Standardization of Parental Assessments of Child Adjustment in Trinidad and Tobago Using Exploratory Multidimensional Item Response Theory

Jessica L. Chao
Paul A. McDermott
University of Pennsylvania, Philadelphia

Marley W. Watkins
Baylor University, Waco

Michael J. Rovine
University of Pennsylvania, Philadelphia

Anna Rhoad-Drogalis
The Ohio State University, Columbus

Frank C. Worrell
University of California, Berkeley

Tracey E. Hall
Center for Applied Special Technology, Wakefield

Abstract

Psychologists consider many sources of information when assessing the needs of children with behavioural difficulties, including child observation by teachers and parents. This study reports on the psychometric properties of the Adjustment Scales for Children and Adolescents-Home Edition (ASCA-H) in Trinidad and Tobago from an item response theory perspective. The ASCA-H is a contextually-based rating scale intended to collect information from parents on behaviours observable in the home that are relevant to identifying sociobehavioural problems. The sample was comprised of students (N = 731) attending government and assisted schools across Trinidad and Tobago. Exploratory full-information factor analysis yielded three robust and meaningful

dimensions: Aggressive, Reticent/Withdrawn, and Irascible/Attention-Seeking. Bayesian scores were used in HLM models to investigate variance explained in measures of academic achievement, learning behaviours, and teacher ratings of behavioural maladjustment. Age, gender, and ethnic differences were also explored. The current article demonstrates the application of multidimensional IRT factor analysis for revealing psychological dimensions of child adjustment and scaling and scoring those dimensions for research and practice in Trinidad and Tobago.

Key words: *Item Response Theory, Trinidad and Tobago, Sociobehavioural*

Received December 14, 2017
Accepted November 3, 2018

This research was supported in part by the Trinidad and Tobago Ministry of Education, the Organization for American States (Fund # TT/AE/ 138101941). Correspondence concerning this article should be addressed to Jessica L. Chao, Graduate School of Education, Human Development and Quantitative Methods Division, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104-6216; email: jec@gse.upenn.edu.

Standardization of Parental Assessments of Child Adjustment in Trinidad and Tobago Using Exploratory Multidimensional Item Response Theory

Behaviour and social skills have long been linked to school success and psychosocial and health outcomes (DiPerna & Elliott, 2002; Malecki & Elliott, 2002; Wentzel, 2009). Identifying and assessing the needs of children with behavioural difficulties is thus a primary concern. With early identification, practitioners can implement interventions and help children avoid later adverse outcomes. Behavioural assessment tools include interviews, rating scales, and child observation. Among these techniques, measures often differ in regard to key elements, including the informant (self, parent, or teacher), intended population for use, and scale composition and format. Standardized behavioural rating scales offer several advantages: they are an objective assessment,

relatively unobtrusive, practical to utilize across different raters and settings, and provide normative information even on low-frequency but emblematic behaviours (Merrell, 2003). The current article is intended to demonstrate an appropriate methodological strategy for the development of a standardized behavioural rating scale informed by parents, for use in Trinidad and Tobago, comprised of multiple dimensions (where the scale may measure several potentially related traits simultaneously).

Source of the Informant: Parents

Although parent ratings are often considered less reliable than teacher ratings (Dupaul et al., 1998), parents can provide valuable additional information considering their unique knowledge of their child in social, community, and learning contexts outside of the classroom (Friedman, Leone, & Friedman, 1999). This is important considering the contextually-selective nature of many social and behavioural problems (Achenbach, McConaughy, & Howell, 1987; Costenbader & Keller, 1990). Teachers are more likely to offer a normative perspective (Konold, Brewster, & Pianta, 2004) but parents provide judgments more specific to the child. As teachers and parents perceive a child's behaviour within different environments, it is important to gain both perspectives for a comprehensive understanding of the child's socioemotional and behavioural functioning.

Commonly used behavioural rating scales for parents include the Revised Conners Parent Rating Scale (CPRS-R; Conners, Sitarenios, Parker, & Epstein, 1998) and the Child Behaviour Checklist for Ages 6-18 (CBCL/6-18; Achenbach & Rescorla, 2001). Both present a list of behaviours for which parents rate their child based on frequency and intensity of manifestation of those behaviours. Although scores on these instruments have demonstrated adequate reliability and validity (Conners et al.,, 1998; Sattler, 1992), they do not inform about the circumstances within which the problematic behaviours occur, nor whether the behaviours emerge across multiple situations.

The Adjustment Scales for Children and Adolescents-Home Edition (ASCA-H), the focus of this article, is a contextually-based behavioural rating scale designed to collect information from parents. It was inspired by the Adjustment Scales for Children and Adolescents (ASCA; McDermott et al., 2015) teacher rating scale, with items included from the original scale if they were relevant to home situations, and new items added for areas where parents

would be particularly knowledgeable. It was developed to complement teachers' classroom observations but is constructed especially to assess children in 28 different *home* situations (relationships with parents, other adults and children; coping with responsibilities; daily living; spare-time activities; schoolwork; discipline). Clinical jargon is avoided. Both the ASCA and ASCA-H were administered in Trinidad and Tobago as part of an initiative to identify children at risk for academic and behaviour problems.

Intended Population: Cultural Context and Use of Behaviour Rating Scales

While the ASCA teacher rating scale was originally developed, standardized, and validated in the U.S. (and later standardized for use in Trinidad and Tobago: see McDermott et al., 2015), the ASCA-H parent rating scale was designed by a group of Trinidadian and American researchers specifically for Trinidad and Tobago. Good practice in scale development compels assessment of the psychometric properties of the scale unique to any population for which use is intended, using a valid representative sample. Recognizing the impact of cultural context is perhaps especially important in the development and assessment of behavioural rating scales, where differences in cultural expectations, local aspects of behavioural styles, and distinct philosophies toward child rearing may cause variation (United Nations Children's Fund, 2013).

Prevalence estimates of diagnosed psychopathology vary widely across societies, attributable at least in part to variations in methodologic elements such as use of dimensional scales outside their societies of origin (Achenbach, Rescorla, & Ivanova, 2012). Differences in cultural expectations and variant levels of concern over certain patterns of behaviour can easily confound measurements functioning across multiple societies when using the same metric. Due to variations in linguistic nuances and conceptual interpretations of constructs, identical items may offer different amounts of psychometric information depending on the target population.

Additionally, although Trinidad and Tobago is considered one of the most developed nations in the Caribbean, the International Monetary Fund still classifies it as a developing economy. Few parent rating scales have been nationally adapted and validated for use in developing nations (Mpofu, Oakland, Ntinda, Seeco, & Maree, 2014). Much of the epidemiological research on childhood mental

health has focused on industrialized countries, but low and middle income countries face different challenges that may impact their conceptualization of child psychopathology and the way it manifests (Atilola, 2015). Trinidad and Tobago faces difficult economic circumstances and high crime rates (Worrell, 2006a), and has distinctive parental disciplinary practices influenced by a blend of British, East Indian, French, and African traditions (Waithe & Worrell, 2003; Worrell 2006b). Families of Caribbean backgrounds emphasize respect for elders, demonstrating intolerance for rude and aggressive behaviour in their children (Ho, Bluestein, & Jenkins, 2008). Moreover, research in Jamaica, another former British colony and therefore perhaps comparable to Trinidad, suggests that some Jamaican parents may consider problems of fearfulness and fighting less unusual and more redeemable than American parents, so that Jamaican informants tend to comparably underreport their children's activities in those areas (Lambert, Weisz, & Knight, 1989; Lambert et al.,1992). Additionally, there is some evidence that there are different syndrome base rates in Jamaica versus the U.S., with Jamaican parents reporting a higher base rate for internalizing problems (Lambert, Lyubansky, & Achenbach, 1998).

Scale Format and Composition: Multidimensionality

When referring to scale format and composition, most often the discussion centers on item format (open-ended, dichotomous, etc.), length, modality (e.g., online, paper and pencil), and content. In scale development and assessment, it is important to establish the structural properties of the scale, called *dimensionality*. Unidimensional scales measure one trait, with the observed set of variables purporting to represent one latent phenomenon (e.g., oppositionality). However, many behavioural rating scales are designed to measure phenomena with multiple dimensions (oppositionality, withdrawal, moodiness, etc.). This violates the assumption of unidimensionality required by traditional factor analysis, the method most commonly used to establish scale dimensionality. Additionally, whereas factor analysis is ideally intended for use on continuous data following the normal distribution, problem behaviour items tend to have non-normal distributions and relatively few response categories, leading to potentially biased findings. Factor analysis uses the covariance matrix of item responses and reduces them to linear combinations of items representing latent variables (factors), often broad

concepts that reflect an observed trait (e.g., reticence). Item response theory (IRT), a newer paradigm, jointly scales items and persons so that trait level is estimated using both an individual's item responses and the item characteristics (as estimated using the responses of all persons) to estimate trait level. This results in population-invariant item calibration, item-invariant meaning for traits and item-referenced meaning for trait-levels, and standard errors of measurement that reflect item appropriateness at trait levels (Embretson & McCollam, 2000).

