Article

Journal of Psychoeducational Assessment
2022, Vol. 40(8) 936–949

DOI: 10.1177/07342829221116807
journals.sagepub.com/home/jpa
SAGE

Evidence of the Internal Structure and Measurement Invariance of the BASC-3 Behavioral and Emotional Screening System Teacher Form

Katie Eklund[1], Stephen P. Kilgus[1], Jessica B. Willenbrink[1], Bri'Anna Collins[1], Noah Gill[1], Madison M. Weist[1], Julia Porter[1], Timothy J. Lewis[2], Barbara Mitchell[3], and Howard Wills[4]

Abstract

The Behavior Assessment System for Children, Third Edition (BASC-3) Behavioral and Emotional Screening System Teacher Form (BESS) is a universal screening measure designed to identify social, emotional, and behavioral risks in students. A series of confirmatory factor analyses (CFAs) were conducted to examine the extent to which the developer-proposed factor model fits the BASC-3 BESS data within an elementary K-5 sample of students (N = 1472). Results suggested a higher-order factor structure, which was specified in accordance with developer recommendations, provided acceptable fit to the BASC-3 BESS data. Measurement invariance of the factor structure across student age levels was then examined via a series of multi-group CFA models. Introduction Multi-group CFA findings supported the configural, metric, and scalar invariance of the factor structure across age levels. Practical implications for educators and school psychologists interested in universal screening and directions for future research are discussed.

Keywords

universal screening, behavior assessment, confirmatory factor analysis

[1]University of Wisconsin-Madison, Madison, WI, USA
[2]University of Missouri, Columbia, MO, USA
[3]Kansas State University in Manhattan, Ks, USA
[4]University of Kansas, Lawrence, Ks, USA

Corresponding Author:
Katie Eklund, Department of Educational Psychology, University of Wisconsin-Madison, 1025 W Johnson St, Madison, WI 53706, USA.
Email: katie.eklund@wisc.edu

■■■

A significant number of children and adolescents experience social, emotional, and behavioral (SEB) challenges that can negatively impact their behavioral and academic functioning across a wide range of areas (Darney et al., 2013; King et al., 2019). Universal screening has been identified as one strategy to identify children who may be experiencing academic and/or behavioral concerns to provide early intervention. However, behavior screening instruments have notably less evidence regarding their psychometric properties and currently do not experience widespread adoption as compared to academic screening measures (Bruhn et al., 2014). As a result, the primary aim of this study was to further investigate the proposed factor structure and measurement invariance of the Behavior Assessment System for Children, Third Edition (BASC-3) Behavioral and Emotional Screening System Teacher Form (BESS; Kamphaus & Reynolds, 2015), a commonly used tool within schools that conduct universal screening for SEB concerns.

Rates of Mental Health Concerns

Research suggests at least 20% of children experience some type of mental health concern throughout their school trajectory; however, less than 50% of youth in need of mental health intervention actually receive treatment (Merikangas et al., 2010). Children with SEB challenges face a myriad of concerns. For example, studies evaluating the outcomes of children with behavioral problems found that they are more likely to experience academic failure, peer rejection, and delinquency than peers that lack behavioral difficulties (Bradley et al., 2004; Reinke et al., 2008; Wagner et al., 2005). In a sample of first grade students who screened positive for behavior concerns, 11 years later these students experienced higher rates of conduct disorder, being placed into special education, and being arrested by 12th grade (Darney et al., 2013). Given the plethora of negative outcomes experienced by young children with SEB concerns, there is a critical need for accurate identification and treatment of these problems early in life. The current study focuses on an elementary sample to further examine universal screening measures that can accurately identify children in order to provide necessary services.

Traditionally, schools have utilized a "wait-to-fail" approach to supporting students, where students are identified as having SEB difficulties only once the symptoms become severe enough to be noticed by school professionals. Two common methods within this approach include (1) reliance on teacher identification and referral of student SEB concerns and (2) monitoring of office discipline referral (ODR) data to support student risk identification. Unfortunately, these approaches have also been found to be most vulnerable to bias, due to differences in teacher expectations of student behavior functioning and variations in understanding how and when to refer students for additional services (Severson et al., 2007). In contrast to this approach, universal approaches where all children are screened in a particular classroom or school provide the opportunity to identify children with SEB needs across a continuum of severity (Eklund & Dowdy, 2014). Thus, universal screening has the additional benefit of identifying children with behavioral and emotional needs that, while not clinically significant, put them at an increased risk of developing more intense problems later in life (Glover & Albers, 2007). Previous research has found that more than half of the children that were at-risk for behavioral and emotional problems were not identified through traditional referral pathways (e.g., teacher referral and school-based problem solving teams), suggesting that screening may cast a wider net and bring students to the attention of educators who might not be otherwise referred (Eklund et al., 2009; Eklund & Dowdy, 2014). In addition to the improved ability to detect subclinical problems early on, evidence also suggests that universal screening provides opportunities to initiate early intervention,

ultimately preventing the worsening of symptoms over time and reducing the time between intervention and the onset of symptoms (Chin et al., 2013).

