

# Examining the BASC-3 BESS Parent Form–Preschool Using Rasch Methodology

Assessment  
2019, Vol. 26(6) 1162–1175  
© The Author(s) 2017  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1073191117723112  
journals.sagepub.com/home/asm



Christine DiStefano<sup>1</sup>, Fred W. Greer<sup>1</sup>, and Erin Dowdy<sup>2</sup>

## Abstract

This study examined the Behavior Assessment System for Children–Third Edition Behavioral and Emotional Screening System, Parent Form–Preschool, which is used to identify children at risk for emotional/behavioral problems. Rasch rating scale methods were used to provide information about psychometric properties of items, parent raters, and the response scale. Using the norming sample of 459 children, we investigated the structure of the screener, performance of the scale, and coverage of the maladaptive behavior construct to identify children with emerging behavioral difficulties. Most items showed good fit to the rating scale methods model with broad coverage of the latent domain. Items measuring internalizing and externalizing problems indicated a greater tendency for at-risk behavior, and items measuring adaptive skills offered a minimal contribution to measuring risk status. No differential item functioning was observed due to child gender; however, differential item functioning was observed between ratings for Caucasian and Hispanic or African American students for select items.

## Keywords

screening, preschools, BASC, Rasch, Response to Intervention

During the last four decades, the number of preschoolers served in center-based programs (e.g., child care centers, Head Start Programs, publicly funded prekindergarten (pre-K) programs, private preschools) has increased dramatically (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015). Present figures suggest that more than 1.3 million children nationwide are attending state-funded preschool programs and another 1.1 million preschoolers are attending Head Start Programs. While these figures are impressive, there is a push to dramatically increase children's access to pre-K services through the "Preschool for All" initiative targeted to provide preschool services to all 4-year-olds in the United States. Thus, projections regarding numbers of children to be served and expenditures dedicated to preschool education are both expected to rise (cf. Barnett et al., 2015).

Concurrent with the expansion of services to young children has been the enrollment of many who exhibit emerging behavioral and social-emotional difficulties (Brown & Conroy, 2011; Webster-Stratton & Reid, 2013). Although variable, contemporary prevalence rates indicate that from 10% of preschoolers (Forness, Freeman, Paparella, Kauffman, & Walker, 2012) to as many as 25% of young children have behavioral and social-emotional difficulties severe enough to impede their social competence and educational experience (e.g., Campbell, 2002; Conroy, Hendrickson, & Hester, 2004; Webster-Stratton, 2005). National information has indicated that preschoolers were

more than 3 times likely to be expelled than older school-age children (Gilliam & Shahar, 2006). The prevalence of preschool children's problem behaviors, if not dealt with early and effectively, is a continuing concern among many practitioners, administrators, and parents, especially as children enroll in greater numbers in preschool programs (cf. Brown & Conroy, 2011; Webster-Stratton & Reid, 2013).

One way to assist young children at risk for problems is early and effective behavioral intervention at school. While prevention and early intervention methods have been implemented with older children, similar services are not routinely extended to preschool-aged children (e.g., Conroy & Brown, 2004; McConnell, Carta, & Greenwood, 2008). Even when early intervention or prevention services are offered, efforts to identify young children with emerging behavioral and social-emotional problems have not been systematically or widely established (Biglan, 2004; Campbell, 2002; Conroy & Brown, 2004; Webster-Stratton & Reid, 2008). However, the success of early intervention or prevention services depends on the ability to accurately identify the needs of the children involved (Ward & Rothlisberg, 2011). Recently, a

<sup>1</sup>University of South Carolina, Columbia, SC, USA

<sup>2</sup>University of California Santa Barbara, Santa Barbara, USA

## Corresponding Author:

Christine DiStefano, Department of Educational Studies, University of South Carolina, Columbia, SC 29208, USA.

Email: distefan@mailbox.sc.edu

Response to Intervention (RtI) approach has been proposed for preschool children (McConnell et al., 2008). The overall goal of RtI is to implement a systematic approach to lessen the overall negative influences and outcomes of children's early school-related problems (i.e., academic, behavioral, and social-emotional) and disabilities. Often, the first stage in a school-based RtI system is to implement activities to assess all children (i.e., Tier 1) to further determine who may be in need of additional, more targeted, or indicated services. Universal screening is frequently conducted for Tier 1 activities, given that it is a relatively quick method to conduct school-wide assessment.

To accomplish universal screening at the preschool level, information may be obtained by preschool teachers or parents at the start of the academic year. By gathering information at entry into preschool, school personnel could be well situated to consider how to best support incoming students. As preschool may be the first place where parents have access to behavioral support services, it may be critically important to gather information from parents within a universal screening practice. Additionally, given that behavioral/emotional instruments typically require that raters evaluate behavior in the past 4 to 6 weeks, universal screening could be completed by parents at the start of, or even before, the academic year. This would provide information for RtI services earlier than could be obtained from teacher raters. Additionally, a comprehensive review of the literature indicates that parents may be the best informants for younger children's internalizing and externalizing functioning (S. R. Smith, 2007).

While there are many instruments available for screening preschool-aged children, we focus on the newly revised edition of the Behavior Assessment System for Children—Third Edition (BASC-3) Behavioral and Emotional Screening System, Parent Form—Preschool (BESS Parent-P; Kamphaus & Reynolds, 2015). This form was recently revised along with the larger family of BASC-3 behavioral instruments (Reynolds & Kamphaus, 2015). In terms of structure, the BESS Parent-P was designed to include nine items for each of three different dimensions of behavior (i.e., Externalizing Problems, Internalizing Problems, Adaptive Skills). The Externalizing Problems dimension measures children's tendencies to display aggressive or hyperactive behaviors toward peers or adults (e.g., "hits other children"). The Internalizing Problems dimension includes children's tendencies to show feelings of anxiety, worry, or stress (e.g., "is easily upset"). This area includes behaviors that are not marked by acting out behaviors but, instead, are associated with dysregulated emotions. Finally, the Adaptive Skills dimension measures how students develop socially and interact with peers and authority figures. These skills represent positive behaviors and appropriate social skills among young children (e.g., "responds appropriately when asked a question"). In addition, two items measuring Attention Problems are included on the

form; these items do not belong to any dimension. However, all items are presumed to load onto one overarching Maladaptive Behavior construct.

The BASC-3 BESS Parent-P has undergone significant revisions to enhance content and clarity. While the number of items on the BASC-3 BESS Parent-P form has only been reduced by one—from 30 to 29 items, the content of over half of the items on the screener has changed since the previous edition of the BASC-2 BESS Parent-P (Kamphaus & Reynolds, 2007). The revision process resulted in dropping items from the BASC-2 BESS Parent-P that were not working acceptably due to issues such as suboptimal psychometric performance and problems due to clarity or interpretability (Kamphaus & Reynolds, 2015). Specifically, the new edition of the BASC-3 BESS Parent-P includes more items representing the Adaptive Skills dimension and fewer items measuring Externalizing Problems and Attention Problems.

To use the screener, parents rate the frequency of observed behaviors over the past 6-week period. Items on the BASC-3 BESS Parent-P are rated using a 4-point Likert-type scale of 0 = *never* (behavior not observed), 1 = *sometimes*, 2 = *often*, and 3 = *almost always*. Screener scores are computed for each of the three dimensions and for the overall scale (including the two Attention Problem items) by creating a sum score of item ratings. The total score is transformed to a *T*-score ( $M = 50$ ,  $SD = 10$ ). BESS Parent-P total scores classify children at risk for developing behavioral and emotional problems, where higher scores represent greater amounts of maladaptive behavior. In addition, a classification system is used, in which total scores are classified into risk levels of *Normal* development (*T*-scores of  $\leq 60$ ), *Elevated* risk (*T*-scores of 61-70), and *Extremely elevated* (*T*-scores of  $\geq 71$ ) risk.

