

# The 10-Item Positive and Negative Affect Schedule for Children, Child and Parent Shortened Versions: Application of Item Response Theory for More Efficient Assessment

Chad Ebesutani · Jennifer Regan · Ashley Smith ·  
Steven Reise · Charmaine Higa-McMillan ·  
Bruce F. Chorpita

Published online: 4 January 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** The Positive and Negative Affect Schedule for Children (PANAS-C/P; child and parent versions) yield positive affect (PA) and negative affect (NA) scales that are clinically useful for identifying youth with anxiety and mood problems. Despite the advantages that item response theory (IRT) offers relative to classical test theory with respect to shortening test instruments, no studies to date have applied IRT methodology to the PANAS-C/P scales. In the present study, we thus applied IRT methodology using a school-based development sample (child sample: N=799; parent sample: N=553) and developed a shortened 5-item PA scale (*joyful, cheerful, happy, lively, proud*) and a 5-item NA scale (*miserable, mad, afraid, scared, sad*) for the sake of simultaneously increasing the assessment efficiency of the PANAS-C/P scales while improving the psychometric properties of the scales. The reduced PA and NA child scales classified relevant diagnostic groups in a separate clinic-

referred validation sample (N=662) just as well as the original PANAS-C child scales and may be used to help identify youth with internalizing disorders in need of mental health services.

**Keywords** PANAS-C · Positive Affect · Negative Affect · Item Response Theory · Depression · Psychometric properties

Positive Affect (PA) is a broad affective dimension characterized by feelings of enthusiasm, alertness, and activity, with high PA characteristic of “high energy, full concentration, and pleasurable engagement” with one’s environment (Watson et al. 1988, p. 1063). PA has become an increasingly important concept in clinical psychology given its close association with depression in relation to the tripartite model (Clark and Watson 1991), as well as its ability to identify individuals with mood problems (Chorpita and Daleiden 2002). Given the importance of PA in understanding and identifying youths with depressive problems, affective neuroscience researchers have also recently begun researching PA and how the neural processes of PA relate to depression in youth (Forbes and Dahl 2005). This increased attention towards better understanding and measuring PA is fitting given previous research showing that PA assessment provides one of the best methods for identifying and differentiating youths with depression apart from youths with other internalizing and externalizing problems (e.g., Chorpita and Daleiden 2002). The PA scale is thus considered a clinically useful tool to aid in differential diagnosis in the clinical assessment process.

Related to PA is negative affect (NA), another temperamental factor. NA is also related to mood states, but is more directly associated with emotionally distressing experiences

---

Adapted from Watson, D. & Clark, L.A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule—Expanded form-Revised. Copyright 1994 by D. Watson and L. A. Clark; all rights reserved. PANAS-X adapted with permission.

---

C. Ebesutani (✉)  
Department of Psychology, Yonsei University,  
50 Yonsei-ro, Seodaemun-gu,  
Seoul, South Korea 120-749  
e-mail: ebesutani@yonsei.ac.kr

J. Regan · A. Smith · S. Reise · B. F. Chorpita  
Psychology Department, University of California at Los Angeles,  
Los Angeles, CA, USA

C. Higa-McMillan  
Psychology Department, University of Hawaii,  
Hilo, HI, USA

such as feelings of sadness, fear, guilt and anger. PA and NA have been shown to be related via the tripartite model (Clark and Watson 1991) whereby NA is considered to be a shared component of anxiety and depression. On the other hand, (low) PA is specifically related to depression, and as noted above, yields clinically useful scale scores that may help differentiate depressed youths from youths with other internalizing problems, including anxiety and other experiences related to negative affect (e.g., Chorpita and Daleiden 2002).

The Positive and Negative Affect Schedule for Children, child (PANAS-C; Laurent et al. 1999) and parent versions (PANAS-C-P; Ebesutani et al. 2011a) were developed to measure PA and NA in youth from child and parent perspectives. The PANAS-C (child version) has evidenced favorable psychometric properties for its 12-item PA scale and 15-item NA scale across various independent clinical and non-clinical samples (e.g., Laurent et al. 1999; Chorpita and Daleiden 2002; Hughes, and Kendall 2009). The PANAS-C has also been used for research purposes to further understand the relationship between anxiety and depression (e.g., Jacques, and Mash 2004). The PANAS-C-P (parent version) has also recently been shown to have favorable psychometric support in a school sample of 606 children and adolescents, including support for a two factor structure of NA and PA, high internal consistency for both scales and significant convergent/divergent validity statistics (Ebesutani et al. 2011a).

Despite the support for the psychometric properties of the PANAS-C and PANAS-C-P, these scales might prove more useful for certain contexts (e.g., school-based mental health screenings, community clinics) if further shortened given the time constraints imposed in these contexts as well as the large amount of assessment measures typically given to clients and students alike in a single assessment administration. School screenings, for example, are becoming increasingly more ubiquitous for the purpose of identifying youth in need of mental health services as well as those at risk of developing various forms of psychopathology associated with (educational) impairment in the future. The Behavioral Vital Signs (BVS) project, for example, is a recent state-wide initiative in Mississippi whereby empirically-based mental health screenings are administered to school-aged youths in grades 2–12. The purpose of the BVS project is to provide school administration with information regarding their students' mental health statuses to inform future intervention initiatives to reduce various problems that may be impairing school functioning, including anxiety, depression, and substance use. Such school screenings, however, typically involve the administration of multiple questionnaires to assess various relevant domains. School screenings are also typically administered during school hours, which should be considered when compiling assessment batteries.

Managed care has also begun imposing constraints on the amount of time clinicians may bill for working with clients (e.

g., Christensen and Jacobson 1994; Richardson and Austad 1991). In several states, for example, clinicians are only reimbursed for a 50–60 min initial assessment intake interview—during which they need to collect all needed assessment information, including information often coming from multiple, different self-report questionnaires, to formulate a diagnostic summary and establish a treatment plan. Having shorter versions of the PANAS-C (child and parent versions) that yield scale scores to aid in the identification of NA-elevated youth (e.g., youth with anxiety and/or depression) and in the differential diagnosis of youth with affective-problems (e.g., depression) from youth with other NA-related problems (e.g., anxiety, anger) would provide mental health practitioners with an additional assessment tool to help meet the challenges of providing adequate assessment in a relatively short amount of time. Finding ways to further reduce assessment batteries for use in both applied and research settings is thus needed.