Full information item factor analysis (FIFA) stands at the intersection between factor analysis and IRT methods, with an emphasis on using item parameters to define factors (as in traditional factor analysis) rather than as characterizations of the interaction between persons and items (as in item response theory). The FIFA approach has the advantage of estimating parameters from the full information contained in item-response patterns, drawing on both the relationships between items represented by the correlation matrix and each respondent's actual pattern of responses. As the approach is not solely dependent upon the correlation matrix, this method also avoids false factors (called difficulty factors; Bernstein & Teng, 1989) that emerge when items reflect behaviour that is observed as either presence or absence, and are thus dichotomous in nature. Exploratory IRT models with multiple dimensions are thus considered appropriate for examining the structure of binary or non-continuous data (Embretson & McCollam, 2000). Additionally, these parameters can be used to scale children on each latent trait, weighing both behaviour prevalence and assessment precision (discrimination power), and using more accurate Bayesian scoring rather than the traditional factor scores (weighted sums, regression, etc.).

The Current Study

The current article focuses on Trinidad and Tobago, where discipline is a notable form of parent-child interaction, and obedience and respect of adults is highly valued (Barrow, 2008; Evans & Davies, 1997). The purpose of the research was to evaluate the psychometric properties of ASCA-H scores from an IRT perspective, demonstrating the application of exploratory multidimensional IRT factoring, calibration, Bayesian scoring, and HLM validity analyses for contextually-based home child observations. It builds on a preliminary study conducted by Pearson (2007) by taking advantage of modern methodological

practices to more fully account for the multidimensional nature of the data, as well as IRT scoring methods to enhance reliability and generalizability, and multilevel modeling with other scales to provide estimates of construct validity.

This is inherently an exploratory study, so affirmative research hypotheses would not be appropriate. Instead, two overall research questions are addressed. First, what are the psychometric properties of ASCA-H for elementary school children in Trinidad and Tobago? Previous research in Trinidad and Tobago with teacher rating scales (McDermott et al., 2015), as well as with other international samples, have generally revealed a very robust internalizing/ externalizing paradigm in child behaviour scales (Canivez & Beran, 2009; Canivez & Bohan, 2006; Canivez & Sprouls, 2010), so we would expect to see two or more dimensions (factors) reflecting that pattern. The internalizing dimension is generally characterized by shy or timid behaviour, while the externalizing dimension consists of acting out and uninhibited behaviours. Second, do the resulting scores demonstrate adequate construct validity and are there differences related to age, ethnicity, or gender? Prior research on the teacher-rating ASCA with Trinidadian students found that students of East Indian descent and males had higher scores on aggressive syndromes and younger students displayed more avoidant and oppositional defiant problems (Grim, 2002), so we might expect to see a similar trend.

Method

Participants

The sample included students aged 4 to 15 years (M = 8.0, SD = 2.0) enrolled in grades Infant 1 (approximately age 5) to Standard 5 at 74 government and assisted elementary schools in Trinidad and Tobago. Data on various measures of behavioural adjustment and academic achievement were collected on a stratified random sample of students through the efforts of Trinidad and Tobago's Ministry of Education and a contracted team of Trinidadian and American researchers based at Pennsylvania State University (Watkins, Hall, & Worrell, 2014). Students were stratified by school region, grade, and gender.

ASCA-H was completed for 731 students. Parents completed 81.2% of ASCA-H forms, with other respondents including guardians, grandparents, elder siblings, or other relatives, with 8%

of respondents not reporting their relationship with the student. The sample of students was 50.5% female, and comparable to national ethnic distributions in Trinidad and Tobago (i.e., 34.2% of African descent, 35.4% East Indian descent, and 24.3% mixed descent; Central Intelligence Agency, 2014), with 38.2% of African, 39.7% of East Indian, and 22.1% of mixed descent.

Instruments

Home social-emotional behaviour. ASCA-H is a parent-rating scale that includes 203 behavioural indicators in 28 situational contexts. It contains descriptions of observable behaviours in a dichotomous item response format, where a parent endorses whether or not their child has exhibited a behaviour with reference to specific home situational contexts over a two-month period. Parents may endorse multiple items within each context. For example, within the context of coping with homework, the parent may describe typical behaviour as: "Listless, too unconcerned to do it", "Generally does homework on her own", "Asks for help even when it is not needed", "Wastes time during homework", "Destroys or defaces her school materials", "Often loses or forgets her books", and "Only works when watched". Each situational context presents at least one positive or healthy behavioural indicator (e.g., "Generally does homework on her own" in the previous example) to mitigate the possibility of response sets associated with offering exclusively negative or problematic behaviours (see LeBoeuf, Fantuzzo, & Lopez, 2010). Thus the scale provides information on problem behaviours both from a situational perspective and from a phenological perspective. As the ASCA-H is the focal instrument of this study, the psychometric properties will be reported in the results.

Classroom social-emotional adjustment. ASCA is a teacher rating scale containing 156 behavioural indicators in relation to 29 classroom situational contexts (McDermott, Stott, & Marston, 1993). Like the ASCA-H, it allows investigation of the situational context of behaviours as well as phenological types of behaviours, and has a similar response format. The teacher endorses behavioural indicators relevant to each context which reflect the student's observed behaviour over a two-month period. Three reliable contextual dimensions ($\alpha \geq .75$; peer, learning, and teacher context problems) have been observed in both the U.S. and Trinidad and Tobago normative samples (McDermott, Steinberg, & Angelo, 2005; McDermott et al., 2016). ASCA also provides

information on two broad-band phenological syndromes: Overactivity (i.e., externalizing problems) and Underactivity (i.e., internalizing problems). These two dimensions have been replicated in the U.S. and in populations such as Hispanic/Latino, Native American, and Canadian children (Canivez & Beran, 2009; Canivez & Bohan, 2006; Canivez & Sprouls, 2005, 2010; McDermott, 1993), as well as recently in Trinidad and Tobago ($\alpha >$ .70) (McDermott et al., 2015). There is substantial evidence of internal consistency, convergent and discriminant validity, and factorial validity of the ASCA phenological dimensions across age, gender, and ethnicity (McDermott et al., 2005).

Classroom learning behaviour. The Learning Behaviours Scale (LBS; McDermott, 1999) is a teacher rating scale comprised of 29 items intended to assess approaches to learning tasks. The teacher observes the student for at least 50 days and rates behaviours on a three-point Likert scale (*most often [2], sometimes[1],* or *does not apply[0]*). Item examples include "Responds in a manner that shows attention", "Has enterprising ideas which don't often work out", and "Gets aggressive when frustrated or when work is corrected". For U.S. application, the measure yields a total score as well as four subscores assessing distinct dimensions of learning behaviours, including Competence Motivation, Attitude toward Learning, Attention-Persistence, and Strategy/ Flexibility. Convergent and divergent validity evidence has been established with classroom behaviorbehaviour using the ASCA (McDermott et al., 1993), academic performance using the Basic Achievement Skills Individual Screener (The Psychological Corporation, 1983), and intellectual functioning using the Differential Ability Scales (Elliot, 1990). Support for the factor structure of LBS scores as well as factorial invariance has been reported in multiple contexts including among Canadian youths (Canivez & Beran, 2011; Canivez, Willenborg, & Kearney, 2006; Worrell, Vandiver, & Watkins, 2001). Recent work in Trinidad and Tobago indicates that the measure yields two reliable dimensions for that population ($\alpha >$ .80): Competence Motivation and Strategy/Flexibility (Chao et al., 2018).