The BASC-3 BESS manual reports validity coefficients between the screener scores, by total and by dimension, with established measures of behavioral and emotional problems for young children. Correlation values were moderate and in the expected direction with scores from similar scales on the Child Behavior Checklist for Ages 1½-5 (.59-.71; Achenbach & Rescorla, 2001). In addition, reliability estimates illustrated consistency of responses, with values of at least .84 for internal consistency, at least .81 for test-retest reliability (average of 3 weeks between testings), and at least .63 for interrater reliability among parents.

Given that the BASC-3 BESS Parent-P is a newly revised form, investigations of the structure and performance of the form are warranted (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The manual provides only factor analytic evidence relating strength of the item relative to the latent construct (i.e., loading values). While factor analysis is often used in scale development, some researchers have noted limitations to this method (E. V. Smith, Conrad, Chang, & Piazza, 2002). For example, under factor analysis, estimates of a child's

placement on the latent construct are item- and sample-dependent. Although factor analysis is widely used to provide validity evidence, this method does not identify the coverage of items along the latent domain, examine acceptability of the response scale, or predict how an individual may respond to an item due to his or her level of the construct of interest. In addition, total scores for the BESS Parent-P are created by summing ordinal Likert-type items. This procedure may be inappropriate, as the assumption of equal intervals required for this operation is unlikely to be met with Likert-type data (E. V. Smith et al., 2002).

Current psychometric practice encourages use of the Rasch model to provide additional information about characteristics of items and samples (e.g., Kahler, Strong, & Read, 2005; Thomas, 2011). This methodology can provide greater information than just information on the relation between an item and latent factor. In addition, analyses of the psychometric properties of new behavior rating scales and screening systems should also examine group equivalence (Tyson, 2004). Differential item functioning (DIF) is present when characteristics of an individual item vary across members of the subgroups who have similar mean levels of the latent trait (Bond & Fox, 2007).

In a previous study, the BASC-2 BESS Teacher Form–Preschool (BESS Teacher-P) was examined using Rasch methodology (DiStefano, Greer, Kamphaus, & Brown, 2014). The findings suggested that the scale was interpreted acceptably by teacher raters. In addition, the items selected for the BESS Teacher-P were adequate for a screening measure, with internalizing behaviors and externalizing problems contributing most to behavioral risk. However, teachers also reported that Internalizing Problem items were the hardest to observe.

While the findings showed support for the BASC-2 BESS Teacher-P, there is need to examine the parent form as well. The BASC-3 BESS has been newly revised, and the Parent-Preschool form has not been investigated in its revised, or its original, format. Thus, examination of the form is warranted to support its use by parents and practitioners for identifying emotional and behavioral risk with young children. Therefore, the purpose of this study is to examine the BASC-3 BESS Parent-P using Rasch rating scale methods (RSM). RSM was used to investigate the structure of the screener, the severity of items relative to maladaptive behaviors, the performance of the Likert-type rating scale, coverage of the latent dimension, and probability of score response. DIF tests of item invariance were also conducted to examine item functioning across gender and racial/ethnic subgroups.

## Method

### Sample

The BASC-3 Parent Rating Scale–Preschool (PRS-P) norming data set used to create the revised BESS Parent-P was used to investigate the scale using the Rasch RSM. As

**Table 1.** Demographic Characteristics of Parents and Children; BASC-3 BESS Parent–Preschool Norming Sample (N = 459).

	Frequency	%
Region of the United States		
Midwest	91	19.8
Northeast	79	17.2
South	174	37.9
West	115	25.1
Parent education		
Less than high school	67	14.6
High school graduate	98	21.4
1-3 years post high school	151	32.9
College graduate or beyond	143	31.2
Child gender		
Female	233	50.8
Male	226	49.2
Child age in years		
3	159	34.6
4	148	32.2
5	152	33.1
Child ethnicity		
Caucasian	235	51.2
Hispanic	118	25.7
African American	63	13.7
Asian	18	3.9
Other	25	5.4
Reported diagnosis		
No	425	92.6
Yes	34	7.4

Note. BASC-3 = Behavior Assessment System for Children–Third Edition; BESS = Behavioral and Emotional Screening System.

noted in the manual, the norming program took place across 44 states (Reynolds & Kamphaus, 2015). The 459 children were between 3 and 5 years of age and were rated by their parent or primary caregiver. Table 1 provides demographic information about parents and children included in the BASC-3 PRS-P norm sample.

The sample of parents were obtained from across the United States, with the majority of parents from the South. Parents varied in their level of education, with most parents (33%) reporting 1 to 3 years post high school education. Of the children rated, the average age was 4.5 years ( $SD = 10.3$  months). The sample was approximately equal across sexes, including approximately 49% boys and 51% girls. While the sample was diverse, slightly over half of the children rated were Caucasian (51.2%). The majority of children in the sample (approximately 93%) did not have a prior clinical diagnosis (e.g., autism, attention deficit hyperactivity disorder) from a doctor or psychologist. Using the total BESS Parent-P scores from the norming database, 87% of the preschoolers were classified with *Normal* behavioral and emotional risk and 13% exhibited at least an *Elevated* level of risk.

## Rasch Methodology Overview

Rasch methods refer to a family of mathematical models that compute the probability of a certain response to each item given the amount of the latent construct the individual possesses (i.e., “ability” in Rasch terminology) and the relation between an item and the construct (i.e., “item severity”). The method scales both persons and items according to the strength of an individual’s relation with the latent construct. The Rasch model produces scores for each person and each item on a common, interval level scale, called a logit (i.e., log-odds) scale. These common scores are called measures, and the process of putting both parameters on the same scale is called calibration. A distinguishing feature of the Rasch model is that a person’s ability and item severity can be separated, making it possible to estimate a person’s level of the latent construct free of the distribution of the individual items and to estimate an item’s difficulty level free from the distribution of people used in the sample (Schumaker, 2004). With the BESS Parent-P, it was of interest to estimate the amount of the behavioral and emotional risk construct that each child possessed and the severity of each item relative to measuring the construct.

The Rasch RSM is a specialized model used with polytomously scored items and ordinal data, such as responses from a Likert-type scale. To accommodate ordinal data, a threshold value is included in the item estimation process. Thresholds can be thought of as the point that moves a person from one category to an adjacent category on the Likert-type scale. The number of thresholds is equal to the number of scale categories ( $k$ ) minus 1. RSM estimates the probability that a person will be observed in a specific category of the rating scale, given both a child’s level of the construct and the endorsability of the item.

The RSM provides scores for both items and persons on a logit (i.e., log-odds) scale. The logit will vary if the probability is computed across all children for an item (item logit) or across items to compute the score for an individual (person logit). In the current situation, higher logit values represent higher levels of risk for maladaptive behavior for children or an item more indicative of emotional or behavioral problems. For more detail about Rasch modeling, there are several excellent texts available (e.g., Bond & Fox, 2007; E. V. Smith & Smith, 2004).