It is also notable that recent studies of both the PANAS-C and PANAS-C-P have shown that not all NA and PA items were associated with ideal psychometric properties (Ebesutani et al. 2011a; b). The “*calm*” PA item, for example, lacked face validity and also evidenced relatively low item-total correlations and factor loadings on the PA factor. The “*jittery*” NA item also showed relatively weaker association with the NA latent construct relative to the other NA items. There thus remains ways in which these scale may be refined, such as through eliminating these relatively poorly performing items. Lastly, item response theory (IRT; Embretson and Reise 2000) is a powerful statistical methodology that has recently begun to be applied to in the field of clinical psychology to shorten assessment scales (Reise and Waller 2009). IRT is especially useful for these purposes because, unlike classical test theory (CTT), it focuses on individual test items rather than observed scores on entire instruments, allowing researchers to gather information on an individual's underlying trait score using fewer test items. CTT posits that an individual's observed score is made up of a true score, the average score the individual would receive over multiple administrations of the exact same test, and error, the deviation from the score on one of these administrations and the average score. Given this model, in CTT, person measurement and quantitative indices of the instrument itself are dependent on specific instrument and sample characteristics. In addition, CTT is based on the largely implausible assumptions that, in a population, error scores are uncorrelated with true scores and that errors on one instrument are uncorrelated with true scores and errors on a different instrument. IRT models, on the other hand, are based on mathematical functions that characterize the relationship between an individual's trait level and the probability of responding to an item in a specific category. In IRT, scaling of the latent trait does not depend on a particular set of items. No studies to date, however, have applied IRT-based

methodology to the PANAS-C or PANAS-C-P scales to shorten these instruments. In addition to being able to identify the most discriminating items among an item set, IRT techniques also have the ability to successfully shorten scales without compromising assessment precision (e.g., Reise 2009).

### The Present Study

The purpose of the present study was to use IRT analyses to shorten the PANAS-C/P scales (child and parent versions) as well as to improve the psychometric properties of the NA and PA scales by identifying and retaining the most discriminating and informative items and eliminating the remaining items with less discriminating properties and less “item information.” Although IRT techniques have thus far been applied primarily to refine educational and licensure-related tests, they have recently been shown to also be applicable to psychological instruments (Reise 2009). We thus hypothesized that IRT analyses would be able to identify highly discriminating and informative NA and PA items to inform the development of a reduced version of the PANAS-C (child and parent versions) for more efficient use in applied and research settings. We predicted that “*calm*” (of the PA items) and “*jittery*” (of the NA items) would be identified for removal given that these items have evidenced somewhat weaker psychometric properties in previous studies relative to the other scale items. We also hypothesized that the reduced PA and NA scales of the PANAS-C and PANAS-C-P would meet benchmarks for Cronbach’s alpha coefficients. In terms of the comparative performance of the original and reduced PA scales, we hypothesized that the normalized Test Information Curves (TICs) of the reduced NA and PA scales (for both the PANAS-C and PANAS-C-P) would be comparable to the normalized TICs of the original scales due to the ability of IRT techniques to identify the most informative items. Lastly, we hypothesized that, using a separate clinic-referred validation sample, the shortened PANAS-C NA scale would (a) significantly discriminate between youths with anxiety and/or depression from youths without anxiety or depression just as well as the original 15-item NA scale. Regarding the PANAS-C PA scale, we hypothesized that the shortened PANAS-C PA scale would be able to discriminate youths who met criteria for a mood disorder from (b) youths who did not meet criteria for a mood disorder, (c) youths who met criteria for an externalizing disorder and no mood disorder and (d) youths who met criteria for an anxiety disorder and no mood disorder (cf. Chorpita and Daleiden 2002) just as well as the original 12-item PA scale.

### Method

#### Participants

The development sample consisted of youth (N=799) and their parents (N=553) who consented to participate in a large, school-based study in private and public schools across Hawaii. Criteria for inclusion in the development (school-based) sample included youth being between 6 and 18 years old and having PANAS-C or PANAS-C-P forms available with no missing data. The validation (clinic-referred) sample consisted of youth (N=662) consecutively referred to receive mental health assessments at a CBT clinic in Hawaii. Criteria for inclusion in the clinic-referred validation sample also included youth being between 6 and 18 years old and having PANAS-C and diagnostic data available. Youth and caregiver demographic information appears in Table 1 for both the clinic-referred and school sample. Diagnostic information for the clinic-referred sample appears in Table 2.

**Table 1** Youth and Caregiver Demographic Information

	Clinic-referred Sample		School Sample	
	<i>n</i>	Percentage	<i>n</i>	Percentage
Youth Gender				
Boys	433	65.4	323	40.4
Girls	227	34.3	456	57.1
Missing	2	0.3	20	2.5
Youth Ethnicity				
Multiethnic	306	46.2	185	23.2
White	81	12.2	16	2.0
African American	10	1.5	12	1.5
Asian American	129	19.5	266	33.3
Latino/Hispanic	8	1.2	2	0.3
Other	89	13.4	9	1.1
Missing	39	5.9	7	0.9
Caregiver Marital Status				
Married	285	43.1	391	48.9
Divorced, separated	179	27.0	50	6.3
Widowed	17	2.6	4	0.5
Single	91	13.7	31	3.9
Missing	90	13.6	0	0.0
Family Income				
\$0–\$30,000	399	60.3	60	7.9
\$30,001–\$60,000	114	17.2	106	13.3
\$60,001–\$90,000	67	10.1	106	13.3
\$90,001 or more	53	8.0	185	23.2
Missing	29	4.4	0	0.0

**Table 2** Number of Anywhere and Primary Diagnoses Among the Clinic-referred Youth (N=662)

Diagnoses	Anywhere	Primary
Depressive Disorders	61	40
Major depressive disorder	37	26
Dysthymic disorder	18	11
Depressive Disorder NOS	6	3
Anxiety Disorders	298	146
Panic disorder and/or agoraphobia	11	7
Specific Phobia	49	10
Generalized anxiety disorder	42	20
Separation anxiety disorder	32	18
Social phobia	99	54
Obsessive-compulsive disorder	21	12
PTSD/ASD	32	19
Anxiety NOS	12	6
ADHD	185	122
ADHD-Combined Type	84	52
ADHD-PI	85	62
ADHD-PH	8	3
ADHD-NOS	8	5
Disruptive Behavior Disorders	280	209
Oppositional defiant disorder	148	93
Conduct disorder	97	86
Disruptive behavior disorder NOS	35	30
Schizophrenia	0	0
Bipolar	3	3
Substance Abuse/Dependence	37	10
PDD	11	11
Adjustment Disorders	22	19
Other	71	31
No Diagnosis	122	122
Missing	4	4

*Anywhere* a diagnosis that appears anywhere in a child's diagnostic profile; *Primary* a child's primary diagnosis *NOS* not otherwise specified; *PTSD* post-traumatic stress disorder; *ASD* acute stress disorder; *ADHD* attention-deficit/hyperactivity disorder; *PI* Primarily Inattentive Type; *PH* Primarily Hyperactive Type; *PDD* pervasive developmental disorder. A small portion of youth (n=55; 8.3%) received co-primary diagnoses, accounting for the sum of all primary diagnoses being greater than the sample size of 662

## Measures

*Positive and Negative Affect Schedule for Children* (PANAS-C, Laurent et al. 1999). The PANAS-C is a 27-item youth self-report measure used in child and adolescent populations to measure PA and NA. The PANAS-C asks youth to rate adjectives of varying mood states based on how often they have felt that way in the past few weeks using a 5-point Likert scale ranging from 1 ("very slightly or not at all") to 5 ("extremely"). Laurent and colleagues'

(1999) study revealed that the 12-item PA and 15-item NA scale scores demonstrated support for convergent and divergent validity and also yielded internal consistency estimates ( $\alpha_{PA}=.89$ ;  $\alpha_{NA}=.92$ ) that met Nunnally and Bernstein's (1994) cut-off for "good" internal consistency (i.e., Cronbach's alpha >.80).

