RMSEA = .06, SRMR = .05). In addition, ALTITUDE latent factors were found to have both convergent and discriminant evidence in relation to concurrent fidelity of school-wide positive behavioral interventions and supports (SWPBIS) implementation at Tiers 1, 2, and 3. This construct validity evidence will support the use of the ALTITUDE in identifying the strongest tier-general and tier-specific predictors of sustained implementation of Tier 2 and 3 behavior support systems.

Keywords: SWPBIS; school-wide positive behavior interventions and supports; sustainability; Tier 2; Tier 3

Development and Validation of a Measure Assessing Sustainability of Tier 2 and 3 Behavior Support Systems

Districts and schools expend considerable resources each year in efforts to sustain evidence-based practices (EBPs). Unfortunately, the failure to sustain EBPs is a significant and reoccurring problem for schools (Nadeem & Ringle, 2016; Turri et al., 2016). As described in the field of implementation science, which is defined as the "study of the uptake of research findings and evidence-based practices" (Forman et al., 2013, p. 79), successful EBP implementation involves progression through several implementation stages (e.g., dissemination, adoption, initial implementation, and sustainability; Forman et al., 2013). Some of the reasons why school personnel struggle to sustain EBPs include a lack of administrator support and staff buy-in, limited resources (e.g., money, time), poor implementation fidelity, turnover in school personnel, and competing initiatives (Forman et al., 2013; Kittelman et al., 2020; Nadeem & Ringle, 2016; Turri et al., 2016).

School-Wide Positive Behavioral Interventions and Supports as a Framework for Sustaining EBPs

One approach that shows promise for sustaining EBPs in schools is implementing them within multi-tiered systems of support, such as school-wide positive behavioral interventions and supports (SWPBIS; Sugai & Horner, 2019). Within each tier, systems (e.g., teaming, data, professional development; McIntosh & Goodman, 2016) are intentionally designed to enhance the delivery of EBPs for all students (i.e., Tier 1), students requiring more support (i.e., Tier 2), and students with intensive and individualized needs (i.e., Tier 3; Horner & Sugai, 2015; Sugai & Horner, 2019). Prior work has identified a number of factors at each tier that may be related to the sustained implementation of behavior support systems (Debnam et al., 2012; Hagermoser

Sanetti et al., 2015; Lyon et al., 2019; McIntosh et al., 2013; McIntosh et al., 2018; Reddy et al., 2019).

Assessing Sustainability of Tier 1 Behavior Systems

To assist in identifying Tier 1 factors, a measure with validity evidence for assessing the sustainability of Tier 1 behavior support systems was developed (Hume & McIntosh, 2013; McIntosh et al., 2011). The measure, the *School-wide Universal Behavior Sustainability Index – School Teams* (SUBSIST; McIntosh et al., 2011), consists of 39 items that measure school (e.g., "A vast majority of school personnel [80% or more] support SWPBIS") and district factors (e.g., "There are adequate district resources [funding and time] allocated for SWPBIS") related to SWPBIS implementation. Using this measure, researchers identified factors related to the sustained implementation of Tier 1 behavior support systems (i.e., school teams using data for decision-making, district and school priority, and district capacity building; McIntosh et al., 2013; Mercer et al., 2014). For example, team use of data predicted sustained implementation of Tier 1 behavior support systems three years later, after controlling for initial implementation fidelity and school demographic characteristics (McIntosh et al., 2018).

Differences in Sustaining Tier 1 vs. Tier 2 and 3 Systems

Although it is possible that some of the factors predicting sustained implementation of Tier 1 could also predict sustained implementation of Tier 2 and 3 behavior support systems, Tier 2 and 3 systems and practices may differ enough from Tier 1 that other factors are more predictive. For example, Tier 2 and 3 behavior support systems include more frequent and intensive data collection and more tailored interventions, requiring more frequent meetings and specialized behavior expertise (Algozzine et al., 2014; Loman et al., 2010; McIntosh & Goodman, 2016). As described, team use of data has been shown in multiple studies to predict

sustained implementation of Tier 1 behavior support systems. It may be that this element generalizes to Tier 2 and 3 systems in that the data may change (e.g., screening vs. individual progress monitoring data), but the general actions and skills (i.e., regularly generating and reviewing graphs of implementation and student outcome data and making data-based decisions on whether to continue or modify supports when needed) still pertain. If so, the use of data would be a tier-general predictor of sustained implementation at all three tiers. However, some aspects of data use may be specific to Tiers 2 or 3, such as efficiently monitoring groups of students receiving Tier 2 support, or using functional behavior assessment to match interventions to student need and progress monitor intervention implementation daily (Algozzine et al., 2014). In these cases, these factors might be tier-specific predictors. For these reasons, a strong tool should be able to assess both tier-general and tier-specific sustainability factors.

Research from schools implementing SWPBIS has found that fewer schools measure implementation of Tier 2 and 3 systems and those schools that measure this implementation have lower scores than at Tier 1 (Debnam et al., 2012; Kittelman et al., 2018). As an example, during the 2016-17 school year, 8,467 schools assessed Tier 1 implementation fidelity using a measure that assesses fidelity across all three tiers (SWPBIS Tiered Fidelity Inventory; Algozzine et al., 2014); however, only 5,343 schools assessed Tier 2 fidelity, and only 3,327 schools assessed Tier 3 fidelity. Moreover, items that assessed implementation of behavior support systems across tiers (i.e., professional development systems, evaluation systems) were rated lower for Tier 2 than Tier 1 and lower for Tier 3 than Tier 2 (Kittelman et al., 2018). This provides some preliminary evidence that Tier 2 and Tier 3 behavior support systems are more difficult to implement and sustain and that more empirical research is needed. Additionally, although it seems likely that some factors at Tier 1 are predictive of Tier 2 and 3 behavior systems (i.e.,

professional development systems, evaluation systems), the intensity and resources required to sustain them are likely greater.

Identifying Potential Tier 2 and 3 Sustainability Factors

Over the last decade, several descriptive studies have identified unique factors that may be predictive of the sustainability of Tier 2 and 3 EBPs and behavior support systems (Bambara et al., 2012; Bambara et al., 2009; Loman et al., 2010; Woodbridge et al., 2014). For example, Loman and colleagues (2010) conducted interviews with 29 school personnel from different elementary schools to identify factors perceived to be relegated to the sustained implementation of a Tier 2 EBP (First Step to Success; Walker et al., 1998). Loman et al. found that schools who sustained implementation of the Tier 2 EBP had higher ratings of school capacity (e.g., staffing, budget, orientation, documentation) and screening systems (e.g., request for assistance, selection of students for Tier 2 EBPs) as compared to schools that did not sustain in the first year of implementation. In addition, two of the features rated as most powerful for sustaining the Tier 2 EBP were dedicated resources (e.g., funding, release time) and training/orientation activities. Loman et al. provided additional evidence that factors beyond Tier 1 are significant for sustaining Tier 2 practices and behavior support systems. For example, as all students are expected to receive Tier 1 (universal) support, screening systems become critical for identifying and referring students for EBPs beyond Tier 1. Also, school capacity is likely to be more critical for sustaining Tier 2 and 3 behavior support systems due to the additional resources needed to implement moderately intensive EBPs (i.e., Tier 2) and highly intensive EBPs (i.e., Tier 3).

Other descriptive studies have identified additional variables that may predict the sustainability of Tier 3 practices and systems (Bambara et al., 2012; Bambara et al., 2009). For example, Bambara et al. (2012) surveyed 293 school personnel about their experiences

supporting the implementation of Tier 3 systems and practices. Some of the key variables perceived to be the most supportive of Tier 3 EBPs and systems included collecting individual student data for progress monitoring and decision making, having school personnel observe positive student outcomes, and ensuring functional assessment and behavior support strategies are designed to fit classroom and school contexts (Bambara et al., 2012).

Assessing Sustainability of Tier 2 and 3 Behavior Systems

Although these descriptive studies provide initial insights into factors that may be important for the sustainability of all tiers, and unique to Tier 2 and 3 systems, these insights are relatively limited. Specifically, there are no prospective longitudinal studies that have examined factors predicting sustained implementation of Tier 2 or 3 systems. In addition, no studies have examined whether sustainability factors are tier specific (i.e., factors predicting implementation of one tier only) vs. tier general (i.e., factors predicting implementation of all tiers). To conduct research identifying tier-specific and tier-general factors that predict sustained implementation of Tier 2 and 3 behavior support systems, a measure is needed to assess these factors with validity.