Rasch models have criteria that should be met for accurate estimation: (1) construct unidimensionality, (2) a monotonic scale (i.e., higher latent scores represent a higher level of the latent construct), and (3) that the items adequately fit the Rasch model (Bond & Fox, 2007; Sick, 2010). In addition, RSM requires that each response category ( $k$ ) have a minimum frequency of 10, that the rating scale categories increase in difficulty of endorsement (called step values), and that the thresholds for each item are ordered (E. V. Smith et al., 2002). If the requirements

underlying RSM are met, the model offers the following benefits: (1) a common, interval-level metric for calibrated items and person measures; (2) fit statistics to evaluate items and persons that do not align with the Rasch model (i.e., misfit); (3) estimation of (parent) ratings for an individual (child); and (4) evaluation of the breadth of coverage of the latent construct of Maladaptive Behavior. WINSTEPS Version 3.94 (Linacre, 2017) was used for all analyses.

Analyses were conducted to examine characteristics of items, and functioning of items across subgroups. First, the functioning of the BESS Parent-P as a scale was examined. Person and item reliability indices were computed to determine consistency of ratings. These values may be interpreted similarly to traditional estimates of reliability, for which higher values represent more stability across responses/people and values greater than .80 show high consistency (Crocker & Algina, 1986). To assess unidimensionality of the screener, an unrotated principal components analysis (PCA) of standardized residuals was carried out. The analysis determines if there is additional variance to be explained after the Rasch (or latent) construct has been extracted. The Rasch construct should account for at least 50% of the total variance to be explained, and after accounting for the model, the remaining extracted components should account for a small percentage of the remaining variance (less than 5%; Linacre, 1992).

Second, analyses investigated the relations between items and maladaptive behavior. Two fit indices, Infit mean square and Outfit mean square, were used to assess the fit between items and the Rasch model. These values identify potential unexpected response patterns. Outfit statistics place greater emphasis on unexpected responses far from a person’s or item’s measure, and Infit statistics place greater emphasis on unexpected responses near a person’s or item’s measure (Bond & Fox, 2007). Expected Infit or Outfit mean square values for each item is 1.0, with an acceptable range within 0.5 to 1.5 (Linacre, 2004; Sick, 2010). Values outside of these bounds may suggest a lack of fit between the item and model. Examination of item severity was investigated using item probability plots, and content coverage, through person/variable maps (Bond & Fox, 2007; Linacre, 2006). With screening measures, it is common to select more severe items (i.e., higher difficulty values), which are likely to be endorsed for children at the upper end of the latent construct (e.g., Kahler et al., 2005; Thomas, 2011).

The final set of analyses employed DIF analyses to determine if BESS Parent-P screener items showed bias due to a child’s sex or ethnicity/race. When assessing DIF, the average item measure is computed and compared across groups, while the item measure shows the relative severity score for a subgroup. Two values were examined to determine if an item showed significant bias. First, the DIF contrast (i.e., difference in difficulty of the item between groups) needed to be at least 0.5 logits for DIF to be

noticeable (Linacre, 2006). In addition, the probability value illustrates observing this amount of contrast by chance, when there is no systematic item bias effect needed to be significant. Due to the large number of items compared, alpha (set at .05) was controlled when making comparisons using a Bonferroni comparison. Therefore,  $p$  values had to be less than .002 (i.e., .05 divided by 29 items) with a contrast greater than .5 logits to show evidence of DIF.

## Results

While the BESS Parent-P is thought to measure one overarching construct of maladaptive behavior, it was developed by extracting items from interrelated dimensions from the larger item pool of the BASC-3 PRS. Items on the Adaptive Skills dimension were positively worded relative to at-risk behavior (i.e., higher scores represent lower levels of maladaptive behavior) while items from remaining scales were negatively worded (i.e., higher scores representing higher levels of maladaptive behavior). Following recommendations, items from the Adaptive Skills dimension were reversed prior to analyses, and these items were analyzed by allowing a different response scale (e.g., Linacre, 2006). This practice is used to minimize the presence of a response set due to the direction of the item wording.

### Scale Characteristics

For the entire scale, the screener yielded a person reliability index of .89. This value indicates the replicability of person ordering expected if a parallel set of items were used (Wright & Masters, 1982). The item reliability estimate was .98, indicating the replicability of item placement along the latent dimension if another sample were selected from the same population. Both reliability indices can be interpreted following classical test theory guidelines (Crocker & Algina, 1986) and suggested that the BESS Parent-P produces scores that are stable across children and items. Furthermore, the screener is producing scores that are ordered such that children with higher levels of maladaptive behavior show higher measure scores. Estimates showed that children with a previous diagnosis had higher average logit scores for the overall maladaptive behavior construct ( $M = -0.40$ ,  $SD = 1.31$ ) representing greater risk status than undiagnosed children ( $M = -1.70$ ,  $SD = 1.09$ ). Furthermore, children with a BESS Parent-P rating of at least Elevated risk had higher logit scores ( $M = 0.24$ ,  $SD = 0.70$ ) illustrating higher risk than did children within the Normal development range ( $M = -1.88$ ,  $SD = 0.93$ ).

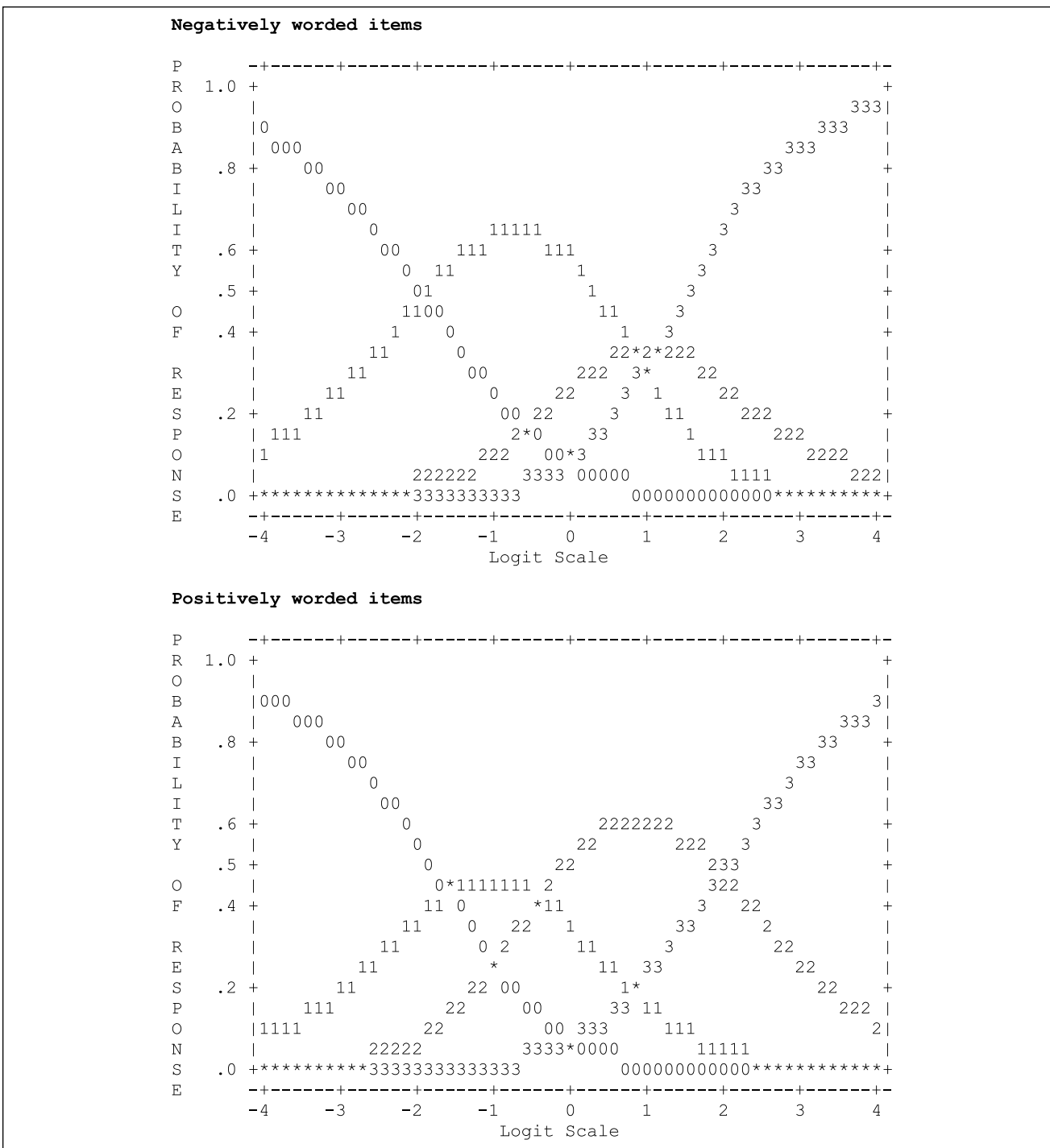
PCA of the standardized residuals showed that the Rasch model accounted for 50.0% of the variance in the observations, just meeting the recommended level. The unexplained variance in the first extracted component