Classroom clinical behaviour. The Disruptive Behaviour Disorders Rating Scale (DBDRS; Pelham, Gnagy, Greenslade, & Milich, 1992) is a teacher rating scale that includes items reflecting criteria from the three disruptive behaviour categories (Attention Deficit Hyperactivity Disorder, Oppositional-Defiant Disorder, Conduct Disorder) described in the Diagnostic and Statistical Manual of Mental Disorders, revised Version III (DSM-III-R;

American Psychiatric Association, 1987). The lists of symptoms for these categories in the DSM-III-R correspond well to those of more recent versions, though some diagnostic criteria have changed (Pelham, Fabiano, & Masseti, 2005). Teachers record the frequency of each symptom on a four-point Likert scale (*Not at all[0], Just a little[1], Pretty much[2], and Very much[3]*). DBDRS yields three reliable scores (α = .91-.96; Inattention, Oppositional/Defiant, and Impulsivity/Overactivity) and has proven an adequate measure for detecting behavioural and pharmacological effects (Pelham et al., 2005). Sufficient predictive and discriminant validity have been established with various populations, including clinical cohorts, males in regular classrooms (Pelham, Gnagy et al., 1992), and special education settings (Pelham, Evans, Gnagy, & Greenslade, 1992).

Academic achievement. Oral reading fluency (ORF) is a measure of the speed and accuracy with which a student reads text and has been normed nationally in the U.S. (Hasbrouck & Tindal, 2006). It is used as a screening and monitoring measure of student reading proficiency, based on the principle that students cannot read faster than they can comprehend (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001). There is evidence supporting both convergent validity of ORF scores with other curriculum-based measures and state reading assessments (Deno, Fuchs, Marston, & Shin, 2001; Stage & Jacobsen, 2001; Wood, 2006) and predictive validity and clinical utility (Hart et al., 2013; Petscher & Kim, 2011). Scores represent the average number of words read correctly from two passages. ORF passages were taken from local grade-level reading texts and administered individually three times over the year in fall (M = 60.9, SE = 38.7), winter (M = 67.5, SE = 39.3), and spring (M = 60.9, SE = 38.7). The average correlation between the two passages was .85.

Analysis

Prior to analysis, several ASCA-H items were deleted due to low prevalence or irrelevance. These included the positive behaviour items, which exist in the scale to reduce response bias associated with presenting exclusively negative or problematic behaviours (see LeBoeuf, Fantuzzo, & Lopez, 2010), and two sets of items under the respective categories of "Troublesome and Illicit Activities," and "Other Behaviours that Cause Concern", which were not presented in specific home contexts but rather were lists of behaviours appended to the end of the scale for clinical information

gathering and not intended to be part of the main instrument. The remaining 131 items were submitted to the dimensional analysis.

Full information factor analysis. Dimensional analysis generally involves first extracting eigenvalues from a correlation matrix, where large eigenvalues indicate association with substantive amounts of variance. These help determine where the proportion of variance attributable to a possible dimension becomes trivial enough that it would preclude interpretation. In unidimensional data we would expect to see one large eigenvalue absorbing much of the variance. The first few eigenvalues extracted from the correlation matrix suggested the multidimensional nature of this data, with the ratio between the first and second being 3.36, below the usual crude criterion value of 5 for unidimensional solutions. Full information item factor analysis (FIFA) was applied in an exploratory manner, where number of dimensions was not assumed known beforehand.

FIFA is an IRT-based marginal maximum likelihood approach that allows estimation of multidimensional IRT model parameters from a smoothed tetrachoric correlation matrix. Tetrachoric correlations are used with dichotomous items because Pearson correlations (normally intended for continuous items) with dichotomous items will produce spurious factors. Smoothing of the matrix is needed because of the problems inherent in applying factor analysis to a large number of items, where the process often results in a non-positive definite matrix (Cullen, 1990). As a positive definite matrix is needed to conduct factor analysis, where the matrix is free of redundant variance (no items whose variance can be predicted by a combination of other items), smoothing adjusts the matrix so the analysis can be performed. FIFA assumes an underlying multiple-factor model where an individual's response to an item is a function of a latent response process. It uses all of the information in the observed case item response patterns to estimate parameters, rather than the partial information from pairwise correlations used in factor analysis or principal components analysis. Since it does not use the correlation matrix to estimate the factor loadings, it avoids the difficulties associated with non-continuous data when zeroes appear in the pairwise joint occurrence frequencies. As the model is fitted by maximizing the marginal likelihood of the thresholds and loadings, the appropriate number of factors can be determined by calculating the increase in likelihood attributable to an additional factor and conducting a likelihood ratio test for significance, or

examining the differences in chi-square fit statistics. The TESTFACT software program (Wood et al., 2003) was specially designed to facilitate this method, where the program will automatically calculate the necessary specialized smoothed tetrachoric matrices and response patterns to be used in the analysis from the raw data. It uses estimates from an initial factor analysis of the tetrachoric matrix as the starting values for full information factor analysis and outputs parameters in both factor analytic and IRT metric. Additionally, it provides Bayes estimates of factor scores. An implementation of the FIFA method is also available using the R package 'mirt' (Chalmers, 2012).

Minimum average partialling (MAP; Velicer, 1976) was used with a smoothed tetrachoric matrix as well as examination of a scree plot to suggest the number of retained dimensions. These are typical preliminary methods used to reveal the general range of the probable number of extractable factors. The G^2 statistics were submitted to an analysis of deviance test for nested models (Bock & Aitkin, 1981) and residual correlation matrices examined for fit. Insignificant G^2 would indicate that any additional factors extracted might be due to sampling variation and uninterpretable. Solution criteria also included item coverage; at least 4 salient (loadings ≥ .40) items per factor; reliable factors (i.e., $\alpha \geq .70$); and theoretical plausibility, parsimony, and concordance with leading research (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Scores were computed through expected a posteriori (EAP) Bayesian estimation using the parameters from the selected multidimensional item response model (Muraki & Engelhard, 1985). The reliability of scores on the ASCA-H dimensions was assessed using both Cronbach's α measure of internal consistency and McDonald's ω.

Construct validity. Bivariate product–moment correlations were computed to determine the direction and magnitude of relationships between scores on each ASCA-H subscale and external criterion variables related to teacher ratings of student behaviour, reading fluency, and student learning behaviours. As data were nested within teachers, HLM was applied, where each ASCA-H subscale served as a group-mean centered predictor in a two-level conditional HLM model, indicating the percentage of between-children within-teacher variance accounted for by respective ASCA-H subscales.

Results

Dimensionality

The scree plot and MAP analysis supported retaining two or three dimensions. Full information analyses were performed for both solutions and compared. The three-factor model fit significantly better than the two-factor model, where P ($\chi^2 \geq 721.0$, 129) $\approx$ 0. Promax (oblique) and varimax (orthogonal) rotated factor loadings were analyzed; as the promax interfactor correlations were between .21 and .43, the promax solution was deemed more suitable.

The three dimensions were named 'Aggressive', 'Reticent/Withdrawn', and 'Irascible/Attention-Seeking'. Names are given to the dimensions on the basis of the items associated with each dimension, as listed in Table 1, such that the items are listed in descending order of strongest relationship with the dimension. In general, 'Aggressive' contains a preponderance of items reflecting confrontational and conduct problem behaviours, 'Reticent/Withdrawn' contains timid, apathetic behaviours, and 'Irascible/Attention-Seeking' contains many items related to acting out for attention. Table 1 presents rotated pattern loadings, final communalities, item-total scale correlations, and prevalence for the Aggressive (35 items; *M* behavioural prevalence = 6.1%), Reticent/Withdrawn (25 items; *M* prevalence = 11.2%) and Irascible/Seeking Attention scales (30 items; *M* prevalence = 19.5%). Empirical reliabilities and αs for internal consistency were all greater than .70. McDonald's ω was just under .70 for Aggressive, but greater than .70 for the other scales.

Construct Validity

EAP scores were used in HLM to investigate construct validity. Table 2 displays the concurrent relationships between ASCA-H scores and independent criterion measures. As the data are nested (children within classrooms), HLM was used to enable more precise estimates of relative criterion-related validity for scores, since ordinary Pearson correlations would partially reflect classroom differences as well as individual differences between children. Thus the last column in Table 2 specifically speaks to variance attributable to children's actual individual differences, with parenthetical entries reflecting how much of that variance is accounted for by a given ASCA-H scale. Significant correlations were in the expected direction between scores on each ASCA-H

dimension and external criterion variables. ASCA-H dimensions showed low to moderate correlations with DBDRS dimensions, with the exception of the near-zero and nonsignificant Reticent/Withdrawn dimension, indicating that ASCA-H has concurrent validity with this measure detecting clinical disturbance in classrooms. Parenthetical values indicating attributable variance for the ASCA-H Aggressive scale ranged 16.9 to 25.1%, suggesting fair correspondence. ORF and LBS correlations were all low but in the expected negative direction, though only Aggressive values tended to be significant. Additionally, Table 2's last column entries for the ORF scales indicate that only 56.0-58.6% of score variance stems from children's individual differences (rather than teacher or classroom characteristics), and parenthetical values of 0.0-1.5 suggest an even smaller amount of association than the -.19 to -.21 correlations.