Universal Screening

Universal screening is the process of assessing many students in a systematic way, allowing school professionals to quickly and accurately identify students at risk for or currently experiencing SEB difficulties (Ikeda et al., 2008). Universal screening measures for SEB concerns are brief, cost-effective rating scales that are designed to be administered to large groups of students across an entire classroom, grade-level, or school (Severson et al., 2007). Measures can be completed by teachers, parents, and/or students. Scores are then used to determine which students are demonstrating elevated concerns for SEB risk. Within Multi-Tiered Systems of Support (MTSS), universal screening is seen as a universal, Tier 1 assessment approach to identifying students who may be in need of additional targeted interventions and supports. These results can then be used to consider universal approaches to addressing student needs that could happen across an entire classroom (e.g., implementation of a social emotional learning curriculum) or may result in more targeted services that could be delivered to a smaller group of students (e.g., social skills instruction). Ultimately, implementing universal screening within an MTSS framework has shown to connect more at-risk students with services and supports (Eklund & Dowdy, 2014).

Screening measures should provide reliable and valid information regarding core constructs that are associated with mental health concerns among school-aged youth (Frick et al., 2009). Edelbrock and Achenbach (1980) identified several key constructs related to student SEB functioning, including externalizing problems, internalizing problems, adaptive behaviors, and attention/learning problems. In order to accurately detect children demonstrating a wide range of concerns, it is important that behavior screening measures accurately capture many of these important constructs.

Several screening measures have been developed to assess SEB risk. One example of a commonly used SEB screener is the BASC-3 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2015). The BASC-3 BESS is a standardized tool designed to assess students' externalizing and internalizing behavior problems, as well as adaptive skills (Reynolds & Kaphaus, 2015), using items drawn from the Behavioral Assessment System for Children, Third Edition (BASC-3; Reynolds & Kaphaus, 2015). The BASC-3 BESS consists of teacher, parent, and student rating forms, which contain items that ask the participant to respond using a four-point Likert-type scale ranging from 0 *(Never)* to 3 *(Almost Always),* with each form taking an average of 5 minutes to complete (Jenkins et al., 2014). The BASC-3 BESS is an efficient measure that has been shown to be a valid predictor of important school-based outcomes such as academic achievement, behavioral engagement, and school climate (Kamphaus & Reynolds, 2007; Kamphaus et al., 2007; Naser & Dever, 2020). The items selected for use with the BESS Teacher form were selected from the standardization items of the Behavior Assessment Scale for Children, Third Edition (BASC-3) Teacher Rating Scale using a test development model that emphasized theory and prior factor analytic work on the full BASC-3 forms. Based on this work, the BASC-3 BESS provides four resulting scores, including an overall Behavioral and Emotional Risk total score, as well as three subindex raw scores: externalizing risk, internalizing risk, and adaptive risk (Kamphaus & Reynolds, 2015). The total score is expressed in terms of a T score (M = 50, SD = 10), while the remaining three scores are raw summed scores. All four continuous scores can be categorized in terms of risk levels based upon distance from the normative mean (i.e., Normal, Elevated, and Extremely Elevated).

Previous research has primarily investigated the factor structure of the Teacher BASC-2 BESS. For example, Wiesner and Schanding (2013) used a structural equation modeling approach to

examine the factor structure of the Teacher Form from a sample of 1885 elementary-aged students. Findings revealed a multi-dimensional factor structure and a preference for a bifactor model that considered a higher order factor measuring overall risk as well as subscales of risk. A second study by Dever et al., (2012) used exploratory factor analyses and confirmatory factor analyses to examine the factor structure of the Teacher BASC-2 BESS among a sample of 2582 children aged 6–12 years. Results suggested a four-factor latent structure consisting of externalizing problems, internalizing problems, school problems, and adaptive skills.

Although previous research examines the measurement invariance of the parent and student self-report forms of the BASC-3 BESS for gender and race (e.g., Kim & Kamphaus, 2018; Edyburn et al., 2020), no previous studies to date were identified exploring measurement invariance for the Teacher form. Moving forward, BESS research should evaluate previously unexamined psychometric characteristics (e.g., measurement invariance) for the various informant reports, while also evaluating previously examined properties (e.g., validity evidence of the internal structure; Kamphaus & Reynolds, 2015). Taken together, such evidence would speak to the broader construct validity of BASC-3 BESS scores (Messick, 1995), and thus their defensibility for use within schools. There is also a need for replication within screening research as a review of extant measures suggests that many universal screening tools may be only supported by a technical manual or a single peer-reviewed paper. Although the well-documented replication crisis within psychology has several far-reaching implications, it is particularly problematic for applied disciplines (e.g., school psychology) that rely on replication research to identify tools and practices that schools can confidently use to best serve children (Makel & Plucker, 2014).