The purpose of the current two studies was to develop and validate the *Advanced Level Tier Interventions Treatment Utilization and Durability Evaluation* (ALTITUDE) for use in assessing sustainability of Tier 2 and 3 behavior systems. We anticipate that this preliminary work on the ALTITUDE will better enable researchers to identify malleable factors that can predict whether schools can sustain implementation of Tier 2 and 3 behavior support systems. Although we view the ALTITUDE as primarily useful for researchers, we believe future research with the ALTITUDE will yield important practical insights, including determining whether tier-general vs. tier-specific factors are most important for sustained implementation, which has implications for staffing, professional development, and other technical assistance activities. If

tier-general factors are less important, then more generic training and coaching associated with data-based decision making may be less effective than training and coaching for tier-specific skills (e.g., how to conduct progress monitoring for individual students on Tier 3 EBPs). Conversely, it is possible that factors promoting sustained implementation of a specific tier are detrimental to sustained implementation of a different tier. For example, personnel receiving professional development on individualizing behavior support plans at Tier 3 might over individualize Tier 2 interventions, leading to inefficient implementation.

Aims of the Current Studies

To conduct sound research identifying factors predicting sustainability of Tier 2 and 3 behavior support systems, the ALTITUDE should have evidence of construct validity (Messick, 1995). According to Messick (1995), there are multiple aspects of construct validity (i.e., content, substantive, structural, generalizability, external, and consequential) that assist in the interpretation and use of scores generated from a particular measure. To evaluate some of these important aspects of construct validity, we report the findings of two sequential studies. In Study 1, we formed an expert panel of researchers and practitioners to conduct an iterative process of developing, refining, and then evaluating the content relevance of ALTITUDE items. In Study 2, we administered the ALTITUDE to a national sample of school personnel to assess the structural and external aspects of construct validity for the ALTITUDE. We hypothesized that the ALTITUDE would have the following properties:

1. There would be evidence of the content aspect of construct validity for the ALTITUDE (Messick, 1995), meaning that the ALTITUDE assesses features experts consider to be important for sustainability.

2. There would be evidence of structural aspects of construct validity for the ALTITUDE, specifically a clear factor structure that can represent both tier-specific and general factors related to Tier 2 and 3 systems.
3. There would be evidence of external aspects of construct validity for the ALTITUDE via significant convergent and smaller discriminant correlations with SWPBIS fidelity of implementation assessments at each tier. Because there is little prior empirical work on the relative importance of tier-general vs. tier-specific sustainability factors in relation to fidelity or sustained implementation, we did not have a specific hypothesis regarding whether tier-general or tier-specific sustainability factors will be more strongly related to Tier 2 or Tier 3 SWPBIS fidelity. However, we expected both tier-general and tier-specific sustainability factors to be related to fidelity. By contrast, because tier-specific sustainability factors logically should be most strongly related to implementation of specific tiers, we had hypotheses for the tier-specific sustainability factors. We hypothesized that Tier 2-specific sustainability factors on the ALTITUDE would be more strongly related to Tier 2 fidelity than Tier 3 or Tier 1 fidelity, and that the Tier 3-specific sustainability factors would be more strongly related to Tier 3 fidelity than Tier 2 or Tier 1 fidelity.

Study 1: Item Development and Content Representativeness

In Study 1, we administered three surveys to an expert panel to develop items for the ALTITUDE. Survey 1 was administered in August of 2018, Survey 2 in September of 2018, and Survey 3 in October of 2018. The goal of Survey 1 was to generate ideas for ALTITUDE item content. We used an open-ended Delphi survey technique to generate as many potential items as possible to ensure maximum coverage, while also limiting potential biases by not including items

the authors had identified (Hasson et al., 2000). In Survey 2, the expert panel provided detailed feedback on the items generated from the Survey 1 and suggested additional items for inclusion. Finally, we used Survey 3 to assess the content representativeness of the ALTITUDE items and measure as a whole.

Participants

Twenty-six individuals with extensive experience supporting the implementation of Tier 2 and 3 behavior support systems and practices participated in the expert panel. Expert panel members were identified and recruited via email invitations by the third author because of their affiliation with the Center on PBIS (www.pbis.org) and/or experience implementing Tier 2 and 3 behavior practices in schools. Criteria for membership on the expert panel consisted of being either (a) a researcher with at least four published articles within the past five years that examined implementation of school-based Tier 2 and 3 behavior practices or (b) a technical assistance provider with at least 10 years of experiences supporting implementation of Tier 2 and 3 practices across multiple schools or districts. Panel membership exceeded the recommended criterion of two to 20 (Gable & Wolf, 1993; Rubio et al., 2003). Fifteen (58%) of the members were female. Twenty-one (81%) had a doctoral degree, four (15%) had a master's degree, and one (4%) did not report education. Most participants identified as White (n = 21, 81%). Fifteen (58%) identified their primary role as being a researcher or project director, six as practitioners (23%), two (8%) as a researcher and practitioner equally, one (4%) as a program director, one (4%) as an evaluator, and one (4%) did not specific their primary role. The average number of years supporting schools in implementing SWPBIS as a practitioner and/or researcher was 15 years (range = 3 – 40 years) and the average number of years supporting schools in implementing Tier 2 and 3 behavior support systems as a practitioner and/or researcher was 13

years (range = 2 – 30 years). Seven (27%) expert panel members worked in the U.S. West region, seven in the Midwest (27%), five in the South (19%), four (15%) in the Northeast, one (4%) in multiple regions, and two did not indicate a region. The was a 100% response rate for Survey 1, a 68% response rate for Survey 2, and an 81% response rate for Survey 3.

Item Generation

Procedure

Expert panel members were recruited through email invitations to participate in a panel focused on developing a measure assessing factors that predict sustained implementation of Tier 2 and 3 systems. A Delphi survey technique was used to gather expert panel feedback during measure development (Hasson et al., 2000). Specifically, the Delphi process occurs over the course of several rounds of survey administrations and assists researchers to gather, summarize, and reach consensus among panel members (Hasson et al., 2000). As is customary in the early rounds of the Delphi process (Hasson et al., 2000), the first round began with open-ended items to assist expert panel members in generating ideas for measure content. We held an in-person meeting (two members participated via videoconference) and provided the expert panel with verbal and written instructions on the aims and objectives of the study, as well as the responsibilities of the expert panel. We then provided expert panel members with individual response packets that included a copy of the items in the Tier 2 and 3 scales of the SWPBIS Tiered Fidelity Inventory (TFI; Algozzine et al., 2014) and two open-ended survey questions: "Given TFI Tier 2 items, what facilitating items would predict high scores on these fidelity items?" and "Given TFI Tier 3 items, what facilitating items would predict high scores on these fidelity items? Please repeat any Tier 2 items that would also apply to Tier 3." The TFI was provided to ensure experts did not generate items redundant with the TFI, given that a measure

of sustainability should predict fidelity but not be an assessment of fidelity (i.e., the measure should not include TFI items). After expert panel members individually completed the response packets, they were invited to share items they generated and discuss ideas as they emerged. The discussion allowed panel members to generate additional items.

Results

The expert panel generated a total of 247 items, suggestions, and comments. We then reviewed these data to combine similar items and delete items that assessed fidelity instead of sustainability. An example of two items that were combined due to similarity included "teacher buy-in and participation in Tier 2 interventions on a daily basis" and "staff buy-in for specific common interventions." An example of two items that were deleted because they directly assessed fidelity included "Implementing Tier 1 with high fidelity" and "classroom practices in place school-wide." This process resulted in a preliminary list of 38 potential ALTITUDE items.

Revisions and Generation of Additional Items

Procedure

Next, expert panel members were invited to participate in an anonymous online survey to rate and generate new items. For each of the potential ALTITUDE items generated in Survey 1, the expert panel was asked to indicate whether to "retain," "remove," or "revise" the item. If expert panel members selected "revise," they were prompted to provide open-ended feedback on how to improve the item (e.g., comments, text edits). In addition, for each of the 38 proposed items, the expert panel was asked whether the item should be asked in reference to only one tier (i.e., Tier 2 or Tier 3) or in reference to both tiers. Specifically, each panel member made one selection per item from the following options: "Tier 2 only (item not relevant to Tier 3)," "Tier 3 only (item not relevant to Tier 2)," "Tier 2 and 3 as separate items (item relevant to both tiers,

and schools are likely to have different responses by tier)," or "Tier 2 and 3, combined in one item (item relevant to both tiers, and schools are likely to respond similarly across tiers)." At the end of the survey, we asked expert panel members to identify any additional items that may predict sustained implementation of Tier 2 and 3 systems.

Results

Based on expert panel responses, we removed four of the 38 items and revised wording for the other 34 items. An example of an item that was removed because 53% of the panel members who completed survey 2 recommended we either "remove" or "revise" the item was "*The Tier 2 and 3 team(s) are willing to discontinue implementation of ineffective practices (to free up resources for more effective options).*" In addition, an example of an item we revised was "*School personnel value partnering with families in Tier 2 and 3 supports and problem solving,*" which was revised to "*School personnel build strong partnerships with families to support students with Tier 2 and 3 behavior needs*." Item revisions and consensus was reached by the authors by collaboratively reviewing survey feedback from the expert panel. In addition, we generated seven new items based on the open-ended feedback from expert panel members. Of the 41 resulting items, 17 items were identified by the expert panel as relevant to both Tiers 2 and 3, 13 items were specific to Tier 2, and 11 items were specific to Tier 3.