was 8.5%, higher than the recommended lower bound of 5%, meaning that there was still meaningful variance to explain after the underlying dimension was extracted. Investigations of PCA residuals and contrasts plots showed that the positively worded items from the Adaptive Skills dimension wanted to “break apart” into a separate factor from items measuring problematic behaviors, even after using a separate rating scale to account for item wording. A lack of unidimensionality suggests that the BESS Parent-P may be measuring emotional and behavioral risk as well as a secondary dimension related to social/adaptive skills. We recognize that the lack of dimensionality is a limitation to the requirements of RSM and note this caveat when interpreting information.

The practicality of the 4-point rating scale was examined for each set of items, grouped by positive or negative wording. For both groups of items, each item showed a sufficient frequency per item, and step measure information showed that the scale points were ordered, with increasing risk as a higher category was selected. Figure 1 illustrates the use of response categories, by wording direction. Both graphs exhibit the recommended pattern, whereby each scale category is the most probable response (i.e., highest peak) for one ability level; although it is noted that for problematic behaviors (negatively worded items), the *often* (2) category was used less frequently than the other options.

As shown in Figure 1, both negatively and positively worded items showed a similar distributional pattern. In other words, parents were not rating items differently due to the wording direction of the statement (i.e., method effect due to item wording; DiStefano, Morgan, & Motl, 2012). For items denoting problematic behaviors (i.e., negatively worded items), the majority of parents reported very few instances of such behaviors occurring, with most ratings stating that the behaviors were *never* (48%) or *sometimes* (42%) observed. Average measures were computed by scale point, showing the average latent score for children with an item rating of a certain category. These values are shown by an asterisk (\*) between curves. For threshold values in the category points of 0 to 1, 1 to 2, and 2 to 3, the average measures were  $-1.92$ ,  $0.81$ , and  $1.11$ , respectively, for negatively worded items.

With items measuring adaptive skills (i.e., positively worded items), most parents rated their children as exhibiting these skills (note that the scale was reversed prior to interpretation), with ratings of *almost always* (36%) and *often* (36%) occurring over the past 6-week period. For positively worded items, average measure thresholds were  $-1.59$ ,  $-0.38$ , and  $1.97$ , respectively. In sum, the set of information suggests that the 0 to 3 rating scale is generally functioning appropriately; however, the lack of respondents to the *often* category does show difficulty for parents to select this option when rating a child.



**Figure 1.** BASC-3 BESS Parent-Preschool rating scale probability curves for the 4-point scale.  
Note. BASC-3 = Behavior Assessment System for Children—Third Edition; BESS = Behavioral and Emotional Screening System. Numbers refer to the response category on the BESS Parent-P, where 0 = never, 1 = sometimes, 2 = often, and 3 = almost always. Positively worded Items were recoded prior to analyses.

**Item Characteristics**

All BESS Parent-P item-to-total correlations were positive, ranging from .35 to .69, suggesting that as a behavior (denoted by item content) was exhibited with greater

frequency, the higher the child’s risk level for maladaptive behavior. Table 2 provides Rasch item statistics for the BESS Parent-P, ordered by severity. We note that the BESS Parent-P items are referred to their number on the screener

**Table 2.** Rasch Fit Statistics, BASC-3 BESS Parent–Preschool ( $N = 459$ ).

Item	Total score	Item severity	Infit MNSQ	Outfit MNSQ	Item-to-total correlation
INT_14	99	1.92	1.32	1.25	.35
INT_19	131	1.51	1.20	1.12	.40
EXT_11	161	1.19	1.03	.84	.63
INT_16	200	0.83	0.88	0.80	.57
INT_8	210	0.74	1.08	1.06	.43
INT_3	223	0.65	1.13	1.16	.42
EXT_17	236	0.53	1.23	1.08	.61
EXT_7	243	0.49	0.91	0.94	.52
INT_23	246	0.47	1.32	1.28	.44
EXT_24	256	0.40	0.73	0.69	.69
EXT_15	276	0.24	0.93	0.86	.66
ADAPT_5	311	0.15	1.22	1.31	.51
EXT_26	293	0.13	0.97	0.96	.53
ADAPT_28	315	0.13	0.82	0.80	.60
INT_25	297	0.11	0.75	0.77	.61
ADAPT_13	367	-0.15	0.99	0.97	.58
EXT_9	342	-0.18	1.14	1.13	.58
ADAPT_4	426	-0.44	1.24	1.29	.50
ATTENT_20	399	-0.51	0.88	0.88	.60
ADAPT_10	467	-0.62	0.92	0.91	.59
ADAPT_29	467	-0.63	1.29	1.40	.50
EXT_12	429	-0.67	0.85	0.87	.61
INT_27	429	-0.70	0.66	0.68	.68
INT_6	446	-0.76	0.75	0.75	.63
ADAPT_1	497	-0.77	1.01	1.03	.56
ADAPT_18	541	-0.96	1.24	1.28	.45
ADAPT_22	555	-1.02	0.98	0.97	.62
EXT_21	501	-1.04	0.98	1.02	.53
ATTENT_2	505	-1.06	0.95	0.99	.55
M	340.3	0.0	1.01	1.00	
SD	126.6	1.01	0.19	0.20	