Current Study

The current study is designed to replicate previous research examining the factor structure of the BASC-2 BESS (Dever et al., 2012; Wiesner & Schanding, 2013) using a large sample of elementary-aged students. More specifically, the first goal is to determine whether prior results would replicate within this investigation, supporting the developer-proposed model described in the BASC-3 BESS manual (Kamphaus & Reynolds, 2015). The second goal is to expand upon the prior investigation by examining the extent to which the BASC-3 BESS is invariant depending on the age of the child. The BASC-3 BESS is used across K-12 settings. This presumes that the proposed factor structure is appropriate to children across varying development levels. If such widespread use is to be appropriate, research should indicate BASC-3 BESS scores can be interpreted in a similar manner across various grade levels. Evidence of measurement invariance would indicate to what extent a particular factor model is structurally valid relative to students across multiple age levels (Pendergast et al., 2017). The current study examined two research questions. First, to what extent does the developer-proposed factor structure fit observed BESS data? Second, to what extent is the factor structure invariant across ages at the elementary level?

Method

Participants

Four K-5 schools across two suburban public school districts in the Midwest participated in this study. A total of 65 teachers completed the BESS for each student in their classroom (total $n = 1472$). Student ages ranged from 5 to 11 with a mean age of 8.14. Age distributions were as follows: 5 years (4.1%, $n = 61$), 6 years (14.7%, $n = 216$), 7 years (18.3%, $n = 270$), 8 years (18.2%, $n = 268$), 9 years (20.3%, $n = 299$), 10 years (18.9%, $n = 278$), and 11 years (5.4%, $n = 80$). The BASC-3 BESS was used as part of a de-identified screening process in evaluating student

eligibility for a randomized controlled trial. Accordingly, individual student demographic data, with the exception of student age, which was collected for BESS scoring purposes, were not collected for the purposes of this investigation. However, school-level demographics were extracted from the National Center for Education Statistics (NCES, 2020) for each participating school. Student demographic data by school for the 2018–2019 school year are provided in Table 1. Screening data were collected the same school year in which participant data are recorded.

Measures

BESS. The BASC-3 BESS Teacher Child/Adolescent Form is a 20-item norm-referenced and commercially available brief behavior rating scale that takes 5 minutes or less to complete per student, allowing for efficient screening of behavioral and emotional concerns for youth in grades K-12 (Kamphaus & Reynolds, 2015). Teachers rate each item using a 4-point Likert scale indicating if a behavior occurs *Never*, *Sometimes*, *Often*, or *Almost Always*. Four scores are then generated, including a total Behavioral and Emotional Risk Index (BERI), as well as three subscale scores: Externalizing Risk Index (ERI), Internalizing Risk Index (IRI), and the Adaptive Skills Risk Index (ARI). To note, 17 of the 20 items are used to compute both the BERI and one of the subscale scores; the remaining three items are only used to compute the BERI score and do not correspond to any subscale.

BERI scores correspond to *T*-scores ($M = 50$, $SD = 10$), which are derived via the use of age-specific norms. Subscale scores correspond to raw scores representative of the sum of item ratings within each scale. Once computed, all scores are compared to age-specific categories to determine

Table 1. Demographic Information.

	School 1		School 2		School 3		School 4		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Enrollment (K-5)	605		1135		653		338		2731	
Teacher participants	11		17		23		14		65	
Student participants	232		373		564		299		1472	
Student age										
5 years	21		14		0		24		61	4.1
6 years	34		66		67		48		216	14.7
7 years	43		54		114		58		270	18.3
8 years	36		55		118		59		268	18.2
9 years	14		120		118		46		299	20.3
10 years	60		55		119		45		278	18.9
11 years	24		9		28		19		80	5.4
Student race/ethnicity										
American Indian/Alaska Native	3	<1	0	0	0	0	0	0	3	<1
Asian	81	13	20	2	6	<1	6	2	113	4
Black	74	12	76	7	75	12	56	17	281	10
Hispanic	44	7	38	3	67	10	59	18	208	8
Native Hawaiian/Pacific Islander	2	<1	0	0	2	<1	4	1	8	<1
White	359	59	931	82	414	63	190	56	1894	69
Two or more races	42	7	69	6	87	13	23	7	221	8
% Female	298	49	527	46	316	48	166	49	1307	48
Free or reduced lunch	204	22	373	22	313	34	219	45	1109	41

each student's behavioral level within each scale. For the BERI, ERI, and IRI scales, scores that are <1 SD above the mean are considered Normal, >1 SD above the mean represent Elevated Risk, and >2 SD above the mean represent Extremely Elevated Risk. For the ARI scale, the scores are reversed so that scores <1 SD below the mean represent Normal, >1 SD below the mean represent Elevated Risk, and >2 SD below the mean represent Extremely Elevated Risk.