Content Representativeness

Procedure

As the last step, we invited expert panel members to participate in Survey 3. We asked expert panel members to rate the extent they perceived each of the items to be predictors of sustainability of Tier 2 and 3 behavior support systems using a sliding scale with the following response anchors: 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Agree*, and 4 = *Strongly Agree*.

Expert panel members could also select "Don't Know/NA" for each item. In addition, there was an optional open-ended prompt for each item for expert panel members to provide any final comments or recommended revisions. At the end of the survey, expert panel members had the opportunity to provide additional comments on the overall measure.

Data Analysis

Content representativeness of the ALTITUDE measure was assessed by calculating a content validity index (CVI) score across the 41 individual items and an overall CVI score for the measure (Rubio et al., 2003). CVI for individual items was calculated by counting the number of expert panel members who rated each item as *Agree* or above and dividing it by the total number of experts, and CVI for the overall measure was calculated by averaging the individual item CVI scores (Rubio et al., 2003). Scores of .80 or higher are considered adequate for new measures (Davis, 1992). If individual item CVI scores were below .80, we inspected them to determine whether to revise or remove those items.

Consistent with the literature (Davis, 1992; Lynn, 1986), interrater agreement (IRA) among expert panel members was calculated by dichotomizing the ratings of *Strongly Disagree* to *Disagree* versus *Agree* to *Strongly Agree*. IRA for individual items was calculated by dichotomizing each expert panel member's score as either disagree or agree, and then reporting the percentage of the two that was higher for each item. As such, the floor for this index was .50. IRA for the overall measure was calculated by counting the number of items with an IRA at or above .80 and dividing it by the total number of items. This total IRA approach is recommended when there are more than five members in an expert panel (Lynn, 1986). Although there were little missing data across expert panel scores on the individual items (2%; 10 items had one, two items had two, and one item had three missing ratings), we conservatively coded all missing

values as disagreements that the item predicted implementation and sustainability of Tier 2 and 3 systems.

Results

Expert panel members' ratings were reliable (IRA = .95), and the overall CVI score for the measure was .93, indicating strong content aspects of construct validity for the ALTITUDE in assessing implementation and sustainability of Tier 2 and 3 systems. In terms of individual item CVI scores, only two items fell below the validity criterion of .80, and we removed both items from the ALTITUDE. Based on open-ended responses from the expert panel, we re-worded or edited 22 of the remaining items for clarity. For example, one item was originally written as "*The school's Tier 3 data are efficient for school personnel to collect"* and was changed to "*Individuals implementing Tier 3 behavior interventions can collect fidelity and student progress data without too much effort."* Table 2 details the ALTITUDE items resulting from the measure development process in Study 1. At the conclusion of Study 1, the ALTITUDE included 39 items, consisting of 15 items related to Tiers 2 and 3, 13 items related to Tier 2 only, and 11 items related to Tier 3 only.

Study 2: Structural Aspects and Relations with Fidelity

Participants

Participants were drawn from a national sample of educators from 646 schools completing the ALTITUDE as part of a longitudinal study. The majority of participants were school SWPBIS team leaders/facilitators/internal coaches (n = 329, 51%), followed by school administrators (n = 172, 27%), school personnel (n = 77, 12%), external/district/regional SWPBIS coaches (n = 54, 8%), PBIS team members (n = 6, 1%), specialists/related service providers (e.g., school psychologist, social worker; n = 5, 1%), other (data coordinator, learning

environment specialist, $n = 2$, > 1%), and a district coordinator ($n = 1$, > 1%). The 646 schools were from 23 U.S. states, representing all four census regions. According to data obtained from the National Center for Educational Statistics (NCES) database for the 2016-17 (most recent available) school year, 50% of the participating schools were elementary schools, and 70% were located in cities or suburbs. Additional NCES school demographic information is in Table 1.

Participants for Study 2 were recruited by state leaders affiliated with the Center on PBIS (www.pbis.org). Recruitment included state leaders working with district PBIS coaches to identify and recruit school participants or by recruiting participants directly during existing SWPBIS training events. Participants needed to meet the following inclusion criteria to participate in this study: (a) be working in or with (i.e., district coach) a public elementary, middle, or high school engaged in SWPBIS implementation (Forman et al., 2013); (b) have knowledge of their school's SWPBIS implementation efforts; and (c) be measuring SWPBIS implementation fidelity.

Measures

ALTITUDE

For each of the 39 items, participants were asked to rate whether each item was true for their schools right at the present time on a sliding scale with the following response anchors: 1 = *Not True*, 2 = *Partially True*, 3 = *Mostly True*, and 4 = *Very True*. Respondents could also indicate a response of "Don't Know/NA" on each item—such responses were treated as missing in the analyses.

SWPBIS Fidelity of Implementation

The TFI (Algozzine et al., 2014) is the most widely-used measure assessing implementation fidelity of SWPBIS at all three tiers (Kittelman et al., 2018). The TFI includes a

total of 45 items (Overall Scale), with 15 items assessing Tier 1 (Scale 1), 13 items assessing Tier 2 (Scale 2), and 17 items assessing Tier 3 (Scale 3). Studies have found the TFI to have strong evidence of content (CVI = .92) and structural aspects of construct validity (i.e., comparative fit index [CFI] = .99, root mean square error of approximation [RMSEA] = .04, across a three-factor CFA model; Massar et al., 2017; McIntosh et al., 2017), reliability (overall measure, α = .96; across individual tiers, α = .87 to .98), and convergent correlations with other SWPBIS implementation fidelity measures (*r*s = .51 to .75). If schools were reporting on implementation fidelity using a measure other than the TFI and only measuring Tier 1 implementation fidelity, their fidelity scores were excluded when examining relations between the ALTITUDE measure and TFI implementation fidelity scores.

Procedure and Data Analysis

To assess the factor structure of the ALTITUDE, we randomly divided the sample of 646 schools into two subsamples of equal size. With the first subsample (*n* = 323 schools), we conducted an exploratory factor analysis (EFA) based on the polychoric correlation matrix to account for the ordered categorial item response format, with an oblique (promax) rotation. Our decision for the number of factors to extract was informed by two variants of parallel analysis using the FACTOR program (Lorenzo-Seva & Ferrando, 2006): (a) Horn's (1965) principal component-based method in which actual eigenvalues are compared to averaged eigenvalues from randomly generated datasets, and (b) Timmerman and Lorenzo-Seva's (2011) method based on minimum rank factor analysis of the polychoric correlation matrix in which the percentages of common factor variance for the actual data set are compared to the average percentages for random permutations of the original data set.

Informed by the EFA results, we deleted seven items, as detailed in the Results. Specific reasons for deletion included: (a) an unusually high amount of missing data (1 item), indicating that participants may have had difficulty answering the item; (b) large cross-loadings on factors that were less logically related to item content (2 items); and (c) content overlap with other items on the same subscale (4 items). Following these decisions, we conducted an ordinal confirmatory factor analysis (CFA) using the second subsample (n = 323). Because the results of the parallel analysis for the EFA were not definitive, we evaluated a unidimensional model plus two- and three-factor models. Also, because we wanted for theoretical reasons to represent general vs. tier-specific aspects of sustainability, we fit Bifactor S-1 models (Eid et al., 2017), in which all ALTITUDE items load on a general Tier 2 and 3 sustainability reference factor, with items assessing tier-specific sustainability factors also loading on correlated specific residual factors (i.e., variance unrelated to the general Tier 2 and 3 sustainability factor). Unlike traditional applications of Bifactor models that have all items loading on both a general and specific factor and some applications of Bifactor S-1 models in which one specific factor is set as the general reference factor for theoretical reasons (Burns et al., 2020), the ALTITUDE has a directly measured general factor, thereby avoiding some limitations of traditional Bifactor models (i.e., insufficiently defined general factors and a tendency to yield anomalous results; Eid, 2020; Eid et al., 2017) and the need to compellingly justify a specific factor as an appropriate reference factor (Burke & Johnston, 2020). Reporting the results of both multidimensional correlated factor models and Bifactor S-1 models has been recommended due to their interpretability for multidimensional assessment instruments (Eid, 2020). Informed by the parallel analysis results, we fit one Bifactor S-1 model with one specific factor beyond the general sustainability factor (see Figure 1) and another Bifactor S-1 model with two specific factors beyond the general factor

(see Figure 2). We report several indicators of model fit, with the following values providing evidence of good model fit to the data (Hu & Bentler, 1999): CFI $\geq$ .95, RMSEA $\leq$ .06, and the standardized root mean squared residual (SRMR) $\leq$ .08.