*Positive and Negative Affect Schedule for Children-Parent Version* (PANAS-C-P; Ebesutani et al. 2011a). The PANAS-C-P is the parent report version of the PANAS-C consisting of the same 27-items as the child version. It is also rated on a 5-point Likert scale ranging from 1 ("very slightly or not at all") to 5 ("extremely"). The items were developed to mirror those on the PANAS-C, but the instructions were adapted to accommodate the perspective of parents. The PANAS-C-P recently evidenced adequate psychometric support in a school-based sample of 606 children and adolescents, including fit indices that supported a two factor structure of NA and PA and consistency with the tripartite model of anxiety and depression (Ebesutani et al. 2011a).

*Anxiety Disorders Interview Schedule for DSM-IV, Child Version* (ADIS-IV-C; Silverman and Albano 1996). The ADIS-IV-C (Silverman and Albano 1996) is a semi-structured diagnostic interview based on *DSM-IV* diagnostic criteria and is used to assess for childhood anxiety, attention, behavioral, and mood problems. The ADIS-IV-C yields clinical psychiatric diagnoses based on the child interview, and it has been shown to have high interrater reliability for principal diagnoses ( $\kappa = .92$ ; Lyneham et al. 2007) and adequate concurrent validity for internalizing disorders (Wood et al. 2002).

## Procedure

The samples described in the current study were derived from two different larger research studies. Both projects received Institutional Review Board approval at the university. The first development sample was drawn from a larger school-based survey of child anxiety and depression. Parental consent was obtained through take-home forms that were signed and returned to school. Child assent was obtained prior to the administration of the PANAS-C as well as a battery of other questionnaires in a group format at school. Assistance was provided if children had difficulty reading and/or filling out questionnaires. Children received a \$5 gift certificate for participating.

The second validation (clinic-referred) sample was drawn from a Hawaii CBT clinic, from which children and adolescents were consecutively referred for mental health assessments. In this sample, parents completed consent forms and children completed assent forms prior to completion of a

battery of questionnaires (including the PANAS-C) and a semi-structured diagnostic interview (the Anxiety Disorders Interview Schedule for DSM-IV child version). The diagnostic interviews were conducted by clinical psychology doctoral student assessors. Assessors in the present study were trained using the ADIS-C. Training in the administration of the ADIS-C followed the standardized procedures as outlined by the authors (Silverman and Albano 1996) and also involved (a) observation of three ADIS-C interviews conducted by trained Center for CBT assessors, (b) conducting a series of five ADIS-C interviews while being observed by a trained assessor, and (c) matching the trained assessor on all clinical diagnoses in at least three of the five interviews. In a subset of 16 randomly selected families in this sample, interrater agreement for principal diagnoses between trainees who administered the ADIS-IV and trainers who observed the administration was found to be excellent ( $\kappa=.77$ ). Assessors were also blind to the PANAS-C scores while formulating diagnoses. Children in this sample were not compensated for participating in the mental health assessment.

#### Data Analytic Approach

*Essential Independence* To justify applying item response theory (IRT) to the NA and PA items in the present study, we first examined the dimensionality of the NA and PA scales to determine whether these scales were both unidimensional to meet the *essential independence* assumption required for IRT analysis (Stout 1990). Although NA and PA are generally known to be separate unidimensional constructs (Bagozzi 1993; Laurent et al. 1999), we nonetheless evaluated the fit of the two separate one-factor models among the NA and PA items using our present dataset with Mplus version 4.21 (Muthén and Muthén 2007). We used the robust maximum likelihood estimator and evaluated model fit using the following fit indices and criteria: the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and the Root Mean Square Error of Approximation (RMSEA). CFI and TLI values of .90 or greater (Bentler, 1990) and RMSEA values of .08 and lower represent good fit (Browne & Cudeck, 1993).

*Item Response Theory* Using the Multivariate Software IRT software package (mvIRT; Multivariate Software, Inc. 2010), we calculated the slope parameters ( $\alpha$ ) from the graded response model (GRM; Samejima 1969) separately for each of the 12 PA items and each of the 15 NA items from the PANAS-C (child version) and PANAS-C-P (parent version) using the development sample to identify the most discriminating items across both PANAS-C and PANAS-C-P instruments. The graded response model is appropriate for use with scales that have response choices of ordered

categories, such as the PANAS-C (e.g., 1, 2, 3, 4, 5 serving as the five response choices corresponding to “*very slightly or not at all*”, “*a little*”, “*moderately*”, “*quite a bit*”, and “*extremely*”, respectively). Slope parameters ( $\alpha_i$ ) are analogous to factor loadings and reflect an item’s discriminating power (i.e., the ability for a given item to differentiate among individuals at different ranges of the latent trait of interest (e.g., Positive Affect). We considered items with slope parameters greater than or equal to 1.7 as highly discriminating items (Baker 2001) and considered these items to be included in the final reduced scales. We also evaluated location parameters ( $\beta$ ) associated with each item. Location parameters for a given item indicate the points on the latent trait scale (e.g., the PA scale) at which an individual has a 50% chance of responding above the threshold between categories (e.g., 1 vs. 2, 3, 4, 5; 1, 2, vs 3, 4, 5). Location parameters are also important to examine to ensure that the included items of a scale provide discrimination at varying levels of the underlying latent trait. However, we also assessed for items associated with extreme (positive or negative) location parameters (e.g., greater than 4 standard deviations above or below the mean of the latent trait of NA or PA) and considered these items for removal. Such items would not be optimal to retain in our reduced scales given that their extreme location parameters indicate that these items would provide discrimination only for those who fall at extreme levels on the latent trait—thus providing discrimination for only very few people who complete the form. Other criteria that had to be met in the scale reduction process were that each scale had to be reduced by at least 50% so as to achieve adequate scale reduction for increased measurement efficiency (resulting in the reduced NA scale consisting of no more than seven items and a reduced PA scale consisting of no more than six items); further, each scale needed to include at least three items to allow for future SEM/CFA techniques to be employed on the scales in future research. We also sought to retain adequate construct coverage of the original NA and PA scale by following Watson and Clark’s (1997) recommendation to include items that tap “fearful/anxious, sad/depressed, and angry/hostile mood, as well as some type of positive affect” (Watson & Clark, 1997, p 294). Lastly, we used normalized Test Information Curves (TICs; described more below) to aid in the item selection process as these TICs allowed us to examine the degree to which the retained items were associated with greater test information relative to the discarded items.

In terms of comparing the properties of the original and reduced NA and PA scales, the item parameters from the graded response model may be converted into Item Information Curves (IICs). “Information” is a function of the square of the item’s discrimination parameter and inversely related to an individual’s standard error of measurement. Therefore, an item’s “information” parameter (often

displayed in the form of an information curve) reflects how precisely an item can measure individuals at each point across the continuum of the latent trait targeted by that item. Also importantly, IICs can be summed across all the items comprising the test to obtain a total Test Information Curve (TIC), indicating how much “information” the entire test provides in measuring that construct for individuals that fall along the continuum of the latent construct (e.g., PA). Relatedly, a *normalized* TIC provides an index of how much information is provided by a test (e.g., the PA scale) controlling for the number of items—normalized TICs are derived by simply dividing TICs by the number of test items.<sup>1</sup> TICs and normalized TICs for different measures (e.g., the original 12-item PA scale versus the shortened PA scale) may then be (visually) compared by overlaying their respective curves on a single graph to determine which test/scale provides more “information” on average across the test/scale items. We plotted and compared the *normalized* TIC of the reduced NA and PA scales relative to the original scales, given that it was expected for the reduced scales to have less overall test information due to TICs, by definition, being the sum of all IIC’s comprising the scale. By plotting and comparing the *normalized* TICs, we were able to examine whether the reduced scale obtains comparable information controlling for the number of items comprising the scale. Given that the purpose of IRT was to identify the most discriminating and informative items, we expected that visual inspection of the curves would show that the normalized TIC of the reduced scales would be at least comparable to, if not greater than, the TIC of the original scales.