Aggressive also correlated more and accounted for a higher percentage of variance with peer and academic context problems than with teacher context problems, indicating that parents and teachers may be observing different behaviours. Aggressive had a low but positive correlation with the teacher rating of Overactivity.

Demographic Trends

Table 3 displays the mean population distribution of Aggressive, Reticent/Withdrawn, and Irascible/Attention-Seeking by student gender and grade level in Trinidad and Tobago, whereas Table 4 shows the distribution by student gender and ethnicity. MANOVA with grade level, gender, and ethnicity as independent variables and the three ASCA-H dimensions as dependent variables was used to indicate whether there were mean differences in dimensional scores across demographic groups. Subsequent univariate ANOVAs and Tukey-Kramer contrasts suggest the means of ASCA-H Aggressive scores vary significantly at $p < .05$ where scores are higher for African children compared to East Indian children and for male children compared to female children. ASCA-H Reticent/Withdrawn scores vary significantly at $p < .05$ where scores are higher for children in Infant 2 than in Standards 3, 4, and 5, for children in Infant 1 than in Standard 5 and for African and Mixed descent children compared to East Indian children. ASCA-H Irascible/Attention-Seeking scores vary significantly at $p < .05$ where scores are higher for East Indian children compared to African children and female compared to male children.

Table 1. *Dimensional Structure and Properties of the Adjustment Scales for Children and Adolescents Home Edition*

Item description[a]	**I**	**II**	**III**	**Comm-unality**	**Item/ scale *r***[c]	**%Pre-valence**[d]
Factor I: Aggression (coefficient α = .79, empirical reliability = .81, ω = .69 (95%CI .62-.75))[e]						
Deliberately destroys others' belongings	*.70*	-.23	.22	.59	.34	.01
Starts fights and rough play	*.68*	.00	.14	.49	.50	.03
Listless, unconcerned about homework	*.66*	-.12	-.07	.45	.22	.04
Lies about assignment	*.65*	-.09	.05	.43	.36	.04
Disobedient, difficult to control	*.65*	-.01	.34	.54	.55	.04
Distant, no effort to relate	*.63*	.08	-.09	.41	.27	.02
Overly rough with smaller children	*.63*	.20	-.03	.43	.42	.03
Answers back aggressively, makes threats	*.62*	-.04	-.04	.39	.35	.02
Disturbs others' fun	*.61*	-.18	.29	.48	.33	.02
Disrupts by fooling around	*.60*	-.11	.27	.44	.36	.04
Uses bad language with other adults	*.58*	.11	.00	.35	.23	.01
Fights physically with others	*.58*	.03	.23	.39	.43	.03
Is rude with other adults	*.58*	-.24	.05	.40	.27	.02
Associates with troublesome children	*.55*	.07	.05	.31	.27	.03
Takes things without permission	*.55*	-.23	.35	.47	.37	.07
Snatches objects away	*.48*	-.06	.32	.34	.30	.03

Table 1 continued

Item description[a]	I	II	III	Comm-unality	Item/ scale r^c	%Pre-valence[d]
Speaks in rude/angry tone	*.47*	.10	.25	.29	.35	.06
Very slow, never finishes	*.47*	.04	.01	.22	.20	.08
Destroys belongings	*.47*	-.07	.24	.28	.25	.05
Destroys school materials	*.46*	-.10	*.41*	.39	.22	.02
Quarrels, provokes others	*.45*	.09	.29	.30	.36	.04
Uses dirty words/off language	*.45*	.30	-.08	.30	.22	.01
Refuses to care for hygiene	*.44*	-.01	.26	.26	.27	.05
Has ruined work on purpose	*.43*	-.12	.18	.24	.23	.01
Fools around when works with hands	*.42*	-.15	*.41*	.37	.31	.06
Stays in bed to avoid responsibilities	*.41*	.28	.08	.25	.28	.04
Unkind to weaker children	*.41*	.22	.07	.22	.26	.01
Often loses belongings	*.41*	-.19	.26	.27	.20	.26
Too disinterested to play	*.41*	.28	-.34	.35	.04	.01
Lacks energy to care for self	*.41*	.18	.04	.20	.26	.04
Often loses or forgets materials/assignment	*.40*	-.04	.27	.24	.28	.06
Unconcerned with attention	*.40*	.10	-.21	.21	.20	.07
Refuses to help	*.40*	.16	.13	.20	.24	.05
Overly fussy about things	*-.42*	*.50*	*.44*	.61	.25	.27
Very concerned about mistakes	*-.55*	.34	.36	.54	-.10	.46
Factor II: Reticence/Withdrawal (coefficient α = .72, empirical reliability = .72, ω = .72 (95%CI .69-.76))[e]						

Table 1 continued

Item description[a]	**I**	**II**	**III**	**Comm-unality**	**Item/ scale *r*[c]**	**%Pre-valence**[d]
Too shy to interact with other adults	-.07	*.66*	-.01	.44	.38	.14
Never any trouble because so timid	-.20	*.64*	-.27	.53	.26	.06
Too shy to greet other adults	-.05	*.61*	-.08	.39	.37	.19
Seems fearful of other adults	.00	*.60*	.06	.36	.30	.05
Feelings easily hurt	-.36	*.58*	.23	.52	.34	.39
Won't get involved	-.07	*.55*	-.13	.32	.14	.02
Too timid to join play	-.24	*.53*	-.39	.48	.11	.04
Overly dependent	-.19	*.51*	.20	.33	.32	.19
Needs encourage to join	-.10	*.50*	-.15	.29	.24	.18
Overly fussy about things	*-.42*	*.50*	*.44*	.61	.25	.27
So timid difficult to get to speak	.28	*.49*	-.17	.34	.26	.02
Will let others push ahead	-.13	*.48*	.06	.25	.26	.09
Tends to go off and play alone	-.02	*.48*	.10	.24	.26	.15
Rarely smiles	.01	*.48*	.07	.23	.23	.03
Distant seldom says anything	.31	*.47*	-.20	.36	.27	.04
Not shy but rarely offers answer	-.12	*.47*	.00	.23	.26	.16
Shy but not unfriendly	-.01	*.46*	-.06	.21	.28	.24
Clings to you or shows tears	-.22	*.44*	.14	.26	.23	.14
Never seeks help even if needed	.19	*.44*	-.04	.23	.28	.05
Responds with angry look	.20	*.44*	.11	.24	.17	.02
Allows self to be bullied	-.10	*.43*	.10	.21	.24	.10

Table 1 continued

Item description[a]	**I**	**II**	**III**	**Comm-unality**	**Item/ scale *r*[c]**	**%Pre-valence**[d]
Too uninterested to notice belongings	-.05	*.42*	.26	.25	.23	.06
Has untalkative moods	.09	*.41*	-.07	.18	.25	.11
Has dejected look	.23	*.41*	.19	.25	.22	.02
Sometimes seeks disapproval	.03	*.40*	.13	.18	.23	.05
Factor III: Irascible/Attention-seeking (coefficient α = .83, empirical reliability = .72, ω = .82 (95%CI .80-.84))[e]						
Has trouble waiting for turn	.00	-.06	*.67*	.44	.44	.14
Uses various ways to get others' attention	.04	-.17	*.66*	.40	.42	.17
Inclined to cheat	.23	-.09	*.61*	.38	.32	.03
Tries to dominate	-.12	-.04	*.57*	.32	.32	.10
Much too talkative	-.07	.13	*.55*	.31	.39	.32
Throws tantrums at bedtime	.01	.06	*.54*	.29	.29	.05
Argues and complains about wait	.12	.05	*.54*	.30	.39	.13
Greets loudly	-.28	.06	*.53*	.36	.30	.32
Seeks help when not needed	-.09	.19	*.53*	.30	.32	.10
Becomes restless and fidgety in line	.03	-.07	*.51*	.25	.37	.35
Wants to dominate play	.18	-.06	*.49*	.27	.38	.12
Constantly restless at meals	.20	-.16	*.49*	.30	.33	.10
Welcomes others loudly	-.17	-.12	*.49*	.26	.23	.23
Upset if does not perform well	-.16	.30	*.47*	.33	.31	.26
Uses various ways to get attention	-.03	.11	*.46*	.22	.34	.40