After evaluating the fit of alternative CFA models, we evaluated subscale reliability based on ordinal α (Zumbo et al., 2007). We also calculated a hierarchical omega (ω_H) reliability coefficient (McDonald, 1999) based on the results of a Bifactor S-1 CFA model to determine the proportion of total variance in the ALTITUDE attributable to a general Tier 2 and 3 sustainability factor, plus hierarchical omega subscale (ω_{HS}) coefficients to determine the proportion of reliable variance in the tier-specific sustainability factors after accounting for variance attributable to a general Tier 2 and 3 sustainability factor. In addition, we evaluated test-retest reliability for each subscale based on a subsample of participants from 16 schools with two ALTITUDE completions at an average of 10 days apart.

We also examined the extent to which ALTITUDE factors correlated with SWPBIS implementation fidelity by tier. TFI implementation fidelity scores were obtained through PBIS Assessment (www.pbisapps.org), which is a free web application maintained at the University of Oregon in the research unit of Educational and Community Supports. PBIS Assessment is used by SWPBIS teams (with assistance from a SWPBIS district/regional coordinator) to enter and review their SWPBIS implementation fidelity scores for implementation progress monitoring. SWPBIS teams can assess and enter their schools' implementation fidelity scores directly into PBIS Assessment at any time during a school year. For the purposes of this study, we used TFI scores from each school's last TFI assessment for the 2018-19 school year (for each tier) that was completed with the guidance of an external SWPBIS coach. Of the 646 schools completing the ALTITUDE, 428 completed the Tier 1 scale of the TFI, 443 completed the Tier 2 scale, and

289 completed the Tier 3 scale. For schools without TFI data on a particular tier, we did not know if data were missing because the school was not implementing that particular tier or because they did not assess fidelity for that tier despite implementing that tier. Thus, we were unable to determine if data were missing not at random (MNAR) or were missing at random (MAR), and, regardless, the percentages of missing data were very high for most missing data handling approaches (e.g., 55% for Tier 3 TFI). For these reasons, we used listwise deletion by TFI tier, based on TFI data availability, as the least limited option for handling missing TFI data. In these analyses, a latent TFI factor (separate models by tier), with items as ordinal indicators, was correlated with latent ALTITUDE factors, also with ordinal item indicators, both as a correlated multidimensional model and Bifactor S-1 model in separate analyses.

All analyses were conducted in M*plus* 8.4 (Muthén & Muthén, 2017) using the mean and variance adjusted weighted least squares (WLSMV) estimator with adjusted standard errors to account for the clustering of schools in districts using the COMPLEX option. Consistent with recommendations for ordinal factor analysis models (Shi et al., 2020), we used multiple imputation (m = 1,000 data sets) to handle item-level missing data on the ALTITUDE (see percentages in Table 3) in the CFAs and correlational models. EFA with multiple imputation is not available in Mplus; thus, we handled missing data using the WLSMV estimator in the EFA.

Results

EFA

Before conducting the EFA, one item (#7) was excluded due to an abnormally large percentage of missing responses (30% compared to 2% and 7% on adjacent items), particularly given its placement early in the measure (see Table 3 for missing data information for all items). It is possible that the large percentage of missing responses for item #7 was because school-level

participants were unaware of what criteria their districts used to hire teachers with skills in implementing Tier 2 and 3 behavior interventions, or their district may not have included this component as part of their hiring procedures.

The EFA (with 38 items) on the first randomly selected subsample (n = 323) indicated a two- or three-factor solution depending on the parallel analysis method. Specifically, Horn's (1965) method indicated that three factors should be extracted, with the first three eigenvalues greater than eigenvalues calculated from randomly generated data (actual eigenvalues = 22.21, 2.48, 1.75, 1.10; eigenvalues on random data = 1.91, 1.79, 1.71, 1.64). By contrast, Timmerman and Lorenzo-Seva's (2011) method indicated that two factors should be extracted, with the actual percentages of common factor variance greater than the percentages based on randomly permuted data for the first two factors (actual percentages = 58.86, 5.77, 4.16; random percentages = 6.05, 5.64, 5.33). In the two-factor solution, items related to general Tier 2 and 3 sustainability and items specific to Tier 2 sustainability both primarily loaded on the first factor, with items specific to Tier 3 sustainability primarily loading on the second factor. In the three-factor solution, items with large loadings on the first factor assessed Tier 2 and Tier 3 systems as a whole, items with large loadings on the second factor assessed Tier 2 specifically, and items with large loadings on the third factor assessed Tier 3 specifically. Promax-rotated loadings for the three-factor solution are presented in Table 3. Both the two- and three-factor solutions were compared in subsequent CFAs.

Item Decisions

Following the EFA, we made decisions on which items to retain in the final ALTITUDE based on the three-factor pattern of EFA loadings and consideration of item content. Based on these deliberations, two items were excluded (#19 and 28) because they had larger loadings on

the combined Tier 2 and 3 and/or specific Tier 3 factor even though the items asked about Tier 2 sustainability. In addition, even though factor loadings were acceptable, four items (#18, 22, 30, and 33) were excluded for parsimony because the items had substantial content overlap with at least one other item on the same subscale. These decisions resulted in the retention of 32 final ALTITUDE items, as detailed in Tables 2 and 3.

CFA

Following these item retention decisions, responses on the 32 items for the second randomly selected subsample (n = 323) were included in five ordinal CFAs with the following factor structures: (a) all items loading on one factor; (b) two correlated factors, with general Tier 2 and 3 items and Tier 2 specific items loading on the same factor, and Tier 3 specific items loading on the other factor; (c) three correlated factors, with each item loading on only one factor, i.e., Tier 2 and 3 combined items loading on a combined factor, Tier 2 items loading on a Tier 2 factor, and Tier 3 items loading on a Tier 3 factor; (d) a Bifactor S-1 model, with all items loading on a general Tier 2 and 3 reference factor and items specific to Tier 3 also loading on a Tier 3 specific residual factor (see Figure 1); and (e) a second Bifactor S-1 model, adding a Tier 2 specific residual factor correlated with the Tier 3 specific residual factor (see Figure 2). Fit indices for the five models are presented in Table 4. In general, fit was better for three-factor models, either in correlated multidimensional or Bifactor S-1 models, compared to two- and one-factor models. Model fit was good for the three-factor correlated model, χ^2 (461) = 1007.26, p < .001; CFI = .98; RMSEA = .06; and SRMR = .05; model fit was also good for the Bifactor S-1 model with the correlated Tier 2 and Tier 3 specific residual factors in addition to the general Tier 2 and 3 factor, χ^2 (445) = 1002.90, p < .001; CFI = .98; RMSEA = .06; and SRMR = .05. Factor loadings for these two CFA models are presented in Table 3.

Reliability

Reliability, based on calculations of ordinal α (Zumbo et al., 2007), was strong for all three sustainability subscales: general Tier 2 and 3 (14 items, α = .95), Tier 2 specific (9 items, α = .95), and Tier 3 specific (9 items, α = .97). Test-retest reliability was strong for the general Tier 2 and 3 subscale (r = .89, p < .001), the Tier 2 specific subscale (r = .90, p < .001), and the Tier 3 specific subscale (r = .95, p < .001). Based on the Bifactor S-1 model with two specific residual factors, we calculated the following hierarchical omega coefficients: ω_H = .92 for the general Tier 2 and 3 factor, ω_{HS} = .22 for the Tier 2 specific residual factor, and ω_{HS} = .40 for the Tier 3 specific residual factor. These coefficients indicated that 92% of the variance on the ALTITUDE can be attributed to a general Tier 2 and 3 sustainability factor, with small amounts of reliable variance remaining on the specific Tier 2 and Tier 3 items after accounting for the general sustainability factor (22 and 40%, respectively).

Relations with Fidelity

Correlations among the ALTITUDE subscales, both as three correlated latent factors and as one general plus two residual specific factors, and TFI fidelity factors at each tier are presented in Table 5.

Tier 1 Fidelity. Fit was good for the model correlating TFI Tier 1 fidelity (n = 428) with the three correlated ALTITUDE factors, χ^2 (1028) = 1457.22, p < .001, CFI = .98, RMSEA = .03, SRMR = .06. Correlations with the TFI Tier 1 factor were small to moderate for the ALTITUDE Tier 2 and 3 general (r = .45, p < .001), Tier 2 specific (r = .46; p < .001), and Tier 3 specific (r = .27; p < .001) factors. Of note, the three ALTITUDE factors were highly correlated (rs = .74 to .85, all p < .001).

Fit was also good for the model correlating the TFI Tier 1 factor with the one general and two tier-specific ALTITUDE factors in a Bifactor S-1 configuration, χ^2 (1012) = 1434.19, p < .001, CFI = .98, RMSEA = .03, SRMR = .06. In this model, the ALTITUDE Tier 2 and 3 general factor correlated with the TFI Tier 1 factor at r = .45 (p < .001), and the Tier 2 specific residual factor had a small correlation with Tier 1 fidelity (r = .15, p = .019). By contrast, the Tier 3 specific residual factor was unrelated to Tier 1 fidelity (r = -.10, p = .062). The ALTITUDE Tier 2 and Tier 3 specific residual factors were positively correlated (r = .38, p < .001).