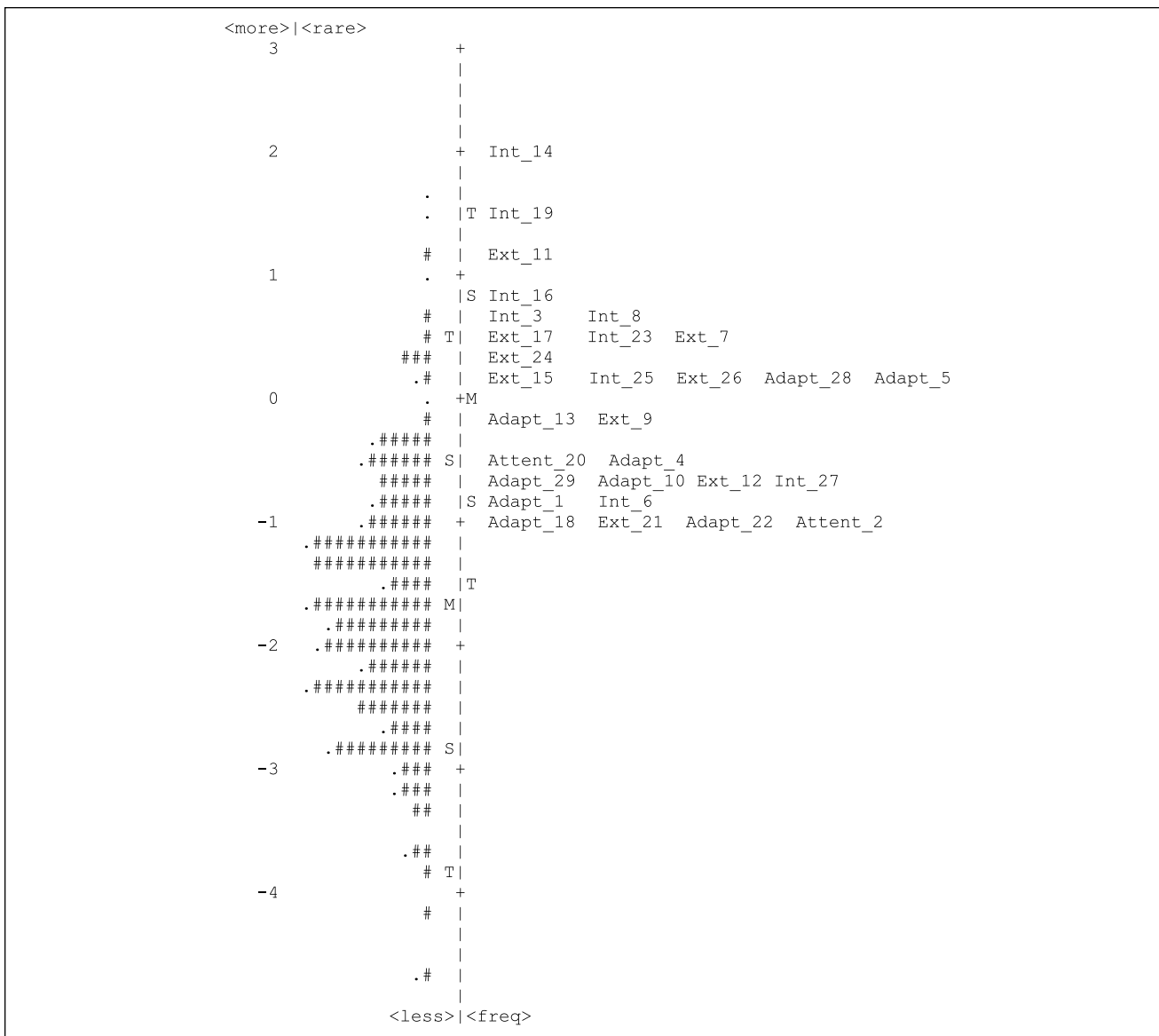
Note. BASC-3 = Behavior Assessment System for Children—Third Edition; BESS = Behavioral and Emotional Screening System; MNSQ = mean square; INT = internalizing problems; EXT = externalizing problems; ADAPT = adaptive skills; ATTENT = attention problems. Total score is computed as the sum of the Likert-type responses for all respondents, with higher scores representing greater frequency of observation.

and the originating dimension (e.g., INT\_14 refers to BESS Parent-P screener Item 14, from the Internalizing Problems dimension). In addition, screener items reported mean squares Infit and Outfit levels within stated guidelines, illustrating acceptable fit to the Rasch RSM.

Calibrated scores for both people and items along the maladaptive behavior construct are provided in a Wright map (Bond & Fox, 2007), shown in Figure 2. On the left side of the map are the person measures, showing the placement of children along the latent dimension of maladaptive behavior according to parent ratings; information about the relation between items and construct is shown along the right side. The map is centered at a score of 0 for the items, and because both sets of measures are on the same scale, person scores can be interpreted relative to the placement of the items. For person and item distributions,

the mean is provided in the center of the distribution with one (S) and two (T) standard deviations from the mean noted.

The left side of the Wright map reports the distribution of measure scores for children in the norm group. The sign shows a child's placement on the dimension, where positive values indicate the presence of more maladaptive behaviors and negative values show fewer behavioral problems. Larger numbers report the greater (or lesser) amounts of at-risk behaviors rated by parents. As seen in Figure 2, risk levels for children in the norm sample are approximately normally distributed with a large spread across the construct. Most preschoolers were rated by parents as well-behaved, as seen by the low average value of the person distribution, reported as  $-1.60$  logits.



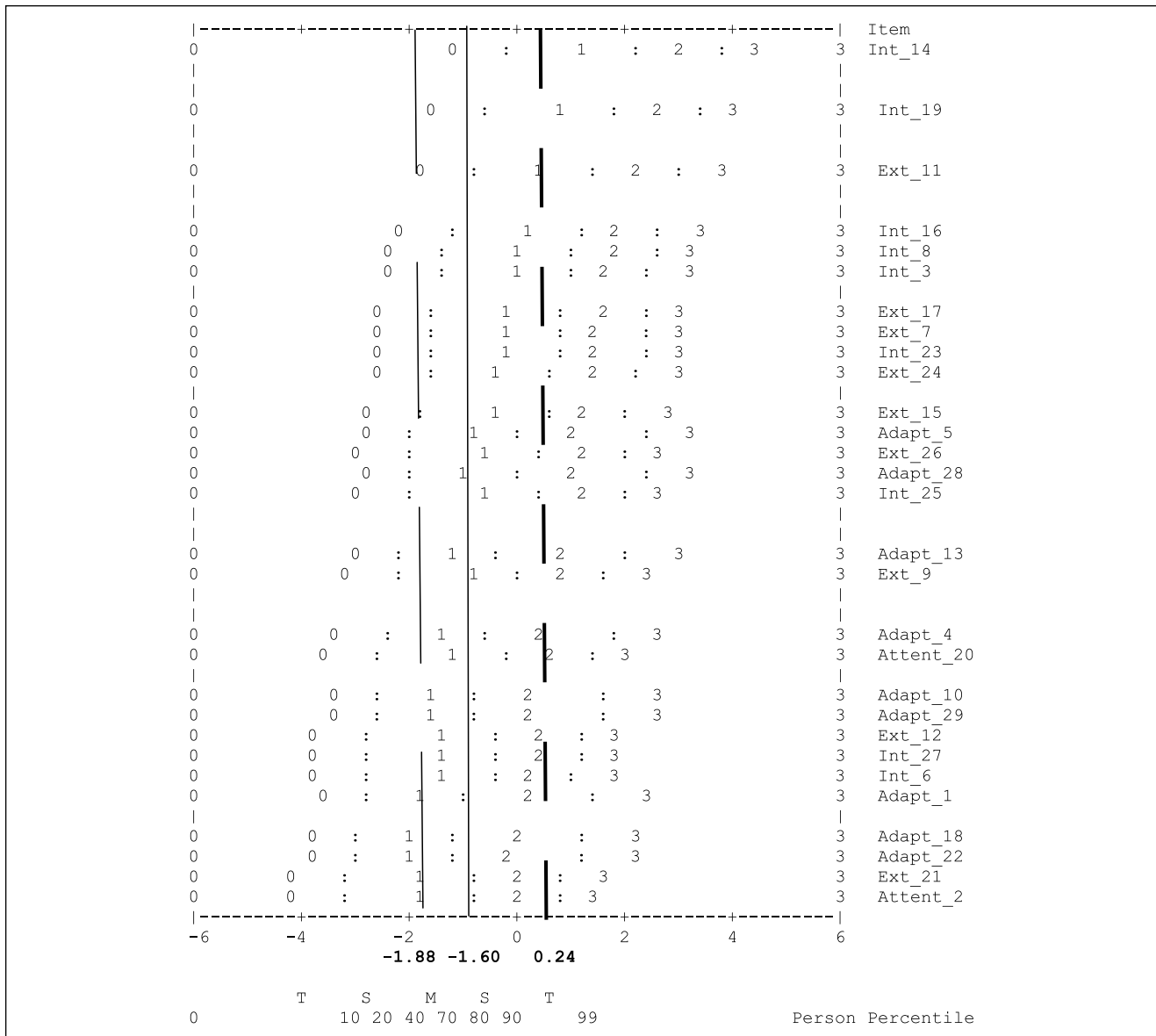
**Figure 2.** Wright map for BASC-3 BESS Parent-Preschool items, 29 items ( $N = 459$ ).  
 Note. BASC-3 = Behavior Assessment System for Children—Third Edition; BESS = Behavioral and Emotional Screening System; Int = internalizing problems; Ext = externalizing problems; Adapt = adaptive skills; Attent = attention problems. Each “#” corresponds to 3 children rated from the Parent Rating Scale—Preschool norm sample, each “.” corresponds to 1 to 2 cases.