Results from the developers' norming and validation process (Kamphaus & Reynolds, 2015) revealed acceptable psychometric properties of the BASC-3 BESS across all forms, including split-half reliability (.81–.96), test–retest reliability (.82–.89), inter-rater reliability (.52–.67), sensitivity (.66–.76), and specificity (.94–.98). Additionally, the BASC-3 BESS demonstrates convergent validity with the BASC-3 Teacher Rating Scales (r = .43–.95), Achenbach System of Empirically Based Assessment (r =.62–.78), Conners 3 (r =.62–.66), Children's Depression Inventory (r =.51–.64), and Revised Children's Manifest Anxiety Scale (r =.39–.51). Further, in a preliminary investigation, Naser and Dever (2020) found validity evidence of BASC-3 BESS concurrent and predictive relations to student reading scores, adding to the empirical support of this measure.

Procedures

Study procedures were approved by the university's Institutional Review Board and school district administrators. After approvals were granted, schools and teachers were recruited and provided detailed information about requirements for participation in the study. All teachers provided active consent, while parents were given the opportunity to opt their child out of participation in the screening portion of the investigation. Students were considered eligible for participation in the screening if a signed opt out form was not received. Consenting teachers then completed the BASC-3 BESS for each student on their roster who were eligible for participation. An electronic version of the BESS (via Qualtrics) was distributed to teachers at two schools. Due to site-specific preferences, a paper version of the BASC-3 BESS was distributed to the remaining two schools and later entered into Qualtrics by members of the research team. With the exception of the students' age, which was required to calculate a BASC-3 BESS score, researchers did not collect students' identifying information. Teachers were given a small monetary compensation for their time.

Data Analysis Plan

Research question 1. A series of confirmatory factor analyses (CFAs) were conducted to examine the extent to which the developer-proposed factor model fits the BASC-3 BESS data. The model represented a higher order factor structure, which was specified in accordance with developer recommendations. A series of items were specified as loading on three narrow first-order factors: Externalizing Risk Index (ERI; six items), Internalizing Risk Index (IRI: six items), and Adaptive Skills Risk Index (ARI; five items). All first-order factors then loaded on a broad second-order factor corresponding to the Behavioral and Emotional Risk Index (BERI). Three additional items then loaded onto the second-order factor but none of the first-order factors. A review of item content suggested these three items represented symptoms of attentional difficulties, which could be aligned with broader behavioral and emotional functioning even if not aligned with the domains assessed via the subscales.

We examined a series of fit statistics to evaluate the extent to which this factor model fits the observed BASC-3 BESS data. Model fit was evaluated through multiple fit statistics, including the chi-square goodness-of-fit test, Tucker–Lewis Index (TLI), Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), and standardized root-mean-square residual

(SRMR). Observed fit statistic values were compared with the following cutoffs in evaluating model fit: chi-square goodness-of-fit test, $p > .05$; RMSEA $\leq .08$; CFI/TLI $\geq .90$; and SRMR $\leq .08$ (Little, 2013). Modification indices were then examined to determine whether any revisions to the factor model might enhance its fit to the data. All potential revisions were considered relative to theory and the BASC-3 BESS conceptual framework in determining their appropriateness. Once a final model was specified and selected, we examined the factor loadings resulting from that model, which represented the standardized (0–1) relationship between each item and factor, as well as each first-order and second-order factor.

A series of omega statistics were calculated to estimate the internal consistency reliability of each first- and second-order factor. The first statistic corresponded to the total omega coefficient (ω), which assumed the unidimensionality of each factor and did not control for variance specific to any other factor. This statistic was calculated for all four of the BASC-3 BESS scales examined within this data analytic plan. The second statistic was hierarchical omega (ω_H), which represents the variance attributable to a general factor after accounting for the variance explained by narrow factors. Within the context of this study, hierarchical omega represented the reliability of the second-order BERI factor after accounting for the three first-order factors (i.e., ERI, IRI, and ARI). The third statistic was omega subscale (ω_S), which represents the variance attributable to a narrow factor after accounting for the variance explained by a general factor. Within the context of this study, subscale coefficients represented the reliability of each first-order factor after accounting for the second-order factor. In accordance with recommended interpretive guidelines, acceptable reliability was defined as coefficients greater than .70 for ω values (Dueber & Toland, 2021). In addition, ω_H and ω_S values > 0.50 were considered acceptable, whereas values >0.75 were considered strong (Reise et al., 2013).