Tier 2 Fidelity. Fit was good for the model correlating TFI Tier 2 fidelity (n = 443) and three correlated ALTITUDE factors, χ^2 (939) = 1415.19, p < .001, CFI = .98, RMSEA = .03, SRMR = .05. In this model, the ALTITUDE Tier 2 factor had the strongest correlation with Tier 2 fidelity (r = .48, p < .001), although correlation magnitude was roughly similar across the ALTITUDE factors, with the general Tier 2 and 3 sustainability factor at r = .43 (p < .001) and the Tier 3 sustainability factor at r = .38 (p < .001).

Fit was also good with the ALTITUDE in a Bifactor S-1 configuration, with χ^2 (923) = 1391.27, p < .001, CFI = .98, RMSEA = .03, SRMR = .05. In the Bifactor S-1 model, the general Tier 2 and 3 sustainability factor (r = .44, p < .001) and the Tier 2 specific residual sustainability factor (r = .20, p = .001) were positively correlated with Tier 2 fidelity, but the Tier 3 specific residual sustainability factor was unrelated to Tier 2 fidelity (r = .09, p = .145).

Tier 3 Fidelity. Fit for the correlated ALTITUDE factors with TFI Tier 3 fidelity (n = 289) was good, with χ^2 (1121) = 1425.15, p < .001, CFI = .98, RMSEA = .03, SRMR = .06. In this model, the ALTITUDE Tier 3 sustainability factor had the largest correlation with Tier 3 fidelity (r = .41, p < .001), with the general Tier 2 and 3 sustainability factor at r = .34 (p < .001) and the Tier 2 sustainability factor at r = .24 (p < .001).

Fit with the ALTITUDE factors in a Bifactor S-1 configuration was also good, with χ^2 (1105) = 1403.41, $p < .001$, CFI = .98, RMSEA = .03, SRMR = .06. In this model, both the ALTITUDE general Tier 2 and 3 sustainability factor ($r = .35$, $p < .001$) and the Tier 3 specific residual factor ($r = .23$, $p = .001$) were related to Tier 3 fidelity, but the Tier 2 specific residual sustainability factor was unrelated to Tier 3 fidelity ($r = -.11$, $p = .186$).

Discussion

The purpose of these two studies was to develop and evaluate the construct validity of the ALTITUDE. This is an important step toward the ALTITUDE being defensibly used to identify contextual variables that predict sustained implementation of Tier 2 and 3 behavior support systems in schools. Overall, the studies provide supportive evidence for the ALTITUDE in relation to several aspects of construct validity.

For Study 1, we utilized an expert panel comprised of researchers and technical assistance providers to support measure development. Using a Delphi survey technique, we conducted three rounds of surveys to develop, refine, and evaluate the ALTITUDE items. Based on the results of Survey 3, we found the measure to contain strong content representativeness (CVI = .93). For Study 2, we administered the ALTITUDE survey to a large national sample of schools implementing SWPBIS. We conducted EFA and CFAs to assess the structural aspects of the measure and examine relations between the ALTITUDE and TFI implementation fidelity subscales (external aspects of validity). We found evidence of the structural aspects and that the ALTITUDE assesses both tier-general and tier-specific Tier 2 and 3 factors related to sustainability. There was also evidence of strong internal consistency for the ALTITUDE factors and some evidence of generalizability of responses over time (i.e., test-retest reliability across 10 days on average). Finally, we also found evidence of predictable relations for the ALTITUDE

with SWPBIS implementation fidelity, including evidence of convergent and discriminant relations with concurrent SWPBIS implementation fidelity assessments at each tier.

Regarding relations with SWPBIS fidelity, the ALTITUDE general Tier 2 and 3, Tier 2 specific, and Tier 3 specific subscales, when specified as correlated factors, all had small to moderate correlations with fidelity at all tiers. We also investigated relations with fidelity with the ALTITUDE factors in a Bifactor S-1 specification, in which the Tier 2 and 3 specific factors represented tier-specific variance in sustainability with variance related to general Tier 2 and 3 sustainability removed. In this configuration, only the general Tier 2 and 3 ALTITUDE factor was related to fidelity of all three SWPBIS tiers, with the Tier 2 specific factor related to both Tier 1 and 2 fidelity and the Tier 3 specific factor only related to Tier 3 fidelity. These patterns of relations for the tier-specific sustainability factors, in the Bifactor S-1 configuration, provide evidence of both convergent and discriminant relations that we expected at the outset of the study.

In sum, these results provide evidence that the subscales of the ALTITUDE assess both tier-specific and tier-general aspects of sustainability in predictable ways. Although model fit was good for most of the alternative CFA models, model fit was strongest for the two 3-factor models (Model 3 and 5; Table 4). In particular, the S-1 model with the correlated Tier 2 and Tier 3 specific residual factors (Model 5) has strong conceptual fit because it demonstrates that the Tier 2 and 3 specific residual factors contained unique variance not shared by the general Tier 2 and 3 sustainability reference factor. This result provides empirical evidence, and aligns with our hypotheses, that factors (e.g., district or school-level) predictive of sustained implementation of Tier 2 and 3 behavior systems can be tier-general or tier-specific.

Findings from these two studies are an important initial step towards extending previous research examining factors predictive of the sustained implementation of Tier 1 behavior support systems. Researchers previously developed a measure (McIntosh et al., 2011; McIntosh et al., 2018) that was later used to identify key district and school factors predictive of Tier 1 behavior support systems in schools (McIntosh et al., 2018). The current study extended this research by developing a research measure that can be used to identify tier-general and tier-specific aspects of Tier 2 and Tier 3 behavior support systems with validity. This is an important area for future research because thousands of schools in the U.S. are implementing these advanced tiers (Kittelman et al., 2018) and much of what we know about variables that are important for the sustained implementation of Tier 2 and 3 behavior support systems has been gathered through descriptive and cross-sectional studies (Bambara et al., 2012; Bambara et al., 2009; Loman et al., 2010; Woodbridge et al., 2014). Using the ALTITUDE, we anticipate that future research will be able to empirically confirm whether these factors are critical for sustained implementation of the advanced tiers and identify other novel district and school features that contribute to sustainability.

Limitations and Future Research

There were limitations across the two current studies that future research may address. The first limitation is the generalizability of the samples of participants in Study 1. We assessed the content representativeness of the ALTITUDE measure using an expert panel of researchers and practitioners with a high investment in implementing Tier 2 and 3 behavior support systems within a SWPBIS framework. Although the number of participants in the expert panel exceeded recommended sample sizes (Gable & Wolf, 1993; Rubio et al., 2003), it is possible that their experiences and perceptions may have overemphasized or underemphasized certain features

related to Tier 2 and 3 systems, especially to multi-tiered systems of support other than SWPBIS. Additionally, the majority of expert panel members were primarily researchers, not practitioners, and therefore these members may have placed a higher importance on certain features of Tier 2 and 3 systems, which would then be reflected in the ALTITUDE items. Because these participants helped to develop the items, these participants may have rated certain items on the measure more favorably than other researchers and practitioners implementing Tier 2 and 3 behavior support systems. In addition, the sample in Study 2 included individuals from schools that were actively implementing SWPBIS, generally with support from strong state SWPBIS networks. Similar to the expert panel, their ratings may have been biased by the investment their schools had made with implementing SWPBIS. Potential recommendations for future research to address these sample limitations would be to replicate the findings with another group of participants at different stages of SWPBIS implementation and examine the measurement invariance of the ALTITUDE. For example, research could examine whether there are significant differences in ALTITUDE scores for schools at different stages of Tier 2 and 3 implementation, such as early adopters versus sustainers (Forman et al., 2013).

Another limitation was the availability of fidelity data, given that fewer than half of the participants that completed the ALTITUDE measure also completed a TFI at Tier 3. For this reason, we used listwise deletion based on availability of TFI data by tier when examining relations between ALTITUDE factors and TFI subscale factors. Although using listwise deletion to handle missing TFI data likely introduced some bias, we believe this was the least limited option for missing data due to the large percentages of missing data (e.g., 55% of cases for Tier 3) and the unclear specific missing data mechanism. Specifically, there was some evidence that data were MAR given that the probability of missing data on the TFI was predictable by

ALTITUDE scores (e.g., schools with Tier 3 TFI scores had higher general Tier 2 and 3, Tier 2 specific, and Tier 3 specific ALTITUDE scores, all $p < .001$), but it is also possible that TFI data were MNAR if schools did not complete the TFI because they were not implementing Tier 2 and 3 behavior systems.