We also calculated Cronbach’s alpha coefficients to examine the internal consistency of the reduced NA and PA scales relative to the original scales for both the PANAS-C and PANAS-C-P in the school development sample. We used .80 as the cut-off for acceptable internal consistency (Nunnally and Bernstein 1994).

Using a separate clinic-referred validation sample, we then examined the degree to which the shortened NA scale was able to discriminate youths with NA-related clinical disorders (i.e., anxiety and/or depression, based on the ADIS-C) from youths without these disorders relative to the original NA scale. We also examined the degree to which the reduced PA scale could discriminate youths who met criteria for a mood disorder from (a) youths who did not meet criteria for a mood disorder, (b) youths who met criteria for an externalizing disorder and no mood disorder and (c) youths who met criteria for an anxiety disorder and no mood disorder. These discriminative analyses were based on a recent study that also examined and compared the

discriminative properties and clinical utility of two measures of PA (cf. Chorpita and Daleiden 2002).

To examine these questions, we conducted group ANOVAs and calculated receiver operating characteristic (ROC) Area Under the Curve (AUC) values for the shortened and original PANAS-C scales<sup>2</sup> (with respect to the diagnostic classifications noted above), using Analyze-It for Microsoft Excel version 2.12 (Analyze-It Software Ltd. 2008). An ROC AUC value indicates the degree to which a scale (e.g., the reduced PA scale) accurately predicts the presence or absence of a clinical diagnosis (e.g., depression). Larger AUC values represent better prediction power of diagnostic status, and AUC values significantly greater than .50 indicate that the scale can classify diagnostic group status better than chance. AUC values may also be interpreted as follows: .50–.70, poor; .70–.80, fair; .80–.90, good; .90–1.00, excellent (c.f. Ferdinand 2008). We evaluated the comparative performance between the reduced and original scales with respect to discriminating the various diagnostic groups noted above via *z*-test comparisons of AUC values (DeLong et al. 1988).

## Results

### Scale Development

*Essential Independence* Both PA and NA scales in the child and parent developmental (school-based) samples evidenced essential independence warranting application of IRT on the NA and PA items. Specifically, as expected, a unidimensional model fit to the PA items evidenced good model fit among the PANAS-C child version (RMSEA=.06; CFI=.94; TLI=.92) and the PANAS-C-P parent version (RMSEA=.08; CFI=.93; TLI=.91). Regarding the NA items, it is notable that Ebesutani and colleagues (2011b) recently found that “fear” and “distress” group factors comprise the NA items of the PANAS-C, along with an overarching “NA” general factor consistent with a “restricted” bifactor model (Gibbons and Hedeker 1992), and that this model fit better than a standard one-factor model. And importantly, scales that evidence a bifactor structure are also often appropriate for IRT modeling (Reise et al. 2010). In such “restricted” bifactor models, each item is allowed to load on (only) one of multiple group factors as well as on the general common factor. The group factors in bifactor models are orthogonal to each other as well as to the general

<sup>1</sup> Normalized TICs are obtained as optional output available in the mvIRT program (Multivariate Software, 2010).

<sup>2</sup> Due to PANAS-C-P (parent report) data not being available in the clinic-referred sample, we were unable to conduct these AUC validity analyses using the reduced PANAS-C-P PA scale among the clinic-referred sample.

factor. A bifactor structure of NA (positing the “fear” and “distress” group factors and the common “NA” factor) was supported in the present sample with the PANAS-C items (RMSEA=.06; CFI=.94; TLI=.92) and the PANAS-C-P items (RMSEA=.06; CFI=.95; TLI=.93). Together, these results provided support for proceeding with IRT analysis, below.

*IRT Slope Parameter and Location Estimates* The IRT item slope parameter estimates ( $\alpha$ ) from the graded response model for the 12 PA items and 15 NA items of the PANAS-C and PANAS-C-P based on the school-based development sample appear in Table 3 (child version) and Table 4 (parent version). Estimated location parameters ( $\beta$ ) also appear in these tables for each of the four “locations” on the PANAS-C and PANAS-C-P.

Based on the item selection criteria noted above, we identified five PA items across both the PANAS-C and the

PANAS-C-P to comprise the reduced PA scale (e.g., *joyful, cheerful, happy, lively, proud*). These items were associated with high discrimination parameters, reduced the scale by at least 50%, and adequately tapped a somewhat broad range of positive affect, including lively, proud and happiness. Although a few other PANAS-C-P (parent) PA items also evidenced high discrimination properties (e.g., *energetic, active, delighted*), we retained the five PANAS-C/P PA items noted above (i.e., *joyful, cheerful, happy, lively, proud*) for the sake of item-consistency across both child and parent versions of the PANAS-C and to reduce both scales by at least 50% to increase measurement efficiency. These five items were also associated with greater normalized TICs (see Figure 1) as well as acceptable location parameter estimates, ranging from -2.70 to 0.81 for the PANAS-C PA scores and -3.18 to 1.28 for the PANAS-C-P PA scores. We discarded the remaining seven items, including “*calm*”, which as predicted, was associated with

**Table 3** Item Response Theory Item Parameters for the PANAS-C PA and NA Items (school sample)

Items	Slope ( $\alpha$ )		Location Parameters ( $\beta$ ) and Standard Errors							
	Slope ( $\alpha$ )	S.E.	1	S.E.	2	S.E.	3	S.E.	4	S.E.
PA items										
Joyful	2.92	.18	-2.03	.35	-1.04	.20	-0.10	.15	0.81	.18
Cheerful	2.78	.18	-2.12	.35	-1.06	.20	-0.15	.14	0.76	.17
Happy	2.52	.18	-2.70	.45	-1.80	.27	-0.76	.16	0.22	.13
Lively	2.42	.18	-2.03	.29	-1.12	.18	-0.19	.13	0.79	.16
Proud	1.99	.13	-1.94	.21	-0.93	.14	-0.01	.11	0.81	.13
Energetic	1.85	.14	-2.29	.23	-1.23	.15	-0.34	.11	0.56	.12
Delighted	1.84	.15	-1.93	.20	-1.03	.14	0.19	.11	1.22	.15
Excited	1.75	.13	-2.62	.26	-1.61	.16	-0.49	.11	0.59	.11
Active	1.58	.12	-2.79	.25	-1.49	.14	-0.40	.10	0.56	.11
Strong	1.36	.12	-2.64	.19	-1.20	.11	0.04	.09	1.25	.12
Interested	0.97	.10	-4.13	.23	-1.73	.11	0.16	.08	2.29	.12
Calm	0.81	.11	-3.75	.17	-1.89	.10	0.05	.08	1.72	.10
NA items										
Miserable	2.33	.19	0.13	.13	0.88	.16	1.43	.21	2.36	.34
Blue	2.19	.15	-0.13	.12	0.84	.15	1.63	.21	2.36	.31
Afraid	2.02	.13	-0.31	.12	0.92	.14	1.61	.18	2.41	.28
Scared	1.95	.13	-0.34	.12	0.76	.13	1.49	.17	2.35	.26
Sad	1.93	.11	-0.86	.13	0.42	.12	1.18	.14	2.30	.24
Mad	1.89	.13	-0.84	.13	0.16	.11	1.06	.14	1.85	.19
Gloomy	1.88	.13	-0.21	.11	0.81	.13	1.61	.17	2.49	.27
Upset	1.77	.12	-1.12	.13	0.13	.11	0.87	.12	2.00	.19
Lonely	1.69	.11	-0.27	.11	0.63	.11	1.36	.14	2.40	.22
Ashamed	1.66	.11	-0.22	.11	0.84	.12	1.77	.16	2.71	.25
Frightened	1.51	.10	-0.22	.10	1.01	.12	1.89	.16	3.25	.30
Disgusted	1.47	.09	-0.27	.10	0.86	.11	1.78	.14	2.58	.20
Guilty	1.32	.09	-0.17	.09	0.96	.10	1.89	.14	3.26	.24
Nervous	1.15	.09	-1.37	.11	0.20	.09	1.34	.11	2.75	.17
Jittery	0.60	.07	-0.88	.08	1.36	.08	3.46	.11	5.42	.18