Research question 2. Measurement invariance was then examined via a series of multi-group CFA models. Of interest were the configural, metric, and scalar invariances of the BASC-3 BESS factor structure. The grouping factor was a binary indicator of *age level*, which represented a roughly even split of the sample in terms of age. Students ages 5–8 were placed in the "Lower" elementary age group (n = 815.55% of the total sample), and students ages 9–11 were placed in the "Upper" elementary age group (n = 657.45% of the total sample). We originally considered using student age as a grouping factor with seven levels. However, we identified empty cells for the "Almost Always" Likert-scale response category for student ages 5 and 11 across four items: Item 1 (66 students), Item 4 (25 students), Item 13 (63 students), and Item 20 (33 students). As empty cells are not permitted within any level of a grouping factor in a multi-group CFA, it was necessary to consider alternative approaches to retain the full sample (as opposed to removing student ages 5 and 11). We ran a multi-group CFA with collapsed categories within the four aforementioned items (DiStefano et al., 2021); however, we encountered model convergence issues. Thus, as an alternative approach, we collapsed the age categories into a single binary age level variable and used this as the grouping factor. This approach permitted examination of the consistency of BASC-3 BESS interpretation across lower and upper elementary students.

The first multi-group CFA examined *configural invariance*, defined as whether the same factor structure fits the data at each age level (i.e., Lower and Upper). This form of invariance was tested by specifying that each item loaded on the same factor (and each first-order factor loaded on the second-order factor) across each age level. The second analysis examined *metric invariance*, defined as the extent to which the magnitude of item-factor relations was equivalent across age levels. To test metric invariance, the model specified that items loaded on the same factor across age levels (i.e., in accordance with configural invariance), while also constraining factor loadings to be equal across groups. The third multi-group CFA examined *scalar/threshold invariance*, or whether groups were equal in terms of the probability of individuals shifting between Likert-type

response categories (e.g., *never* and *sometimes*) given comparable scores on the latent factor. Scalar/threshold invariance was tested by specifying a model that included the same equality constraints as the previous two models while also constraining item thresholds to be equal. Each particular form of invariance was supported if the decline in the RMSEA and CFI fit statistics for a particular model relative to a preceding model did not exceed .01 (i.e., ΔRMSEA $\leq$ .01 and ΔCFI $\geq$ –.01; Keith, 2014).

All CFA and invariance testing were conducted using the CFA function within the lavaan package (Version 0.6-5; Rosseel, 2012) through the computer program R Version 4.0 (R Core Team, 2020). CFAs and multi-group CFAs were conducted using the WLSMV estimator with NLMINB optimization. The fit statistics specifically examined were the robust values resulting from each analysis. Omega coefficients were calculated using the omega function within the psych package in the computer program R (Revelle, 2020).

Results

Research Question 1

A review of fit statistics resulting from the initial CFA suggested the developer-recommended higher-order factor structure was a good fit to the data, with three fit statistics meeting their threshold for acceptability: CFI = .98, TLI = .98, and SRMR = .06. The sole fit statistic that did not suggest adequate fit was RMSEA, which was equal to .10 (90% confidence interval = .09–.10). Factor loadings resulting from this model indicated all items were highly related to their corresponding first-order factors (≥.75), as were the first-order factors to the second-order factor (≥.66; see Table 2). A review of modification indices indicated that permitting Item 8 to cross-load on both the ERI and IRI factors would substantially improve fit. This cross-loading was considered permissible given the nature of the item (i.e., "Changes mood quickly"), which represented a symptom of both externalizing and internalizing concerns. Model fit improved after permitting this cross-loading, with RMSEA and SRMR decreasing by .02 and .01, respectively, while CFI and TLI went unchanged. Given this relatively modest improvement in fit, and that three of the four fit statistics demonstrated acceptable fit for the initial model, the decision was made to retain the initial developer-recommended factor model for measurement invariance testing. A review of correlations among the raw index scores (Table 3) indicated that while the subscale index scores were highly correlated with the total (r = .73–.88), they were only moderately correlated with each other (r = .42–.62). This latter finding speaks to the independence and unique contribution of each subscale index.

A review of total omega coefficients supported the internal consistency reliability of items within each of the factors, with BERI ω = 0.97, ERI ω = 0.96, IRI ω = 0.95, and ARI ω = 0.92. A review of omega hierarchical and subscale omega coefficients suggested that the reliability of scores within factors was reduced after accounting for the other factors. The BERI factor (ω_H = 0.76) continued to demonstrate strong reliability (>.75), while the ARI (ω_S = 0.60) factor demonstrated acceptable reliability (>.50). The remaining two factors did not reach the threshold for acceptable reliability (ERI ω_S = 0.22 and IRI ω_S = 0.33).

Research Question 2

Multi-group CFA findings supported the configural, metric, and scalar invariance of the factor structure across age levels (see Table 4). Rather than decrease, both RMSEA and CFI values remained stable or improved across configural and metric models, suggesting the constraints imposed through each factor model did not diminish fit. Fit decreased only slightly for the scalar

Table 2. Factor Loadings for Each BASC-3 BESS Item and Corresponding Factors.