The right side of the map reports the severity ratings of BESS Parent-P items. Overall, the set of items illustrated a range of the severity of behaviors measured by the screener, with values spanning roughly two standard deviations above and below the mean. Items at the top of the item distribution are more severe or were harder for parents to have observed in the past 6 weeks and items at the bottom of the scale more likely to have been observed.

All of the items on the BESS Parent-P yielded average logit ratings above the mean of the average child measure. This may be interpreted as the screener is better suited to measure children with higher than average levels of

maladaptive behavior (i.e., latent scores at the upper end of the continuum). Examining item content from the Wright map shows that, in general, items measuring Internalizing Problems are at the top of the scale, meaning they are the hardest for parents to frequently observe. Externalizing Problem items are generally in the middle of the distribution; items measuring Adaptive Skills and Attention Problems were much easier for parents to observe. There is also some overlap in item contribution, illustrated as multiple items on the same line of the graph. Items at the same severity level are not providing unique information in terms of measuring the latent construct.





**Figure 3.** Expected scores on the BASC-3 BESS Parent-Preschool by child measure. Note. BASC-3 = Behavior Assessment System for Children—Third Edition; BESS = Behavioral and Emotional Screening System. Dashed line = children rated at a normal level of risk ( $M = -1.88$ ), solid line = average measure score ( $M = -1.60$ ), bolded dashed line = children rated with at least Elevated risk ( $M = 0.24$ ).

Figure 3 presents the expected item endorsements for children at various risk levels. Along the x-axis, preschoolers' risk levels (i.e., logit scores) are shown; along the y-axis are items from the BESS Parent-P, ordered according to item severity values. Values correspond to the ordinal scale used by the BESS Parent-P, and the colons correspond to threshold values, where a parent would mark the next highest category on the rating scale if the threshold is surpassed.

To read the graph, identify a child measure value that is of interest and track scores upward to determine the expected rating for each item. For example, for a child at the

mean of the person distribution of  $-1.60$  (i.e., solid line), his or her parent would be expected to endorse a score of 1 (*sometimes*) for Item 1 on the Adaptive Skills scale and a score of 0 (*never*) for Item 11 on the Externalizing Problems scale. Parent ratings for most Internalizing Problems and Externalizing Problems items are expected to be 0 (*never*) for a child at an average level of maladaptive behavior.

Figure 3 also separately reports the expected responses for children in the normative sample with reported diagnoses and those without diagnoses. For example, the dashed line reports average scores for children without a previous

diagnosis. Examining the responses for items at this logit level provides an estimate for ratings of a “typical” nondiagnosed preschooler. Similar interpretation may be conducted for children with a previous diagnosis by interpreting expected scores close to dotted line. Also, the bottom of the graph relates the percentile of latent scores from the child distribution according to person logits ( $x$ -axis), making it relatively easy to estimate item responses for child scores along the latent dimension and to estimate the percentile associated with a given latent score.

Using the BESS Parent-P method of summing item raw scores to provide a total score, a child at the mean of the distribution would be expected to have a sum score of approximately 25. It is easy to see from Figure 3 what expected parent ratings would be for any of the BESS Parent-P items given different levels of maladaptive behavior. In general, a child’s measure score would need to be greater than a measure of 2 to receive many scores of *often* (2) or *always* (3) on Internalizing and Externalizing Problems items.

### Differential Item Functioning

DIF results for gender showed that BESS Parent-P items could be considered invariant based on the criteria used. There were, however, select items that exhibited DIF between Caucasian children and either African American or Hispanic children, as rated by parents.

Across groups, 7 of the 29 BESS Parent-P items (24%) reported evidence of DIF between ratings of Caucasian children and African American or Hispanic children. One item, measuring Adaptive Skills (“responds appropriately when asked a question”) showed DIF between Caucasian students and both African American and Hispanic children. With African American students, DIF was observed between three Externalizing Problem items and one Adaptive Skills item. For the Externalizing Problem items, African American children were rated by their parents higher than Caucasian children, illustrating higher levels of parent-reported problematic behavior (“throws tantrums,”  $p = .0007$ , contrast =  $-0.64$ ; “loses temper too easily,”  $p = .002$ , contrast =  $-0.98$ ; “argues when denied own way,”  $p = .0001$ , contrast =  $-0.93$ ). The Adaptive Skills item (noted above) illustrated the opposite pattern, that Caucasian children were rated by their parents with higher skill levels (i.e., better social/adaptive skills) than African American children ( $p < .0001$ , contrast =  $0.82$ ).

Considering Hispanic students and the four items yielding DIF, a similar pattern as with African American students was observed. For the two items measuring Internalizing Problems, Hispanic children received higher ratings from their parents (i.e., more risk) than Caucasian children received (“is irritable,”  $p = .0001$ , contrast =  $-0.83$ ; “is easily frustrated,”  $p = .0007$ , contrast =  $-0.63$ ). The two items

measuring Adaptive Skills reported that Caucasian students received higher ratings than Hispanic students (“begins conversations appropriately,”  $p = .0002$ , contrast =  $0.63$ ; “gets along well with others,”  $p = .0008$ , contrast =  $0.50$ ).

### Discussion

This study investigated the BASC-3 BESS Preschool Parent Form using the Rasch RSM modeling framework to investigate item, scale, and subgroup characteristics. This methodology is often used in scale development and allows for in-depth examination of the feasibility of the rating scale categories, item severity, and coverage of the latent dimension (Bond & Fox, 2007; Kahler et al., 2005; Thomas, 2011). With screeners, RSM can be very useful to ensure that an instrument is targeting the intended population—in this study, young children at risk for behavioral and emotional difficulties.

Overall, the scale used for the BESS Parent-P is functioning acceptably with parent raters, and stable ratings were observed. Also, the items reported acceptable fit to the RSM, with severity estimates ranging from  $-1.09$  to  $1.92$ , providing relatively good coverage of the behaviors along the Maladaptive Behavior dimension. From the expected distribution, endorsement of an item at a given severity level indicates that those behaviors with lower severity levels are more likely to be observed by parents. As shown by the Wright map, BESS Parent-P items were spread along the latent dimension but were at the upper end of the latent distribution. This means that many of the items are targeted toward children with higher levels of risk (i.e., higher measure overall scores). Given that the goal of the screener is to provide early identification of behavioral and emotional risk, this pattern is appropriate.