**Table 4** Item Response Theory Item Parameters for the PANAS-C-P PA and NA Items (school sample)

Items	Location Parameters ( $\beta$ ) and Standard Errors		Location Parameters ( $\beta$ ) and Standard Errors							
	Slope ( $\alpha$ )	S.E.	1	S.E.	2	S.E.	3	S.E.	4	S.E.
PA items										
Joyful	4.01	.41	-1.99	.32	-1.29	.36	-0.23	.45	1.02	.99
Cheerful	3.67	.27	-2.5	.14	-1.39	.14	-0.31	.18	0.89	.93
Happy	3.56	.35	-2.33	.51	-1.32	.38	-0.23	.30	0.94	1.00
Lively	2.96	.41	-3.18	.29	-2.13	.27	-0.65	.30	0.81	1.20
Proud	2.48	.24	-1.99	.88	-1.19	.64	-0.03	.43	1.28	.60
Energetic	2.32	.13	-2.06	.36	-1.23	.33	-0.07	.35	1.05	.37
Delighted	2.17	.21	-2.51	.25	-1.47	.27	-0.23	.36	1.04	.45
Excited	2.14	.27	-2.64	.64	-1.64	.54	-0.55	.60	0.78	.49
Active	1.71	.57	-2.51	.27	-1.39	.24	-0.05	.28	1.31	.36
Strong	1.71	.19	-2.56	.59	-1.35	.44	0.02	.50	1.44	.33
Interested	1.51	.77	-3.63	-	-1.89	-	-0.31	-	1.37	.51
Calm	0.68	.37	-4.88	.34	-2.26	.33	0.57	.38	3.14	.23
NA items										
Miserable	2.96	.30	0.82	.27	1.58	.41	2.26	.59	3.12	.92
Blue	2.89	.30	0.68	.24	1.62	.41	2.36	.63	2.83	.78
Afraid	2.49	.20	0.46	.18	1.72	.32	2.37	.45	3.27	.74
Scared	2.44	.22	0.46	.17	1.62	.30	2.44	.46	3.19	.70
Sad	2.23	.16	0.10	.15	1.48	.23	2.33	.35	3.11	.55
Mad	2.17	.19	0.83	.19	1.74	.28	2.54	.42	3.25	.64
Gloomy	2.13	.16	0.84	.18	2.00	.29	2.77	.43	3.77	.78
Upset	2.08	.17	0.67	.17	1.62	.24	2.29	.33	3.23	.55
Lonely	2.00	.16	0.81	.17	1.77	.25	2.55	.36	3.39	.57
Ashamed	1.90	.15	0.59	.15	1.61	.22	2.35	.30	3.33	.51
Frightened	1.84	.14	-0.54	.14	0.67	.14	1.62	.20	2.56	.32
Disgusted	1.78	.13	0.59	.14	1.91	.22	2.74	.33	3.78	.60
Guilty	1.71	.14	-0.48	.13	0.80	.14	1.70	.20	2.57	.29
Nervous	1.62	.12	-0.05	.12	1.38	.16	2.51	.26	3.65	.48
Jittery	1.50	.12	1.19	.15	2.20	.21	3.15	.33	3.75	.44

the lowest discrimination parameters for both the PANAS-C ( $\alpha=0.81$ ) and PANAS-C-P ( $\alpha=0.68$ ) and location parameter estimates that suggested discrimination for youths at extreme locations on the latent trait (e.g., being estimated at 4.88 standard deviations below the mean of positive affect to endorse the first category of the “*calm*” item on the PANAS-C-P).

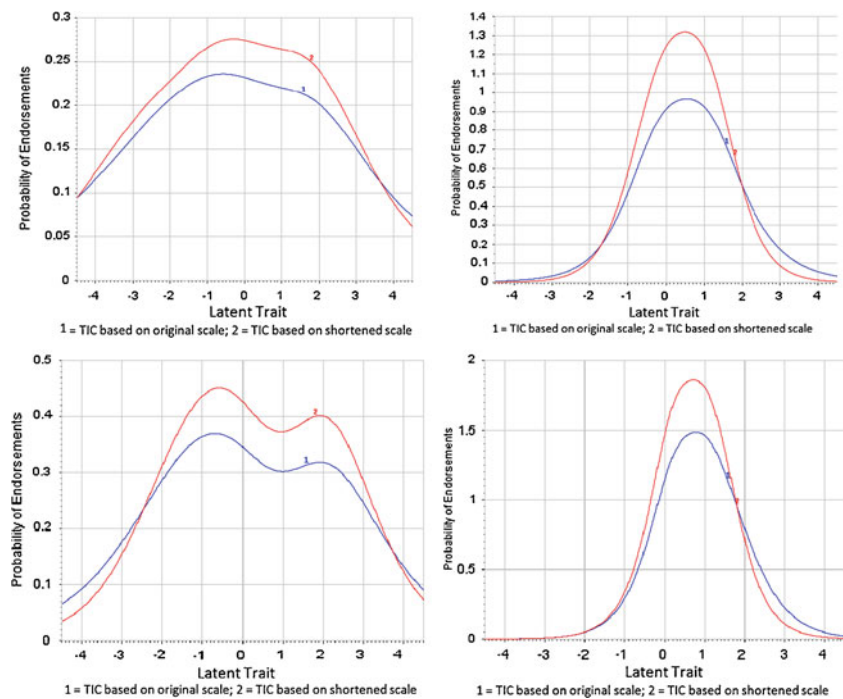
Based on the item selection criteria noted above, we also identified five NA items across both the PANAS-C and the PANAS-C-P to comprise the reduced NA scale (i.e., *miserable, mad, afraid, scared, sad*). These items were associated with high discrimination parameters, reduced the scale by at least 50%, and adequately tapped the recommended domains of “fearful/anxious, sad/depressed, and angry/hostile mood” (Watson and Clark 1997, p. 294). As with the PA items, there were also a few other (parent) NA items with high discrimination properties (e.g., gloomy, upset, lonely); however, we did not include these items in the reduced

version. The retained five NA items (i.e., *miserable, mad, afraid, scared, sad*) were also associated with greater normalized TICs (see Figure 1) as well as acceptable location parameter estimates ranging from  $-0.86$  to  $2.41$  for the PANAS-C NA scores and  $0.10$  to  $3.27$  for the PANAS-C-P NA scores. We discarded the remaining ten NA items, including “*jittery*”, which as predicted, was associated with the lowest discrimination parameters for both the PANAS-C ( $\alpha=0.60$ ) and PANAS-C-P ( $\alpha=1.50$ ) and location parameter estimates that suggested discrimination for youths at extreme locations on the latent trait (e.g.,  $5.42$  standard deviations above the mean of negative affect to endorse the last category of this item on the PANAS-C).