	BERI	ERI	IRI	ARI
Externalizing	.89			
3		.94		
6		.96		
9		.92		
11		.96		
15		.93		
18		.97		
Internalizing	.66			
1			.75	
4			.83	
8			.98	
12			.93	
13			.90	
20			.83	
Adaptive Skills	.83			
2				.92
5				.92
7				.88
17				.90
19				.92
10[a]	.94			
14[a]	.93			
16[a]	.89			

[a]Items not affiliated with any narrow first-order factor.

model (i.e., RMSEA and CFI $\Delta = .01$), thus supporting scalar invariance. Taken together, results suggest all three forms of measurement invariance were present within the dataset across age levels.

Discussion

A large proportion of youth experiencing SEB challenges go undetected or do not receive services to address their concerns (Kataoka et al., 2002). In schools, universal SEB screening may offer a more appropriate, equitable, and efficient means by which to identify youth experiencing SEB challenges (Dowdy et al., 2011a; Eklund et al., 2017). However, the utility of screening tools rests on the assumption these measures accurately and reliably assess constructs relevant to SEB functioning. The BESS is a popular screening measure, designed to assess externalizing, internalizing, and adaptive behaviors among school-aged youth (Reynolds & Kaphaus, 2015). Initial studies have supported the psychometric properties of the BASC-2 BESS (Dever et al., 2012); however, less evidence is available on the third edition, especially examining measurement invariance which is important in determining if the constructs (e.g., internalizing problems, externalizing problems, and adaptive behaviors) measured are consistent across samples and students of various ages (Edyburn et al., 2020).

The first research question examined the extent to which the developer-proposed factor structure (Kamphaus & Reynolds, 2015) fits the BASC-3 BESS data. Results of the initial CFA

Table 3. Intercorrelations Among the BASC-3 BESS Index Scores.

BESS Index	1	2	3	4
1. Externalizing	1.00			
2. Internalizing	0.52	1.00		
3. Adaptive Skills	0.62	0.42	1.00	
4. Total	0.88	0.73	0.84	1.00

Table 4. Fit Statistics From Confirmatory Factor Analyses.

	RMSEA (CI-90)	CFI	TLI	SRMR
Baseline	.10 (.09–.10)	.98	.98	.06
Configural	.09 (.09–.10)	.98		
Metric	.08 (.08–.09)	.99		
Scalar	.09 (.08–.09)	.98		

Note: RMSEA = Root mean square error of approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; SRMR = Standardized root-mean-square residual.

indicated a four-factor structure (i.e., Externalizing Risk, Internalizing Risk, Adaptive Skills Risk, and Behavioral and Emotional Risk) fits the observed BESS data. BASC-3 BESS items loaded onto their respective first-order factors (e.g., Externalizing Risk, Internalizing Risk, and Adaptive Skills Risk) and the first-order factors then loaded onto the broader second-order factor of Behavior and Emotional Risk. Three of the four fit statistics for the initial CFA supported the developer-proposed four-factor structure. Notably, cross-loading one particular item (e.g., Item 8) onto two factors, Externalizing and Internalizing Risk, improved the fit statistics, RMSEA and SRMR. However, this improvement did not extend to other fit statistics, and the researchers decided to retain the original developer-proposed factor structure as a result. Given the modest improvement from this cross-loading, further research should investigate potential cross-loadings to inform factor structure revisions.

Overall, the results from the initial CFA are generally consistent with previous research, supporting a four-factor structure for BESS data (Dowdy et al., 2011a; Dowdy et al., 2011b; Harrell-Williams et al., 2015; Wiesner & Schanding, 2013). When considered in relation to the reliability findings, these results provide evidence that the BASC-3 BESS is a valid measure of general emotional and behavioral risks. In contrast, some concerns are noted for the interpretation of the subscale indexes. Though total omega coefficients reached the threshold for acceptable reliability, two of the subscale omega coefficients did not, including those specific to the ERI and IRI. Such a finding is consistent with recent studies, which have found inconsistent evidence for the interpretation of subscale scores once a total score is considered (Canivez et al., 2020; Kilgus et al., 2021). It is strongly recommended that this issue be considered through additional research on the BASC-3 BESS to offer guidance to schools regarding the most defensible approach to interpreting scores. Documentation of similar findings in the future could ultimately call into question the utility of subscale scores from universal screening measures. This would suggest that practitioners place more emphasis on interpreting overall risk scores versus prioritizing subscale scores.