Many of the items with lower severity scores were measuring Adaptive Skills (e.g., “adjusts easily to new surroundings”). Two items—one from the Externalizing Problems scale (“throws tantrums”) and one item from the Internalizing Problems scale (“is easily upset”)—were also below the item severity mean and, considering the content, represent relatively common behaviors for young children. Thus, this set of items (below the item severity mean) may tap relatively common behaviors for young children but contribute little toward identifying at-risk status.

Items that strongly contributed to measuring risk were largely due to Internalizing and Externalizing Problems. Content of this set of items included the majority of internalizing behaviors (e.g., “is negative about things,” “complains of Physical Behaviors”) and exhibition of externalizing behaviors (e.g., “hits others,” “acts out of Control”) toward peers and authority figures. As expected, these behaviors were not observed as frequently by parents, and relatively few children would be targeted by these items. In summary, it appears that the items on the BESS

Parent-P screener tap a range of behaviors observed by parents, but items above the item severity mean are not as common and thus have a greater contribution to measurement of behavioral and emotional risk.

Findings with the item content and spread of items across the latent distribution for the BASC-3 Parent-P matched findings for teachers conducted with the original BASC-2 versions of the BESS preschool and child/adolescent versions (DiStefano et al., 2014; DiStefano & Morgan, 2010). As with previous research, items measuring Adaptive Skills in the school environment did not contribute greatly to the identification of behavioral and emotional risk. Risk classification was due in larger part to the Externalizing and Internalizing Problem items. A new feature with the BASC-3 BESS is the creation of subscale scores for the three dimensions on the form. Future research may investigate the possibility of children who are noted as at risk in one area (e.g., lack of Adaptive Skills) but do not show elevation in other dimensions, leading to an overall score that does not denote risk.

Items measuring Internalizing Problems generally received high item severity scores, meaning that it was hard, even for parents, to frequently observe these behaviors with young children. Again, similar findings have been observed with investigations of teacher forms of the original BASC-2 BESS (DiStefano et al., 2014; DiStefano & Morgan, 2010). While internalizing behaviors in children have been noted as difficult for adult observers to detect (Achenbach & Rescorla, 2001; Gresham & Kern, 2004), another possibility is that internalizing problems may manifest themselves differently with young children (e.g., lack of pleasure, lack of enjoyment in activities) and may need to be revised to better identify children at risk for internalizing disorders. If the content is not appropriate on the screener, this decreases the likelihood that young children at risk for internalizing problems (e.g., depression, anxiety) are referred for timely help and support. The difficulty in endorsing this set of items may suggest the need to examine the construct and determine if internalizing problems with preschool children are manifested differently than with older children.

Finally, we report that the BESS Parent-P scale, when examined as a unique instrument, showed initial evidence of DIF by race/ethnicity but not by gender. Thus, the form may be considered as invariant across gender. The items are functioning as expected for girls as for boys, which supports the use of combined norms for scoring purposes. DIF was, however, observed between select ratings for Caucasian children and Hispanic or African American children. This finding is similar to a related study of the BASC-2 BESS Teacher Form–Child/Adolescent, which found minimal DIF exhibited for older children (aged 6–18 years) as measured by teachers; Hispanic students were more frequently rated by teachers as exhibiting at-risk behaviors

than were Caucasian students (Dowdy, Dever, DiStefano, & Chin, 2011). In the current study, findings were similar, as parents of Caucasian children were less likely to report problematic behaviors for the seven items exhibiting DIF and were more likely to report positive ratings of their children. The presence of DIF will also affect the scores provided by the BESS Parent-P, as African American and Hispanic children will have higher ratings for maladaptive behaviors on selected items and thus, higher scores. Besides recommending a closer examination of the items that exhibit DIF, it may be of interest to examine the use of culturally appropriate norms for scoring and/or interpretations to facilitate the use of the scores. For example, the items noted with DIF between Caucasian and African American children related to communication. Given that more children of color are traditionally referred for special education services by teachers, it is of interest to see why parents are rating these items as higher. This could be achieved through conducting parent interviews or focus groups to determine if the maladaptive behaviors hold the same interpretation with minority subgroups, or if cultural factors may result in different ratings by subgroup, and not necessarily relate to an expression of poor behavior.

## Limitations and Future Research

While we conducted RSM with the BESS Parent-P norming sample used to create the instrument, we recognize that there are limitations with the present study. As mentioned, there appears to be misfit of select items as well as redundancy of information provided by some of the existing items included on the screener. The values suggest that to fit the RSM, misfitting items could be removed to determine how scores to identify risk vary with, and without, the presence of these items. Similarly, the Wright map showed that many items were not providing unique information to the measurement of maladaptive behavior. In subsequent revisions of the screener, items may be reviewed along with alternative items from the norm pool to ensure that items selected for the screener measure unique content, represent behaviors that are observable in the classroom, and have content that is common across preschool teachers.

In addition, analyses were not overly supportive of the unidimensionality of the screener. Granted, some misfit was expected, given that the BASC-3 BESS Parent-P was developed from the longer form of a much longer, multidimensional instrument. Currently, the cohesiveness of the screener as a stand-alone instrument may be debatable; however, unidimensionality is critical if overall scores from the screener are to be used for decision making (G. T. Smith, McCarthy, & Anderson, 2000). Analyses provided initial evidence favoring the dimensionality of the BESS Parent-P scale. Residual plots (not presented here) suggested that the Internalizing and Externalizing Problems items were distinct from the other

items on the screener. Also, the bimodal distribution presented on the Wright map illustrated the split between items due to content, but this was also related to item wording direction. However, the positively worded items are useful to include, not only to focus a respondent's attention at the task at hand (Reynolds & Kamphaus, 2015) but also statistically. Previous research has shown that negatively worded items tend to contribute lower amounts of test information and have poorer discrimination as compared to positively worded items (Sliter & Zickar, 2014). Given that the construct under consideration (Maladaptive Behavior) is "negatively" focused, it would be useful to examine the impact of item direction on screener scores to determine if positively and negatively worded items contribute equally to scale information and have similar discrimination levels. Additionally, the Internalizing Problem item at the "top" of the Wright map (i.e., "Nobody likes me") was very hard for parents to observe and may not be a behavior that is likely to be demonstrated with frequency. Thus, this item may be a candidate for removal or editing in future revisions of the parent screener.

Finally, differences in item performance observed with race/ethnicity may, in part, be due to differences between cultural norms at home. For example, Hispanic parents rated their children as higher on the two Internalizing Problems items exhibiting DIF. Given the focus on the family that many Hispanic families share (Stein et al., 2014), these parents may be better attuned to fluctuations in their child's behavior. Future research may investigate the influence of family culture and norms to determine how differences across cultural contexts may correspond to ratings.

Another interesting avenue for future study involves the total score computed from the BESS Parent-P. The BESS Parent-P creates a sum score of item level raw data to create a child's risk classification level; however, there could be problems arising from the use of a sum score from a Likert-type scale. Summing responses assumes at least interval level of data; however, this assumption may be questionable when ordinal data are present (Bond & Fox, 2007; E. V. Smith et al., 2002). Furthermore, summed scores do not give additional weight to items that may vary the severity score associated with an item. Along these lines, an interesting avenue for future study will be to examine the relationship between external criteria, such as kindergarten readiness measures or test scores in later grades, and BESS Parent-P sum scores as well as latent scores derived by Rasch methodology. This analysis would help recognize the impact that different items may have on a child's total score and illustrate how these differences affect validity estimates. In addition, alternative cut points could be created using Rasch standard setting procedures (Stone, 2001), by age-group as well as by dimension, and classification results could be compared to existing methods. Additionally, the potential for measurement bias should be studied with more vigorous approaches. While the current study examines the

DIF across groups and provides information on mean differences and the potential for differences in predictive validity, more rigorous studies examining measurement equivalence, including studies of factorial invariance of the underlying structure, are needed.