Relatedly, the sample size in Study 2 was minimally adequate considering the complexity of the models fit to the data (i.e., multiple factors with ordinal items). To address these limitations, future research could replicate and extend the current work by including a more varied sample of schools assessing implementation fidelity at Tiers 2 and 3. Even if schools were not fully implementing Tiers 2 and 3, it would be worthwhile to examine their Tier 2 and 3 fidelity scores and their ALTITUDE scores to identify factors related to the implementation and sustainability of Tier 2 and 3 behavior support systems for schools in the early stages of implementation. It should also be noted that the responses of any single participant may not be representative of a school team or others with knowledge of the schools' systems. Although most participants in Study 2 were in leadership or teaming positions at their schools, it is possible that their perceptions of the features assessed by the ALTITUDE varied considerably from other team members in their schools. Future research could help to examine these potential differences by conducting another large-scale study in which multiple school personnel within each building completed the ALTITUDE and interrater reliability of the ALTITUDE be evaluated. The two current studies discussed were cross-sectional examinations of Tier 2 and 3 behavior support systems. Thus, the availability of the ALTITUDE enables longitudinal examinations of factors predicting sustained implementation of Tier 2 and 3 practices. This new opportunity is important, given that the existing studies on this topic have been descriptive or retrospective (Bambara et al., 2012; Woodbridge et al., 2014). Additionally, future research may point to the need for a tier-

general sustainability factor, based on the relation of general Tier 2 and 3 ALTITUDE factors with Tier 1 TFI fidelity.

Implications for Practice

Although we believe the ALTITUDE will be most useful as a research tool, compared to an action planning tool, future research using the ALTITUDE could have significant implications for district and school SWPBIS implementation teams. As implementing SWPBIS within district and schools is resource-intensive, particularly EBPs and behavior systems within the advanced tiers (Bambara et al., 2012; Bambara et al., 2009; Loman et al., 2010), implementation teams can benefit from learning which key features of the advanced behavior systems are predictive of sustained implementation for both tiers vs. each specific tier. This information could be particularly helpful to SWPBIS teams when structuring their professional development activities and implementation timelines for the advanced tiers. For example, once Tier 1 systems are established and implemented with adequate fidelity, leadership teams may decide to invest first in implementing features that are predictive of sustainability for both Tier 2 and 3 behavior systems (e.g., building strong partnerships with families to support students needing Tier 2 and 3 supports) prior to investing in specific and highly expensive resources that may only affect the sustainability of one advanced tier (e.g., training school personnel to collect Tier 3 implementation fidelity and student progress data efficiently).

References

Algozzine, R. F., Barrett, S., Eber, L., George, H., Horner, R. H., Lewis, T. J., Putnam, R. F., Swain-Bradway, J., McIntosh, K., & Sugai, G. (2014). *SWPBIS Tiered Fidelity Inventory*. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. Available at http://www.pbis.org.

Bambara, L., Goh, A., Kern, L., & Caskie, G. (2012). Perceived barriers and enablers to implementing individualized positive behavioral interventions and supports in school settings. *Journal of Positive Behavior Interventions, 14*(4), 228-240. https://doi.org/10.1177/1098300712437219

Bambara, L., Nonnemacher, S., & Kern, L. (2009). Sustaining school-based individualized Positive Behavior Support: Perceived barriers and enablers. *Journal of Positive Behavior Interventions, 11*(3), 161-178. https://doi.org/10.1177/1098300708330878

Burke, J. D., & Johnston, O. G. (2020). The Bifactor S-1 model: A psychometrically sounder alternative to test the structure of ADHD and ODD? *Journal of Abnormal Child Psychology, 48*(7), 911-915. https://doi.org/10.1007/s10802-020-00645-4

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020). Application of the Bifactor S – 1 model to multisource ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. *Journal of Abnormal Child Psychology, 48*, 881–894. https://doi.org/10.1007/s10802-019-00608-4

Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5*(4), 194-197. https://doi.org/10.1016/s0897-1897(05)80008-4

Debnam, K. J., Pas, E. T., & Bradshaw, C. P. (2012). Secondary and tertiary support systems in schools implementing school-wide positive behavioral interventions and supports: A

preliminary descriptive analysis. *Journal of Positive Behavior Interventions, 14*(3), 142-152. https://doi.org/10.1177/1098300712436844

Eid, M. (2020). Multi-faceted constructs in abnormal psychology: Implications of the Bifactor S - 1 model for individual clinical assessment. *Journal of Abnormal Child Psychology, 48*, 895–900. https://doi.org/10.1007/s10802-020-00624-9

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., Sanetti, L. M. H., & Stoiber, K. C. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28*(2), 77-100. https://doi.org/10.1037/spq0000019

Gable, R. K., & Wolf, J. W. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings.* Kluwer Academic.

Hagermoser Sanetti, L. M., Collier-Meek, M. A., Long, A. C. J., Byron, J., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of School Psychology, 53*(3), 209-229. https://doi.org/10.1016/j.jsp.2015.03.002

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008-1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/bf02289447

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior Analysis in Practice, 8*(1), 80-85. https://doi.org/10.1007/s40617-015-0045-4

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. https://doi.org/https://doi.org/10.1080/10705519909540118

Hume, A. E., & McIntosh, K. (2013). Construct validation of a measure to assess sustainability of school-wide behavior interventions. *Psychology in the Schools, 50*(10), 1003-1014. https://doi.org/10.1002/pits.21722

Kittelman, A., Eliason, B., Dickey, C. R., & McIntosh, K. (2018). *How are Schools Using the SWPBIS Tiered Fidelity Inventory (TFI)?* OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org.

Kittelman, A., Strickland-Cohen, M. K., Pinkelman, S. E., & McIntosh, K. (2020). Variables contributing to abandonment and readoption of SWPBIS. *Journal of Positive Behavior Interventions, 22*(2), 67-77. https://doi.org/10.1177/1098300719888748

Loman, S. L., Rodriguez, B. J., & Horner, R. H. (2010). Sustainability of a targeted intervention package: First Step to Success in Oregon. *Journal of Emotional and Behavioral Disorders, 18*(3), 178-191. https://doi.org/10.1177/1063426610362899

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88–91. https://doi.org/10.3758/BF03192753

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-385.

Lyon, A. R., Cook, C. R., Locke, J., Davis, C., Powell, B. J., & Waltz, T. J. (2019). Importance and feasibility of an adapted set of implementation strategies in schools. *Journal of School Psychology, 76*, 66-77. https://doi.org/10.1016/j.jsp.2019.07.014

Massar, M., McIntosh, K., Mercer, S. H., & Hoselton, R. (2017). Factor analysis of a fidelity of implementation measure for social behavior systems. *Remedial and Special Education, 40*(1), 16-24. https://doi.org/10.1177/0741932517736515

McDonald, R. P. (1999). *Test theory: A unified treatment.* Erlbaum.

McIntosh, K., & Goodman, S. (2016). *Integrated multi-tiered systems of support: Blending RTI and PBIS*. Guilford Press.

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*(4), 208-218. https://doi.org/10.1177/1098300710385348

McIntosh, K., Massar, M. M., Algozzine, R. F., George, H. P., Horner, R. H., J., T. L., & Swain-Bradway, J. (2017). Technical adequacy of the SWPBIS tiered fidelity inventory. *Journal of Positive Behavior Interventions, 19*(1), 3-13. https://doi.org/10.1177/1098300716637193

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*(3), 293-311.

McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., Kittelman, A., Hoselton, R., & Horner, R. H. (2018). Factors predicting sustained implementation of a universal

behavior support framework. *Educational Researcher, 47*(5), 307-316. https://doi.org/10.3102/0013189X18776975

Mercer, S. H., McIntosh, K., Strickland-Cohen, M. K., & Horner, R. H. (2014). Measurement invariance of an instrument assessing sustainability of school-based universal behavior practices. *School Psychology Quarterly, 29*(2), 125-137. https://doi.org/https://doi.org/10.1037/spq0000054

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nadeem, E., & Ringle, V. A. (2016). De-adoption of an evidence-based trauma intervention in schools: A retrospective report from an urban school district. *School Mental Health, 8*, 132-143. https://doi.org/10.1007/s12310-016-9179-y

Reddy, L. A., Shernoff, E., Lekwa, A., Matthews, C., Davis, W., & Dudek, C. M. (2019). Coaching to improve teacher instruction and behavior management in a high poverty school: A case study. *School Psychology, 34*(1), 14-21. https://doi.org/10.1037/spq0000302

Rubio, D. M., Berg-Wagner, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94-104. https://doi.org/10.1093/swr/27.2.94

Shi, D., Lee, T., Fairchild, A. J., & Maydeu-Olivares, A. (2020). Fitting ordinal factor analysis models with missing data: A comparison between pairwise deletion and multiple

imputation. *Educational and Psychological Measurement, 80*(1), 41–66. https://doi.org/10.1177/0013164419845039

Sugai, G., & Horner, R. H. (2019). Sustaining and scaling positive behavioral interventions and supports: Implementation drivers, outcomes, and considerations. *Exceptional Children, 86*(2), 120-136. https://doi.org/10.1177/0014402919855331