Table 1 continued

Item description[a]	I	II	III	Comm-unality	Item/ scale *r*[c]	%Pre-valence[d]
Poor loser	.21	-.01	*.46*	.25	.36	.10
Moody and uncooperative	.21	.27	*.45*	.32	.31	.03
Attacks physically if provoked	.32	.04	*.45*	.29	.37	.09
Misbehaves when you attend other things	.30	-.20	*.44*	.32	.36	.34
Overly fussy about things	-*.42*	*.50*	*.44*	.61	.25	.27
Clowns around, plays silly tricks	.09	.11	*.44*	.21	.37	.22
Loud but not disruptive at play	.04	-.23	*.44*	.24	.27	.40
Constantly distracted not ready	.21	-.11	*.43*	.23	.30	.16
Argues and talks back	.14	-.02	*.42*	.19	.36	.29
Sometimes lies to avoid blame	.35	-.03	*.41*	.29	.40	.47
Sometimes unfriendly mood	.17	.21	*.41*	.24	.34	.12
Destroys school materials	*.46*	-.10	*.41*	.39	.22	.02
Fools around when works with hands	*.42*	-.15	*.41*	.36	.31	.06
Does things knows are wrong	.27	.10	*.40*	.24	.35	.18
Improves but does not last	*.39*	-.12	*.40*	.17	.39	.30

Note: [a]Item descriptions are abbreviated for convenient presentation.

[b]Values are promaxian pattern loadings. Salient pattern loadings (≥ .40) are italicized. *N* = 731.

[c]Each correlation reflects the relationship between an item and the sum of the other items comprising a scale, where distributions were standardized to unit-normal form.

[d]Entries indicate the percentage of children for whom the item behaviour is scored present.

[e]Reliability is based on the sample *N* = 731.

Table 2. *Relationships Between ASCA-H Scores and Concurrent Criterion Measures*

Criterion measure	Aggressive	Reticent/ Withdrawn	Irascible/Attention-Seeking	% Explainable variance[a]
Adjustment Scales for Children and Adolescents, Trinidad and Tobago scale (teacher rating)				
Overactivity (*n* = 689)	.29 (7.7)	10 (5.5)[†]	-.14 (11.5)	88.2
Underactivity (*n* = 689)	.15 (9.2)	.14 (-0.1)	.05[†] (2.5)[†]	91.1
Adjustment Scales for Children and Adolescents, Trinidad and Tobago context scales (teacher rating)				
Peer Context Problems (*n* = 689)	.31 (11.2)	.08 (0.0)[†]	-.08 (8.6)*	74.5
Teacher Context Problems (*n* = 689)	.19 (2.8)	-.08 (0.5)*	.01[†] (14.9)[†]	91.7
Academic Context Problems (*n* = 689)	.37 (12.8)	.03 (0.0)	-.08 (8.2)*	86.0
Disruptive Behaviour Disorder Rating Scale (teacher rating)				
Inattention (*n* = 575)	.36 (25.1)	.01[†] (0.0)[†]	-.12 (0.8)	83.3

Table 2 continued

Criterion measure	**Aggressive**	**Reticent/ Withdrawn**	**Irascible/Attention-Seeking**	**% Explainable variance[a]**
Oppositional/Defiant (*n* = 534)	.27 (19.0)	.05[†] (0.4)[†]	-.14 (16.4)	88.9
Impulsivity/Overactivity (*n* = 531)	.23 (16.9)	.15 (2.4)	-.15 (2.7)	76.7
Oral Reading Fluency (direct assessment)				
Fall ORF Mean of A & B passages (*n* = 554)	-.19 (1.5)	-.05[†] (1.2)[†]	.04[†] (0.0)[†]	56.0
Winter ORF Mean of A & B passages (*n* = 578)	-.21 (0.0)	-.04[†] (3.9)[†]	.04[†] (0.0)[†]	58.6
Spring ORF Mean of A & B passages (*n* = 554)	-.19 (1.5)	-.05[†] (1.2)[†]	-.04[†] (0.0)[†]	56.0
Learning Behaviours Scale (teacher rating), Trinidad and Tobago scale				
Strategy/Flexibility (*n* = 705)	-.29 (7.5)	-.05[†] (0.0)[†]	.09 (5.6)*	71.8
Competence Motivation (*n* = 705)	-.32 (10.0)	.12 (5.5)[†]	.05[†] (8.6)[†]	95.4

Note. Nonparenthetical entries are Pearson product moment correlations. Parenthetical entries indicate the percentage of variance in the

Table 2 continued

Criterion measure	Aggressive	Reticent/ Withdrawn	Irascible/Attention-Seeking	% Explainable variance[a]

respective criterion measure scores between children within classrooms that is accounted for by a given ASCA-H scale score. Values equal 1 - reduction in the intraclass correlation (100) as estimated via hierarchical linear modeling. Each two-level random coefficients model entered a given ASCA-H scale as the covariate. Correlations and fixed effects associated with ASCA-H scales are significant statistically at $p < .01$ unless indicated † (nonsignificant). ASCA = Adjustment Scales for Children and Adolescents, ORF = Oral Reading Fluency, ADH = Attention Deficit and Hyperactivity disorder.

[a]Total percentage of potentially explainable variance between children within classrooms. Values equal 1 – intraclass correlation (100), where the intraclass correlation was estimated via hierarchical linear modeling. Each two-level, unconditional means model applied random intercepts for classrooms, where the random effect was significant at $p < .001$.

Table 3. *Mean Population Distribution by Gender and Grade Level in Trinidad and Tobago*

		Aggressive		Reticent/ Withdrawn		Irascible/Attention -Seeking	
Gender		***M***	***(SD)***	***M***	***(SD)***	***M***	***(SD)***
				Infant 1			
Male	(*n* = 52)	51.4	(9.4)	53.4	(10.9)	50.0	(10.5)
Female	(*n* = 57)	50.3	(9.8)	49.7	(12.1)	50.7	(10.5)
				Infant 2			
Male	(*n* = 52)	51.5	(9.1)	52.4	(10.3)	48.0	(9.5)
Female	(*n* = 52)	49.4	(8.6)	55.0	(10.4)	52.3	(11.2)
				Standard 1			
Male	(*n* = 52)	52.3	(11.0)	50.4	(8.9)	49.9	(9.9)
Female	(*n* = 45)	47.4	(8.4)	51.8	(9.8)	49.7	(11.3)
				Standard 2			
Male	(*n* = 52)	51.4	(9.6)	51.8	(9.1)	48.4	(9.1)
Female	(*n* = 60)	48.0	(9.3)	49.7	(9.1)	50.7	(10.1)
				Standard 3			
Male	(*n* = 50)	50.5	(11.7)	47.4	(9.8)	48.4	(9.4)
Female	(*n* = 51)	50.7	(9.2)	47.4	(9.1)	51.3	(9.8)
				Standard 4			
Male	(*n* = 51)	49.0	(9.5)	47.7	(9.4)	49.7	(8.9)
Female	(*n* = 52)	47.8	(14.4)	48.5	(9.0)	50.6	(10.8)
				Standard 5			
Male	(*n* = 48)	51.3	(8.6)	47.3	(9.5)	49.6	(10.0)
Female	(*n* = 51)	50.4	(9.0)	47.3	(8.7)	49.0	(9.0)

Table 4. *Mean Population Distribution by Gender and Ethnicity in Trinidad and Tobago*

	Aggressive		Reticent/ Withdrawn		Irascible/Attention -Seeking	
Gender	***M***	***(SD)***	***M***	***(SD)***	***M***	***(SD)***
		African descent				
Male (*n* = 133)	51.1	(9.4)	50.3	(9.9)	47.4	(9.2)
Female (*n* = 136)	50.6	(11.2)	51.7	(10.3)	49.1	(9.9)
		East Indian descent				
Male (*n* = 136)	50.6	(10.8)	48.5	(9.3)	50.0	(9.2)
Female (*n* = 142)	46.6	(9.2)	47.6	(10.0)	52.8	(9.8)
		Mixed descent				
Male (*n* = 78)	52.2	(8.5)	52.1	(11.0)	50.2	(10.0)
Female (*n* = 78)	50.3	(8.8)	51.2	(9.5)	50.9	(11.4)

Discussion

The central purpose of this article is the introduction and demonstration of an advanced approach to factor analysis and measurement scaling. In the process, we focused on a new measuring device intended for application with parents of school children in Trinidad and Tobago. In a sense, the article has dual purposes, the primary being the methodological demonstration and the secondary the discoveries regarding the new measure. Thus our discussion is presented in two parts. The first part pertains to the measurement device under development and shows how this work can be understood within the context of extant research and practice. The second part recounts and reflects on the introduced methodology.