In summary, this study investigated the BASC-3 BESS Parent-P screener using Rasch RSM to further investigate psychometric properties associated with the scale, the items, and measuring behavioral risk with young children. Implementing behavioral and emotional risk screenings using parent ratings may be one important avenue for reaching young children with behavioral problems early, before problems have adverse effects. Given the goals associated with universal screening, evaluation of assessment tools to assess need is warranted (Gudino, Lau, Yeh, McCabe, & Hough, 2009). High-quality screening measures of behavioral and emotional risk have promise for ensuring that young children have timely access to comprehensive prevention and intervention services.

### Acknowledgments

We acknowledge Pearson for granting use of the BASC-3 PRS-P norming data set.

### 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 in part by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A160157 to South Carolina Research Foundation. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

### References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA preschool forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015). *The state of preschool 2014: State preschool yearbook*. New Brunswick, NJ: National Institute of Early Education Research.
- Biglan, A. (2004, February). *A public health perspective on early childhood*. Keynote address presented at the fourth biennial Conference on Research Innovations in Early Intervention, San Diego, CA.

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Brown, W. H., & Conroy, M. A. (2011). Social-emotional competence in young children with developmental delays: Our reflection and vision for the future. *Journal of Early Intervention, 33*, 310-320. doi:10.1177/1053815111429969
- Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues* (2nd ed.). New York, NY: Guilford.
- Conroy, M. A., & Brown, W. H. (2004). Early identification, prevention, and early intervention with young children at-risk for emotional or behavioral disorders: Issues, trends, and a call for action. *Behavioral Disorders, 29*, 224-237.
- Conroy, M. A., Hendrickson, J. M., & Hester, P. P. (2004). Early identification and prevention of emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 199-215). New York, NY: Guilford.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart & Winston.
- DiStefano, C., Greer, F. W., Kamphaus, R. W., & Brown, W. H. (2014). Using Rasch rating scale methodology to examine a behavioral screener for preschoolers at risk. *Journal of Early Intervention, 36*, 192-211. doi:10.1177/1053815115573078
- DiStefano, C., & Morgan, G. (2010). Evaluation of the BESS TRS-CA using the Rasch rating scale model. *School Psychology Quarterly, 25*, 202-212. doi:10.1037/a0021509
- DiStefano, C., Morgan, G. B., & Motl, R. W. (2012). Detecting acquiescence underlying Rosenberg's Self Esteem Scale using the Rasch partial credit model. In S. De Wals & K. Meszaros (Eds.), *Handbook on psychology of self esteem*. Hauppauge, NY: Nova Science.
- Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. (2011). Screening for emotional and behavioral risk among students with limited English proficiency. *School Psychology Quarterly, 26*(1), 14-26. doi:10.1037/a0022072
- Forness, S. R., Freeman, S. F. N., Paparella, T., Kauffman, J. M., & Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders, 20*, 4-18. doi:10.1177/1063426611401624
- Gilliam, W. S., & Shahar, G. (2006). Preschool and child care expulsion and suspension: Rates and predictors in one state. *Infants & Young Children, 19*, 228-245. doi:10.1097/00001163-200607000-00007
- Gresham, F. M., & Kern, L. (2004). Internalizing behavior problems in children and adolescents. In R. B. Rutherford, Jr., M. M. Quinn & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 262-281). New York, NY: Guilford.
- Gudino, O. G., Lau, A. S., Yeh, M., McCabe, K. M., & Hough, R. L. (2009). Understanding racial/ethnic disparities in youth mental health services: Do disparities vary by problem type? *Journal of Emotional and Behavioral Disorders, 17*, 3-16. doi:10.1177.1063426608317710
- Kahler, C. W., Strong, D. R., & Read, J. P. (2005). Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: The brief young adult alcohol consequences questionnaire. *Alcoholism, 29*, 1180-1189. doi:10.1097/01.ALC.0000171940.95813.A5
- Kamphaus, R. W., & Reynolds, C. R. (2007). *Behavior Assessment System for Children—Second Edition (BASC-2): Behavioral and Emotional Screening System (BESS)*. Bloomington, MN: Pearson.
- Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior Assessment System for Children—Third Edition (BASC-3): Behavioral and Emotional Screening System (BESS)*. Bloomington, MN: Pearson.
- Linacre, J. M. (1992). *Many facet Rasch measurement*. Chicago, IL: MESA Press.
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. Rasch measurement: The dichotomous model. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 258-278). Maple Grove, MN: JAM Press.
- Linacre, J. M. (2006). *WINSTEPS: Rasch Measurement software*. Chicago, IL: Winsteps.
- Linacre, J. M. (2017). *WINSTEPS: Rasch Measurement software*. Chicago, IL: Winsteps.
- McConnell, S., Carta, J., & Greenwood, C. (2008, February). *Bringing response to intervention (RtI) to into preschool: Bridging research to practice*. Paper presented at the Conference on Research Innovations in Early Intervention, San Diego, CA.
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children—Third Edition*. Bloomington, MN: Pearson.
- Schumaker, R. E. (2004). Rasch measurement: The dichotomous model. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 226-257). Maple Grove, MN: JAM Press.
- Sick, J. (2010). Assumptions and requirements of Rasch measurement. *JALT Testing & Evaluation SIG Newsletter, 14*, 23-29.
- Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement, 74*, 214-226.
- Smith, E. V., Conrad, K. M., Chang, K., & Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement, 10*, 189-206. doi:10.1891/jnum.10.3.189.52562
- Smith, E. V., & Smith, R. M. (2004). *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111. doi:10.1037.1040-3590.12.1.102
- Smith, S. R. (2007). Making sense of multiple informants in child and adolescent psychopathology: Guide for clinicians. *Journal of Psychoeducational Assessment, 25*, 139-149. doi:10.1177/073428906296233
- Stein, G. L., Cupito, A. M., Mendez, J. L., Prandoni, J., Huq, N., & Westerberg, D. (2014). Familism through a developmental lens. *Journal of Latina/o Psychology, 2*, 224-250. doi:10.1037/lat0000025
- Stone, G. E. (2001). Understanding Rasch measurement: Objective standard setting (or truth in advertising). *Journal of Applied Measurement, 2*, 187-201.
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment, 18*, 291-307. doi:10.1177/1073191110374797

- Tyson, E. H. (2004). Ethnic difference in using behavior rating scales to assess the mental health of children: A conceptual and psychometric critique. *Child Psychiatry & Human Development*, 34, 167-201. doi:10.1023/B:CHUD.0000014996.26276.a5
- Ward, K. E., & Rothlisberg, B. A. (2011). Introduction to the special issue: Preschool assessment and intervention. *Psychology in the Schools*, 48, 427-429. doi:10.1002/pits.20564
- Webster-Stratton, C. (2005). The incredible years: A training series for the prevention and treatment of conduct disorders in young children. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (2nd ed., pp. 507-555). Washington, DC: American Psychological Association.
- Webster-Stratton, C., & Reid, M. J. (2008). Strengthening social and emotional competence in young children who are socioeconomically disadvantaged. In W. H. Brown, S. L. Odom & S. R. McConnell (Eds.), *Social competence of young children: Risk, disability, and evidence-based practices* (2nd ed., pp. 185-203). Baltimore, MD: Paul H. Brookes.
- Webster-Stratton, C., & Reid, M. J. (2013). Supporting social and emotional development in preschool children. In V. Buisse & E. S. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 265-281). Baltimore, MD: Paul H Brookes.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.