**Cronbach Alpha Coefficients** Cronbach’s alpha coefficients for the reduced 5-item PA scale and original 12-item PA scale of the PANAS-C (*child version*) were .86



**Fig. 1** Normalized Test Information Curves for the PANAS-C and PANAS-C-P original and shortened NA and PA scales, based on the development school-sample



and .89, respectively. Cronbach's alpha coefficients for the reduced 5-item PA scale and original 12-item PA scale of the PANAS-C-P (*parent version*) were .85 and .88, respectively. As predicted, the original scales' coefficient alpha estimates were larger; however, as predicted, the reduced PA scales met the cut-off for "good" internal consistency (Cronbach's alpha  $>.80$ ; Nunnally and Bernstein 1994). Further, given that the average inter-item correlations for the original scales were smaller than the reduced scales (as shown below), we can deduce that the original scales' coefficient alpha estimates were larger relative to the reduced scales simply due to having a greater number of items. This is because coefficient alpha is a function of two parameters: (a) the number of items and (b) the average inter-item correlation (AIC).

Cronbach's alpha coefficients for the shortened 5-item NA scale and original 15-item NA scale of the PANAS-C (*child version*) were .82 and .90, respectively. Cronbach's alpha coefficients for the reduced 5-item NA scale and original 15-item NA scale of the PANAS-C-P (*parent version*) were .83 and .93, respectively. Again, as predicted, the original NA scales evidenced higher alpha coefficients; however all internal consistency estimates met the aforementioned benchmark for good internal consistency. Together, both sets of results for the NA and PA scales demonstrated acceptable internal consistency estimates for the shortened NA and PA scales.

**Inter-item correlations** The average inter-item correlations for the reduced 5-item PA scale and original 12-item PA

scale were .55 and .39, respectively, for the PANAS-C (*child version*), and .63 and .48, respectively, for the PANAS-C-P (*parent version*) PA scales. The average inter-item correlations for the reduced 5-item NA scale and original 15-item NA scale were .47 and .37, respectively, for the PANAS-C (*child version*), and .50 and .46, respectively, for the PANAS-C-P (*parent version*) NA scales. These results suggest that the items comprising the reduced NA and PA scales are more inter-related and cover a somewhat less broad range of content domains relative to the original scales. For example, the items of the reduced PA scale primarily tap content domains specific to happiness and liveliness, while the original PA scales tap other content domains such as "strong" and "delighted."

**Divergent Validity** Based on the tripartite model and work with adult populations, researchers have contended that NA and PA should be uncorrelated (Clark and Watson 1991). Some work conducted among youth however have found NA and PA to be slightly negatively correlated, ranging from  $-.16$  (Laurent et al. 1999) to  $-.29$  (Chorpita and Daleiden 2002). We thus examined the correlations between the NA and PA scale from the PANAS-C (*child*) and PANAS-C-P (*parent*) using both the original and reduced versions. For the shortened scales to retain their discriminant properties, the correlations between the shortened NA and PA scale scores should still be only slightly (negatively) correlated at most (i.e., not greater than the correlations noted above based on previous studies with children), and should also not be significantly (negatively) greater than the (negative) correlations between the original NA and PA

scale scores. In our present sample, the correlation between the PANAS-C NA and PA (original) parent scales was  $-.13$  ( $p < .01$ ), and the correlation between the PANAS-C NA and PA (reduced) scales was  $-.14$  ( $p < .01$ ). Fisher's z-test for independent correlations revealed that these correlations were not significantly different,  $z = 0.20$ ,  $p = .84$ . The correlation between the PANAS-C NA and PA (original) child scales was  $-.06$  (ns), and the correlation between the PANAS-C NA and PA (reduced) scales was  $-.13$  ( $p < .01$ ). Fisher's z-test for independent correlations also revealed that these correlations were not significantly different,  $z = 1.41$ ,  $p = .16$ .

### Scale Validation

**Test information based on the validation sample** The normalized Test Information Curves (TICs) for the PA and NA scales (comparing the original and shortened scales for the PANAS-C) based on the clinic-referred validation sample, appear in Figure 2. These normalized TIC results visually depict the degree to which the shortened PA and NA scales identified in the development sample provides more *test information* (per test item) relative to the original scales in this separate (clinic-referred) validation sample. As expected, the items comprising the reduced PANAS-C NA and PA scales were associated with substantially greater TICs relative to the discarded items.

**ANOVA and ROC analyses** Results of the ANOVA and ROC analyses evaluating the correspondence of the short

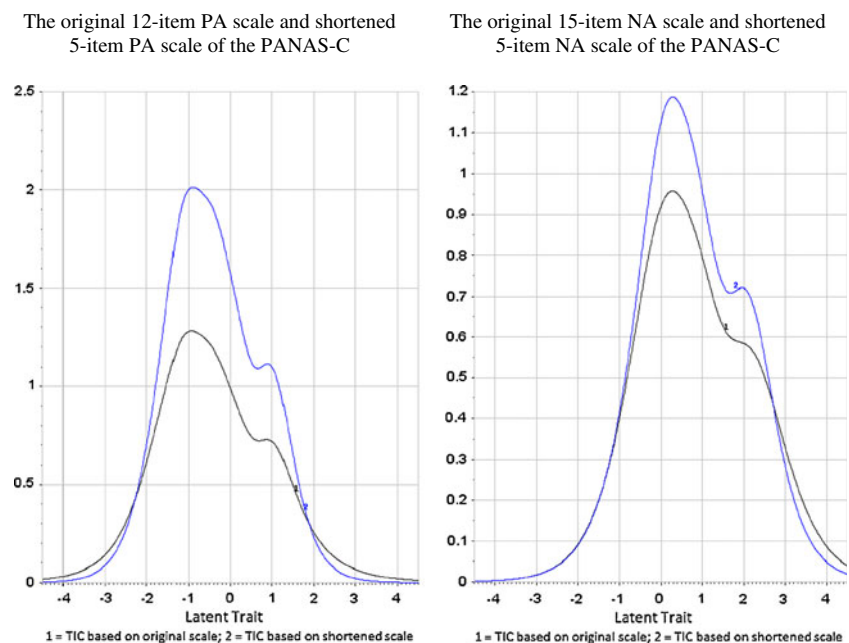
and original NA and PA scales with relevant diagnostic groups based on the validation (clinic-referred) sample appears in Table 5.