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209–220. https://doi.org/10.1037/a0023353

Turri, M. G., Mercer, S. H., McIntosh, K., Nese, R. N. T., Strickland-Cohen, M. K., & Hoselton, R. (2016). Examining barriers to sustained implementation of school-wide prevention practices. *Assessment for Effective Interventions, 42*(1), 6-17. https://doi.org/10.1177/1534508416634624

Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H., & Feil, E. G. (1998). First Step to Success: An early intervention approach for preventing school antisocial behavior *Journal of Emotional and Behavioral Disorders, 6*(2), 66-80. https://doi.org/10.1177/106342669800600201

Woodbridge, M. W., Sumi, W. C., Yu, J., Rouspil, K., Javitz, H. S., Seeley, J. R., & Walker, H. M. (2014). Implementation and sustainability of an evidence-based program: Lessons learned from the PRISM applied to First Step to Success. *Journal of Emotional and Behavioral Disorders, 22*(2), 95-106.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1). https://doi.org/10.22237/jmasm/1177992180

Table 1

School Characteristics for the National Sample of Schools in Study 2

Characteristics	Sample
Number of schools	646
Number of districts/regions/counties	254
Number of U.S. states	23
Student enrollment, *M* (*SD*)	*608* (*375*)
% Non-white students, *M* (*SD*)	*52 (29)*
% Students receiving FARMs, *M* (*SD*)	*60* (*26*)
School type	
% Elementary schools	60
% Middle schools	14
% High schools	10
% Other	16
School locale	
% Schools in cities	35
% Schools in suburbs	35
% Schools in towns	12
% Schools in rural areas	15

Note. FARMs = free and reduced price meals. Missing data across demographic variables was between 4% to 5%.

Table 2

ALTITUDE Items

Tier 2 and 3 related
1. There is adequate communication across all teams providing Tier 1, 2, and 3 behavior support.*
2. The team(s) responsible for Tier 2 and 3 behavior systems have procedures in place to select and/or train team members on supporting these behavior systems (e.g., monitoring fidelity and student performance).*
3. The team(s) and school personnel responsible for implementing Tier 2 and 3 behavior interventions receive acknowledgement for implementation efforts and accomplishments.*
4. School personnel understand the importance of monitoring intervention fidelity for Tier 2 and 3 behavior interventions.*
5. The school principal consistently expresses the importance of implementing Tier 2 and 3 behavior interventions.*
6. District administrators express a commitment that students with intensive needs should be supported within schools instead of being removed to more restrictive settings.*
7. The district uses hiring procedures (e.g., interview questions, hiring criteria) to select teachers who have skills in supporting students through Tier 2 and 3 behavior interventions.
8. The team(s) responsible for implementing Tier 2 and 3 behavior systems are connected to a "community of practice" (i.e., a network of other schools implementing Tier 2 and 3 behavior systems in the district, state, or region).*
9. Materials (e.g., documents, curricula, tools) for implementing Tier 2 and 3 behavior interventions can be used or adapted with ease over time.*
10. School personnel are committed to supporting students in the classroom instead of excluding them from instruction.*
11. School personnel are committed to implementing Tier 2 and 3 behavior interventions.*
12. The team(s) responsible for implementing Tier 2 and 3 behavior systems effectively problem-solve barriers to implementation.*
13. School personnel build strong partnerships with families to support students with Tier 2 and 3 behavior needs.*

14. Classroom teachers regularly receive fidelity and student progress data about their assigned students receiving Tier 2 and 3 behavior interventions.*

15. Parents/caregivers regularly receive progress data about their children and youth participating in Tier 2 and 3 behavior interventions.*

Tier 2 specific

16. All Tier 2 behavior interventions (e.g., Check-In Check-Out, social/emotional skills small groups) are coordinated by one team.*

17. The team responsible for Tier 2 behavior systems has adequate resources (e.g., time, personnel, materials) to implement Tier 2 behavior interventions with fidelity.*

18. School personnel have adequate resources (e.g., time, materials) to implement Tier 2 behavior interventions with fidelity.

19. Students receiving Tier 2 behavior support also receive academic support when needed.

20. The team responsible for Tier 2 behavior systems has adequate access to training in Tier 2 behavior systems and interventions.*

21. The team responsible for Tier 2 behavior systems has adequate access to coaching for Tier 2 behavior systems and interventions.*

22. School personnel implementing Tier 2 behavior interventions receive adequate coaching on fidelity of data collection and plan implementation that includes performance feedback.

23. School personnel are knowledgeable in the logic and practices of Tier 2 behavior interventions (e.g., why the interventions should work, for whom specific interventions are most likely to be effective, specific skills needed to implement the interventions).*

24. School personnel implementing Tier 2 behavior interventions have the necessary skills for collecting fidelity and student progress data accurately.*

25. The team responsible for Tier 2 behavior systems uses fidelity and student progress data to improve behavior systems and outcomes.*

26. Individuals implementing Tier 2 behavior interventions can collect fidelity and student progress data efficiently.*

27. The school's Tier 2 behavior data systems are easy for the team to use for decision making.*

28. District administrators promote a standard set of evidence-based Tier 2 behavior systems, data, and practices (e.g., screening tools, interventions, decision rules) to be used district-wide.

Tier 3 specific
29. The team responsible for Tier 3 behavior systems has adequate resources (e.g., time, personnel, materials) to implement Tier 3 behavior interventions with fidelity.*
30. School personnel have adequate resources (e.g., time, materials) to implement Tier 3 behavior interventions with fidelity.
31. The team responsible for Tier 3 behavior systems has adequate access to training in Tier 3 behavior systems and interventions.*
32. The team responsible for Tier 3 behavior systems has adequate access to coaching for Tier 3 behavior systems and interventions.*
33. School personnel implementing Tier 3 behavior interventions receive adequate coaching on fidelity of data collection and plan implementation that includes performance feedback.
34. School personnel are knowledgeable in the logic and practices of Tier 3 behavior interventions (e.g., why the interventions should work, for whom specific interventions are most likely to be effective, specific skills needed to implement the interventions).*
35. School personnel implementing Tier 3 behavior interventions have the necessary skills for collecting fidelity and student progress data accurately.*
36. The team responsible for Tier 3 behavior systems uses fidelity and student progress data to improve behavior systems and outcomes.*
37. Individuals implementing Tier 3 behavior interventions can collect fidelity and student progress data without too much effort.*
38. The school's Tier 3 behavior data systems are easy for the team to use for decision making.*
39. District administrators promote a standard set of evidence-based Tier 3 behavior systems, data, and practices (e.g., screening tools, interventions, decision rules) to be used district-wide.*

Note. * indicates that the item was retained in the final 32-item ALTITUDE.

Table 3

ALTITUDE Factor Loadings from the EFA and CFA in Study 2

Tiers assessed	Item	% Missing	EFA			CFA Loading	Bifactor S-1 CFA	
			Factor 1	Factor 2	Factor 3		General	Specific
Tier 2 and 3 related	1	2.5	.49	.35	.02	.84	.84	—[d]
	2	3.1	.55	.30	.04	.86	.86	—[d]
	3	5.0	.62	.19	-.06	.72	.72	—[d]
	4	2.0	.69	.13	-.03	.80	.80	—[d]
	5	2.6	.69	.19	-.03	.81	.81	—[d]
	6	2.3	.49	.13	.05	.51	.51	—[d]
	7	30.3	—[a]	—[a]	—[a]	—[a]	—[a]	—[a]
	8	7.0	.39	.31	.02	.68	.68	—[d]
	9	4.5	.49	.29	.06	.82	.82	—[d]
	10	0.6	.97	-.18	-.07	.66	.66	—[d]
	11	2.5	.82	.06	-.06	.85	.85	—[d]
	12	3.3	.71	.10	.17	.87	.87	—[d]
	13	2.9	.81	-.04	.04	.72	.72	—[d]
	14	4.6	.54	.17	.19	.84	.84	—[d]
	15	7.6	.55	.01	.18	.76	.76	—[d]
Tier 2 specific	16	5.0	.18	.65	-.15	.71	.66	.22
	17	5.3	-.01	.84	.10	.85	.78	.29
	18	5.0	-.03	.86	.06	—[b]	—[b]	—[b]
	19	6.2	.38	.35	.01	—[c]	—[c]	—[c]
	20	6.0	.12	.80	.00	.86	.74	.45
	21	6.5	-.03	.83	.08	.82	.68	.50
	22	7.3	.19	.62	.16	—[b]	—[b]	—[b]
	23	5.9	.32	.52	.11	.88	.77	.42
	24	5.7	.24	.54	.14	.88	.72	.55
	25	6.3	.33	.59	.04	.91	.81	.40
	26	6.5	.34	.54	.06	.92	.81	.42
	27	7.0	.38	.45	.14	.85	.75	.39
	28	10.8	.37	.16	.28	—[c]	—[c]	—[c]
Tier 3 specific	29	19.5	-.26	.43	.82	.87	.71	.48
	30	18.4	-.24	.41	.83	—[b]	—[b]	—[b]
	31	19.5	-.01	.24	.77	.90	.66	.63
	32	19.8	-.06	.27	.77	.92	.70	.59
	33	19.8	.10	.09	.78	—[b]	—[b]	—[b]
	34	18.6	.23	-.01	.78	.90	.66	.63
	35	19.3	.17	.04	.77	.92	.67	.64
	36	20.4	.29	.05	.69	.93	.73	.56
	37	20.7	.41	-.20	.75	.91	.70	.58
	38	21.1	.44	-.20	.76	.92	.72	.57
	39	22.3	.29	-.09	.65	.76	.60	.44

Note. Two independent subsamples of n = 323 schools (n = 646) were used for the EFA and CFA. — indicates that the item was not included in the analysis.