The Measure

Research that provides evidence for the reliability and validity of a new instrument should examine the results in the light of expiating theoretical and empirical literature. The present results of a scale with three dimensions: Reticent/Withdrawn, Attention-

Seeking, and Aggressive, do reflect elements of the robust internalizing/externalizing pattern described in the introduction. Correlations with other scales suggested some convergent validity, with Aggressive positively correlating with the DBDRS indicators of Inattention, Oppositional/Defiant, Impulsivity/Overactivity, and the ASCA Overactivity measure. Though the correspondence with ASCA teacher ratings was generally low, that was not surprising as parents and teachers have the opportunity to observe the child in different settings and many behaviours are context-specific (Dinnebeil et al., 2013; Hartley, Zakriski, & Wright, 2011; Lane, Paynter, & Sharman, 2013). This result supports the idea that in order to adequately assess and design effective interventions for students with behavioural difficulties, consideration of information from multiple sources is needed (Rescorla et al., 2014). Rather than looking for disagreement between parents and teachers, it might instead be beneficial to note children whose behaviours have been flagged by both. Thus, the number of settings in which behavioural difficulties are noted as well as the variety of contexts in which behavioural difficulties are apparent may both be clinically significant (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). As problem behaviours are contextually-based phenomena, this number is likely to be small. In the current study sample, the number of children identified by both parents and teachers as exhibiting behavioural difficulties was approximately 3%, a value reminiscent of that segment of the normative population beyond the 2nd *SD* above the mean.

When comparing parent and teacher ratings, it is also important to consider that the degree to which they differ may also depend on the nature of the behaviour being observed. For instance, there is some evidence that cross-informant agreement between teachers and parents tends to be lower when rating internalizing constructs rather than externalizing constructs (Lambert, Knight, Taylor, & Newell, 1992). Additionally, studies suggest that externalizing problems are more directly related to impacts on academic achievement than are internalizing problems (Nelson, Benner, Lane, & Smith, 2004), which may help to account for the lack of demonstrated association with academic achievement for Reticent/Withdrawn.

There are also important implications for child development as we explore age and gender differences. Prior research on the teacher-rating ASCA with Trinidadian students found that students of East Indian descent and males had higher scores on aggressive

syndromes and younger students displayed more avoidant and oppositional defiant problems (Grim, 2002). The current investigation suggested a similar gender effect for the ASCA-H Aggressive syndrome, a similar ethnic effect for the Irascible/Attention-Seeking syndrome, and age effect for Reticent/Withdrawn. Caution should be taken in interpreting these results as this study was not designed to explore the causes of age, gender, or ethnic differences in depth. For example, although the gender effect for Aggressive syndrome was not unexpected, as research indicates that school-age boys tend to have more aggression-related behaviour problems than girls (Beaman, Wheldall, & Kemp, 2006, Bertrand & Pan 2013), there is also research that concludes that girls simply express their aggression differently (Bjorkqvist, Lagerspetz, & Kaukiainen, 1992). Indeed there are many types of aggression (e.g., indirect, physical, social, relational), and extant literature suggests that possible gender biases may be inherent in behavioural scales including or excluding items representing certain of those types (Bjorkqvist & Niemela, 1992; Crick & Grotpeter, 1995). Differential results related to demographics would be a recommended area of future research.

The Methodology

ASCA-H was standardized for use with children in Trinidad and Tobago using an exploratory multidimensional IRT technique. While traditional factor analysis uses one source of information, FIFA relies on two kinds of information: the relationship between items (correlation matrix) and the pattern of responses from the informant about the child. This is a neat solution to the problem of a large number of items that is common to instruments that are intended to measure multiple dimensions of child behaviours. Large correlation matrices typically mean that various sets of items will tend to predict perfectly the responses to other items. That circumstance effectively precludes the basic mathematics required for factor analysis (Morrison, 1976). The Bock, Gibbons, & Muraki (1988) method of smoothing a tetrachoric correlation matrix, inherent in the FIFA procedure viewed here, produces a workable correlation matrix. Moreover, the tetrachoric correlations themselves are important features of the FIFA procedure because they prevent the spurious and inaccurate factors commonly associated with ordinary Pearson product-moment correlations when applied with dichotomous item response data (Bock et al., 1988; Waller, 2001).

Where the focus of this article was exploratory multidimensional IRT, the same solutions for large numbers of dichotomous items are available for confirmatory multidimensional IRT. Ordinary confirmatory factor analysis is usually not computationally feasible with such large numbers of items and has led to numerous attempts to work around the problem (Bandalos, 2002; Hall, Snell, & Singer Foust, 1999; Hau & Marsh, 2004; Little, Cunningham, Shahar, & Widaman, 2002; Sass & Smuth, 2006; Thompson & Melancon, 1996; Wilkinson, 2007). The work-arounds are essentially forced to abandon any factor analysis based on the original items of interest. However, full information factor analysis is perfectly suited for confirmatory work as well as the exploratory work demonstrated in this article, and such confirmatory work would be an important next step in assessing the construct stability of the ASCA-H in Trinidad and Tobago. Such confirmatory FIFA would require another large independent sample.

Perhaps just as important as the choice of the correct type of factor analysis for the Trinidad and Tobago data, were the steps that preceded factor analysis. Specifically, Fabrigar et al. (1999) have emphasized the importance of beginning with a large representative sample; that is, representative of the focal population or nation. That is why we began with a stratified random sample of the elementary school population instead of a sample of convenience. Additionally, Goldberg and Velicer (2006) stressed the importance of preliminary procedures to estimate the general range of the number of factors that might be extracted from a given correlation matrix. Here we applied the MAP method to guide the factor analytic procedure (Velicer, 1976). This procedure is designed to reduce the likelihood that the researcher would be led to overestimate the number of viable factors.

The ability to implement the exploratory FIFA method is not restricted to one software program. TESTFACT was chosen here because of the availability of extensive documentation and examples, and because it performs the exact analyses required for our problem solution. It also spares the researcher the complexity of mastering a larger multi-purpose program. TESTFACT carries the process all the way through resolving dimensionality, item calibration, scaling, and scoring. Another program, IRTPRO (Cai, Thissen, & duToit, 2011) offers an exploratory FIFA procedure that was similarly based on Bock et al. (1988) for multidimensional dichotomous items. The flexmirt program (Cai, 2013) and the R

package 'mirt' (Chalmers, 2012) also provide procedures capable of implementing these methods.

Conclusion

This study was limited by the instruments and measures available for the Trinidad and Tobago data collection, where the ASCA-H was the only parent rating scale and the ORF the only measure of academic achievement. As the low correspondence between teacher and parent ratings is common, it is difficult to establish whether the discrepancies are due to the differing demands of the relationships (child/teacher or child/parent) or different behaviours manifesting in different environmental contexts (school or home), or whether the correlations might be higher using another criterion scale. A future investigation would enhance the criterion validity by including other parental rating measures in order to determine whether the ASCA-H measures the same psychological adjustment constructs, and additional standardized academic measures to further inform validity analyses for academic achievement. Since the size of the national sample and large number of items in the studied instrument precluded division of the sample to enable subsequent confirmatory factor analysis, future research should also include an independent validation sample and confirmatory IRT procedure.

ASCA-H is the only available, contextually-based, parent rating scale for child problem behaviours in Trinidad and Tobago (Carrington-Blaides & Ramoutar, 2017). Our study aids in establishing the dimensionality, validity, and reliability of the ASCA-H for assessing children's problem behaviours in a home context by utilizing a large nationally representative sample, appropriate modern methodology, and an analysis in conjunction with other instruments established for application in Trinidad and Tobago.