**The NA Scale: Discriminating youth with anxiety and/or mood disorders from youth without any anxiety or mood disorders** As predicted, youths with anxiety and/or mood disorders scored significantly higher on both the short and original versions of the PANAS-C NA scales compared to youths without any anxiety or mood disorders (Table 5, first set of rows). Based on ROC analysis, classification accuracy of both the short and original NA scales fell in "fair" range (i.e.,  $AUC > .70$  for both scales) and their AUC values did not significantly differ.

**The PA Scale: Discriminating youth with mood disorders from youth with no mood disorder** As expected, the original PANAS-C 12-item PA scale and shortened 5-item PA scale were both able to discriminate youths with *mood disorders* from youths with no mood disorder, as evidenced by significant ANOVAs and AUC values significantly greater than chance level (Table 5, second set of rows). Classification accuracy of both short and original PA scales also fell in "fair" range (i.e.,  $AUC > .70$ ) and did not significantly differ.

**The PA Scale: Discriminating youth with mood disorders from youth with an externalizing disorder and no mood disorder** As predicted, the original PANAS-C 12-item PA scale and the shortened 5-item PA scale were both able to discriminate youths with *mood disorders* from youths with an externalizing disorder and no mood

**Fig. 2** Normalized Test Information Curves for the PANAS-C original and shortened NA and PA scales, based on the validation clinical-sample



**Table 5** ANOVA and ROC results for the original and reduced PANAS-C PA and NA scales for each diagnostic group

PANAS-C	Diagnostic Group	ANOVA						ROC			
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>	<i>Partial</i> $\eta^2$	AUC	<i>SE</i>	<i>z</i>	<i>p</i>
NA short	Anxiety/Depression	276	12.6	4.9							
	No Anxiety/No Depression	386	9.0	4.2	103.6	<.001	.136	.73	.020		
NA original	Anxiety/Depression	276	36.6	13.0							
	No Anxiety/No Depression	386	26.7	11.0	110.6	<.001	.144	.72	.020	0.68	.50
PA short	Depression	65	11.3	5.5							
	No Depression	597	16.9	5.5	61.8	<.001	.086	.77	.032		
PA original	Depression	65	28.8	11.3							
	No Depression	597	39.5	11.2	54.5	<.001	.076	.76	.033	1.49	.14
PA short	Depression	65	11.3	5.5							
	Ext. Dx (No Depression)	197	16.9	5.3	53.2	<.001	.170	.77	.035		
PA original	Depression	65	28.8	11.3							
	Ext. Dx (No Depression)	197	39.5	10.9	46.3	<.001	.151	.76	.036	1.16	.25
PA short	Depression	65	11.3	5.5							
	Anxiety Dx (No Depression)	211	17.0	5.1	60.7	<.001	.181	.78	.035		
PA original	Depression	65	28.8	11.3							
	Anxiety Dx (No Depression)	211	39.3	10.1	50.5	<.001	.156	.76	.037	1.93	.05

ROC Receiver Operating Characteristic; AUC Area Under the Curve. Note. PANAS-C Positive and Negative Affect Schedule for Children; **Anxiety** included youths with separation anxiety disorder, generalized anxiety disorder, specific phobia, obsessive-compulsive disorder, post traumatic stress disorder, panic disorder, social phobia, and/or anxiety disorder not otherwise specified; **Depression** included youths with major depressive disorder, dysthymic disorder or mood disorder not otherwise specified; **Ext. Dx**=Externalizing disorders, which included youths with any ADHD disorder, oppositional defiant disorder, conduct disorder or disruptive behavior disorder not otherwise specified

disorder, as evidenced by significant ANOVAs and AUC values significantly greater than chance level (Table 5, third set of rows). Classification accuracy of both the short and original PA scales for this diagnostic classification also fell in “fair” range and did not significantly differ from each other.

*The PA Scale: Discriminating youth with mood disorders from youth with an anxiety disorder and no mood disorder* As predicted, the original PANAS-C 12-item PA scale and the shortened 5-item PA scale were both able to discriminate youths with *mood disorders* from youths with an anxiety disorder and no mood disorder, as evidenced by significant ANOVAs and AUC values significantly greater than chance level (Table 5, fourth set of rows). Classification accuracy of both the short and original PA scales for this diagnostic classification also fell in “fair” range (i.e., AUC > .70) and did not significantly differ from each other.

Together, these ANOVA and ROC analyses comparing the classification accuracy of both the original and shortened NA and PA scales revealed that the assessment precision and clinical utility of the PANAS-C NA and PA shortened scales were not compromised with respect to being able to aid in these clinically-relevant differential diagnostic classifications relative to the original scales.

**Discussion**

Through applying IRT-based methodology in the present study, we simultaneously developed a shorter 10-item version of the PANAS-C and PANAS-C-P (including a 5-item NA scale and 5-item PA scale) that can assess NA and PA more efficiently (due to shorter scale length) with refined psychometric properties. We discarded items with relatively weaker properties based on IRT analysis. For example, recent studies identified questionable psychometric properties of the “calm” and “jittery” items (Ebesutani et al. 2011a; b), and as predicted, these items were discarded in the present study due to evidencing relatively weaker discrimination and item information parameters. Using a separate clinic-referred sample of youth, we also demonstrated that the reduced 5-item PA scale and 5-item NA scale were able to classify and differentiate youths with relevant clinical disorders apart from youths with non-targeted emotional and behavioral problems just as well as the original scales.

Implications for Research, Policy, and Practice

The classification tests conducted in the present study importantly demonstrated that the short versions of the NA and PA scales still provided meaningful information to aid

in the differential diagnosis process involving internalizing and externalizing disorders and may thus be used in school or clinic settings for such identification purposes. It is worth noting however that the items of these reduced NA and PA scales tap a somewhat less broad range of negative and positive affective markers as compared to the original PANAS-C (child version; Laurent et al. 1999) and the adult PANAS (Watson et al. 1988). For example, although the reduced NA scale does include items that tap the recommended NA-related domains of “fearful/anxious, sad/depressed, and angry/hostile mood” (Watson and Clark 1997, p. 294), the original NA scales do include items that tap additional NA-related domains such as feeling guilty and ashamed. The extent to which the removal of these items from the NA scale affect the ability of the reduced scales to identify relevant youth subgroups should be further studied. Nonetheless, the reduced NA and PA scales have demonstrated adequate performance in the present clinical sample with respect to differentiating meaningful subsamples of disordered youth.

In addition to shortening the PANAS-C and PANAS-C-P PA scales for more efficient assessment, the present study also serves as another successful demonstration of applying IRT models to psychological assessment instruments. Researchers and clinicians looking for the most efficient assessment instruments to utilize should be aware of the role that IRT can play in further refining and shortening test instruments. Although IRT models may not be applicable to every psychological test instruments (e.g., due to having to meet the unidimensionality criterion), scales that have been refined and developed via IRT-based methodology may provide more efficient options when selecting scales for research and/or clinical practice. Future assessment-related research efforts should thus also strongly consider incorporating IRT-based methodology (e.g., examining IRT sloping parameters, computing and comparing “Information Curves”) when developing and/or refining scales, particularly given the noted advantages of IRT over classical test theory (see Reise and Henson 2003).