[a] indicates that the item was excluded due to unusually high percentage of missing responses (30%).

[b] the item was excluded due to content redundancy with other items retained on the ALTITUDE.

[c] the item was excluded because the loading on the target factor was lower than for another factor in the EFA.

[d] the ALTITUDE Tier 2 and 3 related factor served as the general reference factor in the Bifactor S-1 analyses, thus, there are no specific factor loadings presented.

Table 4

Model Fit for CFA Models in Study 2

Model	χ^2			CFI	RMSEA	SRMR
	Value	*df*	*p*			
1	1858.16	464	<.001	.94	.10	.08
2	1209.91	463	<.001	.97	.07	.06
3	1007.26	461	<.001	.98	.06	.05
4	1220.53	455	<.001	.97	.07	.06
5	1002.90	445	<.001	.98	.06	.05

Note. Models were fit with the mean and variance adjusted weighted least squares estimator to obtain χ^2, CFI, RMSEA, and SRMR values, and with robust maximum likelihood to obtain AIC and BIC values. Model 1 = 1 factor, Model 2 = 2 correlated factors, Model 3 = 3 correlated factors, Model 4 = Bifactor S-1 model with 1 specific factor, Model 5 = Bifactor S-1 model with 2 specific factors. CFI = comparative fit index, RMSEA = root-mean-square error of approximation, SRMR = standardized root mean squared residual.

Table 5

Correlations among Sustainability (ALTITUTDE) and Fidelity (TFI) Factors

Factor	1	2	3	4	5	6
1. ALT T23	—	0[a]	0[a]	.45***	.44***	.35***
2. ALT T2	.85***	—	.38***	.15*	.20**	-.11
3. ALT T3	.74***	.76***	—	-.10	.09	.23**
4. TFI T1	.45***	.46***	.27***	—	[b]	[b]
5. TFI T2	.43***	.48***	.38***	[b]	—	[b]
6. TFI T3	.34***	.24***	.41***	[b]	[b]	—

Note. Correlations above the diagonal are from Bifactor S-1 models with all ALTITUDE items loading on a general Tier 2 and 3 sustainability factor, Tier 2 items also loading on a specific Tier 2 sustainability factor, and Tier 3 items also loading on a specific Tier 3 sustainability factor. Correlations below the diagonal are from models with three correlated ALTITUDE factors. Correlations among ALTITUDE factors and with the Tier 1 TFI are based on models with $n = 428$, correlations with the Tier 2 TFI are based on models with $n = 443$, and correlations with the Tier 3 TFI are based on models with $n = 289$.

[a] Correlations between the general and specific factors are constrained to zero in Bifactor S-1 models.

[b] Correlations among TFI factors were not estimated because separate models were fit by TFI tier. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

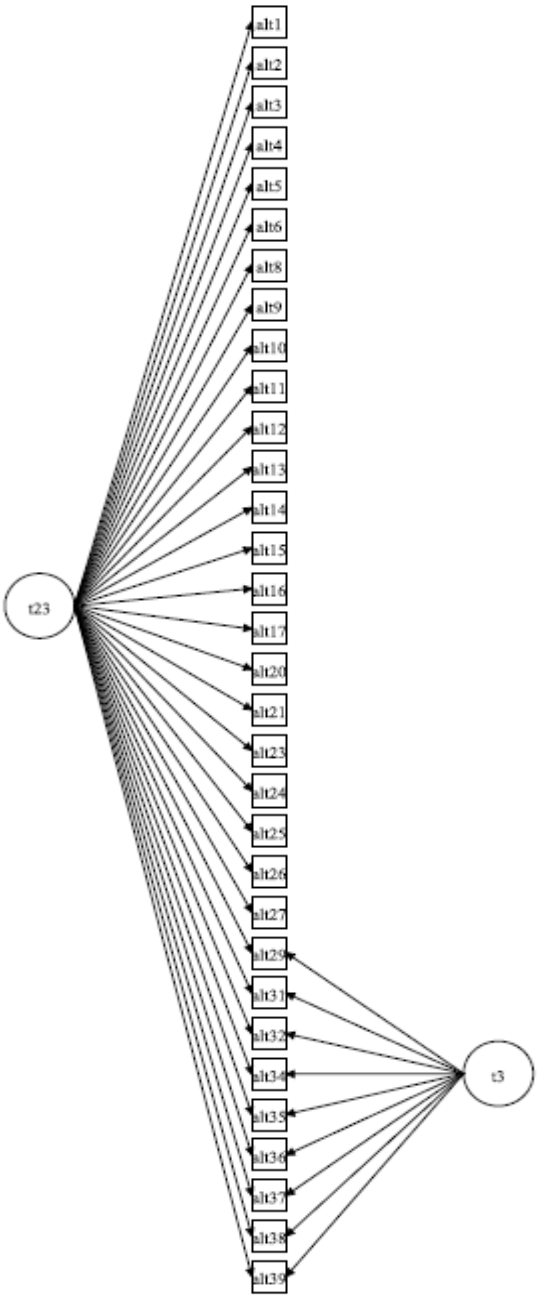

Figure 1. Bifactor S-1 Model for the ALTITUDE with One Specific Residual Factor. T23 = general Tier 2 and 3 sustainability factor, T3 = Tier 3 specific residual sustainability factor. alt[] = ALTITUDE item number.

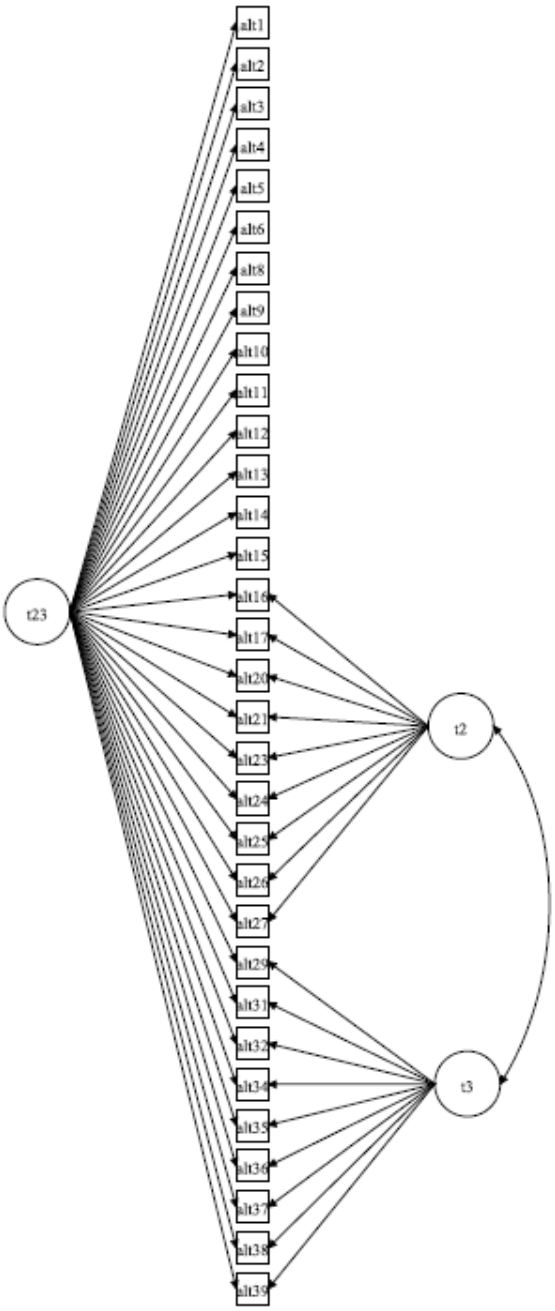

Figure 2. Bifactor S-1 Model for the ALTITUDE with Two Specific Residual Factors. T23 = general Tier 2 and 3 sustainability factor, T2 = Tier 2 specific residual sustainability factor, T3 = Tier 3 specific residual sustainability factor. alt[] = ALTITUDE item number.