About the Authors

Jessica L. Chao, Graduate School of Education, University of Pennsylvania (jec@gse.upenn.edu)

Paul A. McDermott, Graduate School of Education, University of Pennsylvania (paulmcd@gse.upenn.edu)

Marley W. Watkins, Department of Educational Psychology, Baylor University (Marley_Watkins@baylor.edu)

Michael J. Rovine, Graduate School of Education, University of Pennsylvania (mrovine@upenn.edu)

Anna Rhoad-Drogalis, Crane Center for Early Childhood Research and Policy, The Ohio State University (Rhoad-Drogalis.1@osu.edu)

Frank C. Worrell, Graduate School of Education, University of California, Berkeley (frankc@berkeley.edu)

Tracey E. Hall, Center for Applied Special Technology, Wakefield, MA (thall@cast.org).

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment*. University of Vermont. Research Center for Children, Youth & Families, Burlington, VT.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232. doi:10.1037/0033-2909.101.2.213

Achenbach, T. M., Rescorla, L. A., & Ivanova, M. Y. (2012). International epidemiology of child and adolescent psychopathology I: diagnoses, dimensions, and conceptual issues. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(12), 1261-1272. doi: 10.1016/j.jaac.2012.09.010

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., text rev.). Washington, DC: Author.

Atilola, O. (2015). Cross-cultural child and adolescent psychiatry research in developing countries. *Global Mental Health*, 2, e5. doi:10.1017/gmh.2015.8

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, *9*(1), 78-102. doi: 10.1207/S15328007SEM0901_5

Barrow, C. (2008). Early childhood in the Caribbean. *Working papers in early childhood development*, No. 47. The Hague, The Netherlands: Bernard van Leer Foundation. Retrieved from: https://eric.ed.gov/?id=ED522741

Beaman, R., Wheldall, K., & Kemp, C. R. (2006). Recent research on troublesome classroom behavior: A review. *Australasian Journal of Special Education, 31*(1), 45-60. doi: 10.1080/10300110701189014

Bernstein, I. H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105*(3), 467-477. doi: 10.1037/0033-2909.105.3.467

Bertrand, M. & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics, 5*(1), 32-64. Retrieved from http://www.jstor.org/stable/43189418

Bjorkqvist, K., Lagerspetz, M. J. & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression, *Aggressive Behaviour, 18*(2), 117-127. doi: 10.1002/1098-2337(1992)18:2<117::AID-AB2480180205>3.0.CO;2-3

Bjorkqvist, K. & Niemela, P. (1992). New trends in the study of female aggression. In K. Bjorkqvist and P. Niemela (Eds.), *Of mice and women: Aspects of female aggression* (pp 3-16). Academic Press; San Diego, CA.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443-459. doi: 10.1007/BF02293801

Bock, D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*(3), 261-280. doi:10.1177/014662168801200305

Cai, L. (2013). flexMIRT: A numerical engine for flexible multilevel multidimensional item analysis and test scoring (Version 2.0) [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Cai, L., Thissen, D., and du Toit, S. H. C. (2011). IRTPRO for Windows [Computer Software]. Lincolnwood, IL: Scientific Software International.

Canivez, G. L., & Beran, T. N. (2009). Adjustment Scales for Children and Adolescents: Factorial validity in a Canadian sample. *Canadian Journal of School Psychology, 24*(4), 284–302. doi:10.1177/0829573509344344

Canivez, G. L., & Beran, T. N. (2011). Learning Behaviors Scale and Canadian youths: Factorial validity generalization and comparisons to the U.S. standardization sample. *Canadian Journal of School Psychology*, *26*(3), 193–208. doi:10.1177/0829573511416440

Canivez, G. L., & Bohan, K. J. (2006). Adjustment Scales for Children and Adolescents and Native American Indians: Factorial validity generalization for Yavapai Apache youths. *Journal of Psychoeducational Assessment, 24*(4), 329–341. doi:10.1177/0734282906291397

Canivez, G. L., & Sprouls, K. (2005). Assessing the construct validity of the Adjustment Scales for Children and Adolescents. *Journal of Psychoeducational Assessment, 23*(1), 3–14. doi:10.1177/073428290502300101

Canivez, G. L., & Sprouls, K. (2010). Adjustment Scales for Children and Adolescents: Factorial validity generalization with Hispanic/Latino youths. *Journal of Psychoeducational Assessment, 28*(3), 209–221. doi:10.1177/0734282909349213

Canivez, G. L., Willenborg, E., & Kearney, A. (2006). Replication of the Learning Behaviors Scale factor structure with an independent sample. *Journal of Psychoeducational Assessment, 24*(2), 97–111. doi:10.1177/0734282905285239

Carrington-Blaides, E. & Ramoutar, A. S. (2017). Prevalence estimates of behavior problems in primary schools in Trinidad and Tobago: A baseline inquiry. In C. Bissessar (Ed.), *Assessing the current state of education in the Caribbean*. Hershey, PA: IGI Global.

Central Intelligence Agency. (2014). *The world factbook, 2013-2014, Trinidad and Tobago*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/index.html

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1-29. doi:10.18637/jss.v048.i06

Chao, J. L., McDermott, P. A., Watkins, M. W., Drogalis, A. R., Worrell, F. C., & Hall, T. E. (in press). The Learning Behaviors Scale: National standardization in Trinidad and Tobago. *International Journal of School & Educational Psychology*. doi: 10.1080/21683603.2016.1261055

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal*

Child Psychology, *26*(4), 257-268. doi: 10.1023/A:1022602400621

Costenbader, V. K., & Keller, H. R. (1990). Behavioral ratings of emotionally handicapped, learning disabled, and nonreferred children: Scale and source consistency. *Journal of Psychoeducational Assessment*, *8*(4), 485–496. doi:10.1177/073428299000800404

Crick, N. R. & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment, *Child Development, 66*(3), 710-722. doi: 10.2307/1131945

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507–524.

Dinnebeil, L. A., Sawyer, B. E., Logan, J., Dynia, J. M., Cancio, E., & Justice, L. M. (2013). Influences on the congruence between parents' and teachers' ratings of young children's social skills and problem behaviors*. Early Childhood Research Quarterly, 28*(1), 144–152. doi:10.1016/j.ecresq.2012.03.001

DiPerna, J., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievement: An introduction to the mini-series. *School Psychology Review, 31*(3), 293-297.

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior – theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry, 53*(5)*,* 558-574. doi:10.1111/j.1469-7610.2012.02537.x

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment*, *20*(1), 83-102. doi: 10.1023/A:1023087410712

Elliott, C. D. (1990). *Differential Abilities Scale: Introductory and technical handbook*. San Antonio, TX: Psychological Corporation.

Embretson, S. E., & Schmidt McCollam, K. M. (2000). Psychometric approaches to understanding and measuring intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 423-444). New York, NY: Cambridge University Press.

Evans, H., & Davies, R. (1997). Overview issues in childhood socialization in the Caribbean. In J. L. Roopnarine & J. Brown (Eds.), *Caribbean families: Diversity among ethnic groups* (pp. 1-24*)*. Greenwich, CT: Ablex.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. doi:10.1037/1082-989X.4.3.272

Friedman, K. A., Leone, P. E., & Friedman, P. (1999). Strengths-based assessment of children with SED: Consistency of reporting by teachers and parents. *Journal of Child and Family Studies*, *8*(2), 169–180. doi:10.1023/A:1022035817456

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256. doi:10.1207/S1532799XSSR0503_3

Goldberg, L. R. & Velicer, W. F. (2006). Principles of exploratory factor analysis. In Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed., pp. 209-237). New York: Springer.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257–288. doi: 10.1207/S1532799XSSR0503_4

Grim, S. E. (2002) *The assessment of youth psychopathology in Trinidad and Tobago: A cross -cultural construct validity study of the Adjustment Scales for Children and Adolescents (ASCA)* (Doctoral dissertation). Dissertations available from ProQuest. AAI3054946. https://repository.upenn.edu/dissertations/AAI3054946/

Hall, R. J., Snell, A. F., & Singer Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*(3), 232-256. doi: 10.1177/109442819923002

Hart, S. A., Logan, J. A. R., Soden-Hensler, B., Kershaw, S., Taylor, J., & Schatschneider, C. (2013). How nature and nurture affect the development of reading: An analysis of the Florida twin project on reading. *Developmental Psychology, 49*(10), 1971–1981. doi:10.1037/a0031348

Hartley, A. G., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Contextual influences on

divergence and convergence. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 54-66. doi: 10.1080/15374416.2011.533404

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*(7), 636-644. doi:10.1598/RT.59.7.3