There were also limitations to the present study that have implications for future research. First, the lack of PANAS-C-P (parent version) data in the clinic-referred sample precluded our ability to test the degree to which the PANAS-C-P reduced NA and PA scales were able to differentiate relevant diagnostic groups. It is thus recommended that future research examine the discriminant properties of the shortened PANAS-C-P PA and NA scales (parent version) using clinical samples. The present sample was also comprised of a Hawaii-based sample, which may limit the generalizability of the present findings. Although the present sample did include an ethnically diverse group of youths and parents, there were, for example, few African American

and Hispanic participants. The present findings may thus not generalize as well to populations and regions with youths less represented in our sample. More research is thus needed to determine the degree to which our findings remain applicable to other settings and ethnic groups. It is also important to note that Emons et al. (2007) cautioned against the use of short tests, particularly when stakes are high. Although we did find comparable classification accuracy rates between the original and shortened versions of the NA and PA scales with respect to identifying youths with related clinical diagnoses (e.g., depression, anxiety), more research is needed to more fully explore the limitations associated with these shortened scales, such as when measurement error approaches unacceptable levels to discourage their use in any contexts. It is also worth noting that the reduced NA and PA scales included items targeting happiness and sadness—items that are sometimes considered to be problematic when assessing NA and PA given that these types of items tend to be negatively correlated (see Green et al. 1993; Watson and Clark 1997) thereby reducing the discriminant validity of the NA and PA scale scores. Although the reduced NA and PA scales were not significantly more negatively correlated compared to the original NA and PA scales, the effects of retaining happiness- and sadness-related items on measuring NA and PA should continue to be assessed in future research.

Although Emons and colleagues’ (2007) statement of caution is important and should guide decisions made by both test developers (e.g., researchers) and test administrators (e.g., clinicians, school officials), it is important to note that if scales are to include a sufficient number of items to reach “acceptable” levels of measurement error and classification accuracy rates, a test may become too long to be transportable and adopted by “real-world” settings, such as community mental health clinics and schools. If this is the case, a test may be highly accurate and precise, but in the end, may not be utilized due to its length. As shortened versions of scales are developed, it is this tension between scale length and assessment accuracy that should be balanced if instruments are to be used in “real-world” settings. All things being equal, a shortened scale will be less reliable than the longer version; however, a shorter scale will be more easily transportable, adopted and utilized in “real-world” settings. As researchers and test developers continue to conduct research on instruments, an important question is how do we develop instruments that (a) produce reliable and valid scores, yet (b) are short enough so that they are adopted and utilized in “real-world” settings. IRT may be one statistical tool to aid in this process.

In summary, the present study developed and found support for using the reduced 10-item PANAS-C and PANAS-C-P for both research and clinical purposes. The 10-item PANAS-C and PANAS-C-P NA and PA scales both

met benchmark for acceptable internal consistency despite their shortened length, and they also contained more than 50% fewer items relative to the original PANAS-C and PANAS-C-P scales. The PANAS-C child version also performed just as well as the original 27-item PANAS-C with respect to classifying anxious and depressed youth in a clinical youth sample. Researchers and practitioners should thus consider using these reduced 10-item PANAS-C and PANAS-C-P instruments developed in the present study when seeking to measure NA and PA in youth, particularly when time and battery length are of concern.

## References

- Analyse-it Software, Ltd. (2008). Analyse-it for Microsoft Excel (version 2.12).
- Baker F (2001). The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland College Park, MD.
- Bagozzi, R. (1993). An examination of the psychometric properties of measures of negative affect in the PANAS-X scales. *Journal of Personality and Social Psychology*, *65*, 836–851.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Chorpita, B. F., & Daleiden, E. L. (2002). Tripartite dimensions of emotion in a child clinical sample: Measurement strategies and implications for clinical utility. *Journal of Consulting and Clinical Psychology*, *70*, 1150–1160.
- Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science*, *5*, 8–14.
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316–336.
- DeLong, E., DeLong, D., & Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*, 837–845.
- Ebesutani, C., Okamura, K., Higa-McMillan, C., & Chorpita, B. F. (2011a). A Psychometric analysis of the Positive and Negative Affect Schedule for Children – Parent Version in a school sample. *Psychological Assessment*.
- Ebesutani, C., Smith, A., Bernstein, A., Chorpita, B. F., Higa-McMillan, C., & Nakamura, B. J. (2011b). A bifactor model of negative emotion: Negative affectivity comprised of fear and distress among older youth. *Psychological Assessment*.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Emons, W., Sijtsma, K., & Meijer, R. (2007). On the Consistency of Individual Classification Using Short Scales. *Psychological Methods*, *12*, 105–120.
- Ferdinand, R. (2008). Validity of the CBCL/YSR DSM-IV scales Anxiety Problems and Affective Problems. *Journal of Anxiety Disorders*, *22*, 126–134.
- Forbes, E., & Dahl, R. (2005). Neural systems of positive affect: Relevance to understanding child and adolescent depression? *Development and Psychopathology*, *17*, 827–850.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bifactor analysis. *Psychometrika*, *57*, 423–436.
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, *64*, 1029–1041.
- Hughes, A. A., & Kendall, P. C. (2009). Psychometric properties of the Positive and Negative Affect Scale for Children (PANAS-C) in children with anxiety disorders. *Child Psychiatry and Human Development*, *40*, 343–352.
- Jacques, H., & Mash, E. (2004). A test of the tripartite model of anxiety and depression in elementary and high school boys and girls. *Journal of Abnormal Child Psychology*, *32*, 13–25.
- Laurent, J., Catanzaro, S. J., Joiner, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., et al. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment*, *11*, 326–338.
- Lyneham, H. J., Abbott, M. J., & Rapee, R. M. (2007). Interrater reliability of the Anxiety Disorders Interview Schedule for DSM-IV: child and parent version. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*, 731–736.
- Multivariate Software, Inc. (2010). *mvIRT – A User-Friendly IRT Program*. Encino, CA: Multivariate Software, Inc.
- Muthén, B., & Muthén, L. (2007). *Mplus 4.21*. Los Angeles, CA: Muthén & Muthén.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- Reise, S. (2009). Bifactor and Item Response Theory Analyses of Interviewer Report Scales of Cognitive Functioning in Schizophrenia. Submitted for publication.
- Reise, S., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, *81*, 93–103.
- Reise, S., Moore, T., & Haviland, M. (2010). Bifactor Models and Rotations: Exploring the Extent to which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment*, *92*, 544–559.
- Reise, S., & Waller, N. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48.
- Richardson, L. M., & Austad, C. S. (1991). Realities of mental health practice in managed-care settings. *Professional Psychology, Research and Practice*, *22*, 52–59.
- Samejima F (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17.
- Silverman, W. K., & Albano, A. M. (1996). *Anxiety Disorders Interview Schedule for DSM-IV, Child and Parent Versions*. San Antonio, TX: Psychological Corporation.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293–325.
- Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment*, *68*, 267–296.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.
- Wood, J. J., Piacentini, J. C., Bergman, R. L., McCracken, J., & Barrios, V. (2002). Concurrent validity of the anxiety disorders section of the Anxiety Disorders Interview Schedule for DSM-IV: child and parent versions. *Journal of Clinical Child and Adolescent Psychology*, *31*, 335–342.