The results of the multi-group CFA supported the configural, scalar, and metric invariances of the developer-proposed four-factor structure across elementary-aged student groups (i.e., Upper

and Lower). These results indicate the items on the BASC-3 BESS function similarly for elementary school-aged children, regardless of their age. These results are consistent with other screening research results which indicate student characteristics do not contribute to measurement invariance (Harrell-Williams et al., 2015). Together, the results of the multi-group CFA extend previous resource of the BESS to include the Teacher Form, suggesting the BASC-3 BESS Teacher Form is useful in assessing SEB risk among a sample of elementary school-aged students.

Implications for Practice and Research

Educators and school psychologists tasked with selecting universal screening measures within MTSS frameworks must make key decisions in selection, including what core constructs they are interested in identifying. The current study provides initial evidence that the BASC-3 BESS Teacher Form measures three primary areas of SEB functioning, including Internalizing Behaviors, Externalizing Behaviors, and Adaptive Functioning. These areas are important to assessing core constructs of children's functioning (Frick et al., 2009). The results from the current study provide further support of practitioner's use of the developer-proposed factor structure and interpreting data generated from screening measures in the fashion currently described in the test manual. Finally, present study results revealed that the factor structure held up across multiple grade levels so it will work as well for children in early elementary grades (e.g., kindergarten and first grade) as it will for those in later elementary grade levels. This aligns with previous research that has examined the parent and student self-report data (Edyburn et al., 2020; Kim & Kamphaus, 2018). However, additional research is needed to continue to investigate the BASC-3 BESS Teacher Form across other elementary student samples, as well as to investigate its use in middle and high school student samples.

Limitations

The current study examined an elementary student population from two school districts; therefore, future studies including adolescent samples are also needed as the BASC-3 BESS is designed as a K-12 measure. Second, findings also may not generalize to students from other schools with different characteristics (e.g., racial/ethnic composition) of the student body, so additional research with more diverse student samples is warranted. Third, additional research is needed to determine discriminant and convergent validity evidence for the BASC-3 BESS, and for potential subscale scores to further support score inferences. Researchers are also encouraged to examine validity evidence of the relation between the BASC-3 BESS and other variables to determine if it predicts important outcomes, such as behavioral and academic outcomes (e.g., office disciplinary referrals, grades, and test scores). Regardless of the need for future research, current results indicate that the teacher-rated screener can provide valid information about the core constructs associated with SEB concerns among elementary-aged youth.

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 R324A190046 to the University of Wisconsin-Madison. .

Author Note

The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence concerning this article should be addressed to Katie Eklund, Ph.D., Department of Educational Psychology, University of Wisconsin-Madison, 1025 W. Johnson Street, Madison, WI 53706; katie.eklund@wisc.edu; (303) 246–1968.

ORCID iD

Katie Eklund https://orcid.org/0000-0001-5286-9291

References

Bradley, R., Henderson, K., & Monfore, D. A. (2004). A national perspective on children with emotional disorders. *Behavioral Disorders*, *29*(3), 211–223. https://doi.org/10.1177/019874290402900301

Bruhn, A., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavior screening practices in K-12 schools. *Education and Treatment of Children*, *37*(4), 611–634. http://doi.org/10.1353.etc.2014.0039

Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC-V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, *27*(2), 274–296. https://doi.org/10.1177/1073191118811609

Chin, J. K., Dowdy, E., & Quirk, M. P. (2013). Universal screening in middle school: Examining the behavioral and emotional screening system. *Journal of Psychoeducational Assessment*, *31*(1), 53–60. http://doi.org/10.1177/0734282912448137

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology*, *51*(1), 117–128. https://doi.org/10.1016/j.jsp.2012.09.005

Dever, B. V., Mays, K. L., Kamphaus, R. W., & Dowdy, E. (2012). The factor structure of the BASC-2 behavioral and emotional screening system teacher form, child/adolescent. *Journal of Psychoeducational Assessment*, *30*(5), 488–495. https://doi.org/10.1177/0734282912438869

DiStefano, C., Shi, D., & Morgan, G. B. (2021). Collapsing categories is often more advantageous than modeling sparse data: Investigations in the CFA framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(2), 237–249. https://doi.org/10.1080/10705511.2020.1803073

Dowdy, E., Chin, J. K., Twyford, J. M., & Dever, B. V. (2011a). A factor analytic investigation of the BASC-2 behavioral and emotional screening system arent form: Psychometric properties, practical implications, and future directions. *Journal of School Psychology*, *49*(3), 265–280. https://doi.org/10.1016/j.jsp.2011.03.005

Dowdy, E., Twyford, J. M., Chin, J. K., DiStefano, C. A., Kamphaus, R. W., & Mays, K. L. (2011b). Factor structure of the BASC–2 behavioral and emotional screening system student form. *Psychological Assessment*, *23*(2), 379. https://doi.org/10.1037/a0021843

Dueber, D. M., & Toland, M. D. (2021). A bifactor approach to subscore assessment. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000459

Edelbrock, C., & Achenbach, T. M. (1980). A typology of child behavior profile patterns: Distribution and correlates for disturbed children aged 6–16. *Journal of Abnormal Child Psychology*, *8*(4), 441–470.