Hau, K., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, *57*(2), 327-351. doi: 10.1111/j.2044-8317.2004.tb00142.x

Ho, C., Bluestein, D. N., & Jenkins, J. M. (2008). Cultural differences in the relationship between parenting and children's behavior. *Developmental Psychology*, *44*(2), 507-522. doi: 10.1037/0012-1649.44.2.507

Konold, T. R., Brewster, J. C., & Pianta, R. C. (2004). The behavior of child behavior ratings: Measurement structure of the Child Behavior Checklist across time, informants, and child gender. *Journal of Behavioral Disorders, 29*(4), 372-383. Retrieved from http://www.jstor.org/stable/23889529

Lambert, M. C., Lyubansky, M., & Achenbach, T. M. (1998). Behavioral and emotional problems among adolescents of Jamaica and the United States: Parent, teacher, and self-reports for ages 12 to 18. *Journal of Emotional and Behavioral Disorders*, *6*(3), 180-187. doi: 10.1177/106342669800600306

Lambert, M. C., Weisz, J. R., & Knight, F. (1989). Over-and undercontrolled clinic referral problems of Jamaican and American children and adolescents: The culture general and the culture specific. *Journal of Consulting and Clinical Psychology*, *57*(4), 467-472. doi: 10.1037/0022-006X.57.4.467

Lambert, M. C., Weisz, J. R., Knight, F., Desrosiers, M. F., Overly, K., & Thesiger, C. (1992). Jamaican and American adult perspectives on child psychopathology: Further exploration of the threshold model. *Journal of Consulting and Clinical Psychology*, *60*(1), 146-149. doi: 10.1037/0022-006X.60.1.146

Lane, B. R., Paynter, J., & Sharman, R. (2013). Parent and teacher ratings of adaptive and challenging behaviours in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *7*(10), 1196-1203. doi: 10.1016/j.rasd.2013.07.011

LeBoeuf, W. A., Fantuzzo, J. W., & Lopez, M. L. (2010). Measurement and population miss-fits: A case study on the importance of using appropriate measures to evaluate early childhood

interventions. *Applied Developmental Science, 14*(1), 45–53. doi:10.1080/10888690903510349

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the questions, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173. doi: 10.1207/S15328007SEM0902_1

Malecki, C., & Elliott, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*(1), 1–23. doi: 10.1521/scpq.17.1.1.19902

McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment, 5*(4), 413–424. doi:10.1037/1040-3590.5.4.413

McDermott, P. A. (1999). National scales of differential learning behaviors among children and adolescents. *School Psychology Review, 28*(2)*,* 280-291.

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (1999). *Learning Behaviors Scale.* Philadelphia, PA: Edumetric and Clinical Science.

McDermott, P. A., Steinberg, C. M., & Angelo, L. E. (2005). Situational specificity makes the difference in assessment of youth behavior disorders. *Psychology in the Schools, 42*(2), 121–156. doi:10.1002/pits.20050

McDermott, P. A., Stott, D. H., & Marston, N. C. (1993). *Adjustment Scales for Children and Adolescents*. Philadelphia, PA: Edumetric and Clinical Science.

McDermott, P. A., Watkins, M. W., Rhoad, A. M., Chao, J. L., Worrell, F. C., & Hall, T. E. (2015). Trinidad and Tobago national standardization of the Adjustment Scales for Children and Adolescents. *International Journal of School and Educational Psychology, 3*(4)*,* 278-292. doi:10.1080/21683603.2015.1067873

McDermott, P. A., Watkins, M. W., Rhoad, A. M., Chao, J. L., Worrell, F. C., & Hall, T. E. (2016). Classroom contexts as the framework for assessing social-emotional adjustment: A National study in Trinidad and Tobago. *Psychology in the Schools*, *53*(6), 626-640. doi: 10.1002/pits.21930

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents*. Mahwah, NJ: Erlbaum.

Morrison, D. F. (1976). *Multivariate statistical methods*. New York: McGraw-Hill.

Mpofu, E., Oakland, T., Ntinda, K., Seeco, E., & Maree, J. G. (2014). Constructing a framework for the use of tests within a developing nation's school system. *International Perspectives in Psychology: Research, Practice, Consultation, 3*(2), 106–122. doi:10.1037/ipp0000015

Muraki, E., & Engelhard Jr, G. (1985). Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement*, *9*(4), 417-430. doi: 10.1177/014662168500900411

Pearson, K. (2007). *Parent rating of youth behavior in the Republic of Trinidad and Tobago on the Adjustment Scales for Children and Adolescents: Parent Edition.* (Unpublished doctoral dissertation). Pennsylvania State University, State College, PA.

Pelham, W. E., Jr., Evans, S. W., Gnagy, E. N., & Greenslade, K. E. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders: Prevalence, factor analyses, and conditional probabilities in a special education sample. *School Psychology Review, 21*(2), 285–299.

Pelham, Jr, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 449-476. doi:10.1207/s15374424jccp3403_5

Pelham, W. E. Jr., Gnagy, E. N., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child & Adolescent Psychology, 31*(6), 210–218. doi:10.1097/00004583-199203000-00006

Petscher, Y., & Kim, Y. S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology, 49*(1), 107–129. doi:10.1016/j.jsp.2010.09.004

Rescorla, L. A., Bochicchio, L., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I....Verhulst, F. C. (2014). Parent-teacher agreement on children's problems in 21 societies. *Journal of Clinical Child & Adolescent Psychology, 43*(4), 627-642. doi:10.1080/15374416.2014.900719

Sass, D. A., & Smuth, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling*, *13*(4), 566-586. doi: 10.1207/s15328007sem1304_4

Sattler, J. M. (1992). *Assessment of children: WISC—III and WPPSI—R supplement*. San Diego, CA, England: Jerome M. Sattler.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, *30*(3), 407–419.

The Psychological Corporation. (1983). *Basic Achievement Skills Individual Screener*. New York, NY: Author.

Thompson, B., & Melancon, J. (1996, November). *Using item "testlets/parcels" in confirmatory factor analysis: An example using the PPDP-78*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Tuscaloosa, AL. (ERIC Document Reproduction Service No. ED 404 349)

United Nations Children's Fund. (2013). *The State of the world's children 2013*. New York, NY: Author.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321–327. doi:10.1007/BF02293557

Waller, N. G. (2001). MicroFACT 2.0: A microcomputer factor analysis program for ordered polytomous data and mainframe size problems. St. Paul, MN: Assessment Systems Corporation.

Watkins, M. W., Hall, T. E., & Worrell, F. C. (2014). From Central Guidance Unit to Student Support Services Unit: The outcome of a consultation process in Trinidad and Tobago. *Journal of Educational and Psychological Consultation, 24*(4), 283–306. doi:10.1080/10474412.2014.929962

Wentzel, K. R. (2009). Peers and academic functioning at school. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships and groups* (pp. 531–547). New York, NY: Guilford.

Waithe, D., & Worrell, F. C. (2003). The development of the steel band in Trinidad and Tobago. *Juniata Voices*, 41–47.

Wilkinson, W. W. (2007). The structure of the Lawrence locus of control scale in young adults: Comparing item and parcel indicator models. *Personality and Individual Differences, 43*(6), 416-1425.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*(2), 85-104. doi:10.1207/s15326977ea1102_1

Wood R, Wilson D.T., Gibbons R.D., Schilling S.G., Muraki E., Bock R.D. (2003). TESTFACT 4 for Windows: Test Scoring, Item

Statistics, and Full-Information Item Factor Analysis. Scientific Software International.

Worrell, F. C. (2006a). Children and youth in poverty in Trinidad and Tobago: A lack of commitment in the midst of plenty. In C. C. Yeakey (Series Ed.) & C. C. Yeakey, J. W. Richardson, & J. B. Buck (Vol. Eds.), *Advances in education in diverse communities: Research, policy and praxis: Vol. 4. 'Suffer the little children': National and international dimensions of child poverty and public policy* (pp. 145–175). San Diego, CA: Elsevier.

Worrell, F. C. (2006b). Ethnic and gender differences in self-reported achievement and achievement-related attitudes in secondary school students in Trinidad. *Caribbean Curriculum, 13,* 1–22.

Worrell, F. C., Vandiver, B. J., & Watkins, M. W. (2001). Construct validity of the Learning Behaviors Scale with an independent sample of students. *Psychology in the Schools, 38*(3), 207–215. doi: 10.1002/pits.101