Edyburn, K. L., Dowdy, E., DiStefano, C., Bertone, A., & Greer, F. (2020). Measurement invariance of the English and Spanish BASC-3 behavioral and emotional screening system parent preschool forms. *Early Childhood Research Quarterly*, *51*(2), 307–316. https://doi.org/10.1016/j.ecresq.2019.12.002

Eklund, K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health*, *6*(1), 40–49. http://doi.org/10.1007/s12310-013-9109-1

Eklund, K., Kilgus, S., von der Embse, N., Beardmore, M., & Tanner, N. (2017). Use of universal screening scores to predict distal academic and behavioral outcomes: A multi-level approach. *Psychological Assessment*, *29*(5), 486–499. http://doi.org/10.1037/pas0000355

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *California School Psychologist*, *14*(1), 89–95. https://doi.org/10.1007/BF03340954

Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology*, *21*(4), 1111–1131.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117–135. https://doi.org/10.1016/j.jsp.2006.05.005

Harrell-Williams, L. M., Raines, T. C., Kamphaus, R. W., & Dever, B. V. (2015). Psychometric analysis of the BASC–2 Behavioral and Emotional Screening System (BESS) student form: Results from high school student samples. *Psychological Assessment*, *27*(2), 738. https://doi.org/10.1037/pas0000079

Ikeda, M., Nessen, E., & Witt, J. (2008). Universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1921–1932). National Association of School Psychologists.

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., Setmeyer, A. J., Rodelo, C., Newcomb-McNeal, E., & Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary schools. *Contemporary School Psychology*, *18*(4), 241–254. https://doi.org/10.1007/s40688-014-0026-6

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 behavioral and emotional screening system manual*. Pearson.

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior assessment system for children–Third Edition (BASC-3): Behavioral and emotional screening system (BESS).* Pearson.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement*, *67*(1), 1–15. https://doi.org/10.1177/0013164407067002100 1

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*, *159*(9), 1548–1555. https://doi.org/10.1176/appi.ajp.159.9.1548

Keith, T. Z. (2014). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. Routledge.

Kilgus, S. P., Eklund, K., von der Embse, N. P., Weist, M., Barber, A. J., Kaul, M., & Dodge, S. (2021). Structural validity and reliability of social, academic, and emotional behavior risk screener–student rating scale scores: A replication study. *Assessment for Effective Intervention*, *46*(4), 259–269. https://doi.org/10.1177/1534508420909527

Kim, J., & Kamphaus, R. W. (2018). Investigation of factor structure and measurement invariance by gender for the behavioral and emotional screening system among high school students. *Psychological Assessment*, *30*(2), 231–240. https://doi.org/10.1037/pas0000469

King, K. R., Gonzales, C. R., & Reinke, W. M. (2019). Empirically derived subclasses of academic skill among children at risk for behavior problems and association with distal academic outcomes. *Journal of Emotional and Behavioral Disorders*, *27*(3), 131–142. https://doi.org/10.1177/1063426617754082

Little, P. T. D. (2014). Longitudinal structural equation modeling. Guilford Press.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316. http://doi.org/10.3102/0013189X14545513

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjhet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the

national comorbidity survey replication–adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989. https://doi.org/10.1016/j.jaac.2010.05.017

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

Naser, S. C., & Dever, B. V. (2020). A preliminary investigation of the reliability and validity of the BESS-3 teacher and student forms. *Journal of Psychoeducational Assessment*, *38*(2), 263–269. https://doi.org/10.1177/07.34282919837825

National Center for Educational Statistics, NCES (2020). School search. https://nces.ed.gov/ccd/schoolsearch/

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, *60*, 65–82. http://doi.org/10.1016/j.jsp2016.11.002

R Core, Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: Co-Occurrence and distal outcomes. *Journal of Abnormal Child Psychology*, *36*(1), 759–770. https//doi.org/10.1007/s10802-007-9208-2

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Revelle, W. (2020). Procedures for psychological, psychometric, and personality research (R package Version 2.0.9) [Computer software manual]. http://personality-project.org/

Reynolds, C. R., & Kaphaus, R. W. (2015). *Behavior assessment system for children–third edition (BASC-3)*. Pearson.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology*, *45*(2), 193–223. https://doi.org/10.1016/j.jsp.2006.11.003

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavior Disorders*, *13*(2), 79–96. https://doi.org/10.1177/10634266050130020201

Wiesner, M., & Schanding, G. T. (2013). Exploratory structural equation modeling, bifactor models, and standard confirmatory factor analysis models: Application to the BASC-2 behavioral and emotional screening system teacher form. *Journal of School Psychology*, *51*(6), 751–763. https://doi.org/10.1016/j.jsp.2013.09.001