Published in final edited form as:

Autism. 2016 January ; 20(1): 14–25. doi:10.1177/1362361314564473.

Evaluating the appropriateness of a new computer-administered measure of adaptive function for children and youth with Autism Spectrum Disorders (ASDs)

Wendy J. Coster, PhD, OTR/L *,

Professor, Department of Occupational Therapy and PhD Program in Rehabilitation Sciences, Boston University

Jessica M. Kramer, PhD, OTR/L,

Assistant Professor, Department of Occupational Therapy and PhD Program in Rehabilitation Sciences, Boston University

Feng Tian, PhD,

Health & Disability Research Institute, School of Public Health, Boston University

Meghan Dooley, MSOT, OTR/L,

Department of Occupational Therapy, Boston University

Kendra Liljenquist, BS,

PhD Program in Rehabilitation Sciences, Boston University

Ying-Chia Kao, ScD, OTR, and

Department of Occupational Therapy, Boston University

Pengsheng Ni, MD

Senior Data Analyst/Research Assistant, Health & Disability Research Institute; Professor, School of Public Health, Boston University

Introduction

The growing number of youth with autism spectrum disorders (ASD) underscores the importance of efficient, sensitive, and well-targeted instruments to measure outcomes for this population. Measures are needed that can document variations in capacities and limitations that may distinguish meaningful subgroups, to examine the associations among symptoms, impairments, and function, and to evaluate the effectiveness of interventions for improving long-term outcomes. The marked heterogeneity of the population argues for the importance of conducting research with a variety of measures since some may be better suited to certain research questions or clinical purposes than others.

One important focus of research in ASD is the effectiveness of interventions to improve functional and adaptive outcomes, that is, the capacity of the person to perform daily life

*Corresponding author: Wendy J. Coster, PhD, OTR/L, Professor, Department of Occupational Therapy, College of Health and Rehabilitation Sciences: Sargent College, Boston University, 635 Commonwealth Ave (SAR 503), Boston, MA 02215, wjcoster@bu.edu, Phone: 617-353-7518, Fax: (617)-353-2926.

skills that are expected of similar-aged peers including practical skills, social skills, and conceptual skills (Schalock, et al. 2010) Achievement of these skills is associated with the eventual ability of the individual with ASD to transition to adult roles, including independent living, employment, and community participation (Esbensen, et al. 2010, Farley, et al. 2009, Taylor and Seltzer, 2011). Although a growing body of research highlights the significant functional difficulties and poor adult outcomes experienced by individuals on the autism spectrum (Carter et al. 1998; Howlin et al. 2013; Newman, et al., 2009) investigations of the relation between underlying impairments and symptomology and functional performance have yielded inconsistent findings. For example, studies examining the relationship between autism symptom severity and adaptive behavior have resulted in inconsistent findings. One study found severity of autism symptoms is negatively correlated with overall adaptive behavior (Perry, et al. 2009) while another reported that symptoms and behaviors were only correlated for children classified as high functioning autism (Liss, et al. 2001). Findings from research exploring the relation between measures of cognitive functioning and measures of adaptive function also have varied. Adaptive behavior was predicted by IQ levels for children with ASD classified as “low functioning”, while language and memory predicted adaptive behavior scores for children classified as “higher functioning” (Liss, et al. 2001). However, in another study, all children with ASD had adaptive behavior skills that were significantly lower than IQ (Gabriels, et al. 2007). Finally, studies have also reported varied findings across domains of adaptive behavior, particularly whether domains other than socialization are impaired (Liss, et al. 2001; Stone, et al. 1999; Carter, et al., 1998).

The challenges and inconsistencies encountered when attempting to understand the mechanisms underlying successful life outcomes for individuals with ASDs and the interaction between symptom severity, underlying impairments, and functional performance suggest that the field may benefit from other approaches to measuring functional performance of everyday activities. To date most investigations have used instruments developed within the framework of adaptive behavior in the field of intellectual disabilities (Schalock et al., 2010), in particular the Vineland Adaptive Behavior Scales (VABS-II) (Sparrow, et al. 2005) and the Adaptive Behavior Assessment System (ABAS-2) (Harrison and Oakland, 2003). This conceptualization of adaptive behavior identifies two general aspects, personal independence and social responsibility, with three clusters of domains: conceptual, social, and practical. Within adaptive behavior instruments, the domains and subdomains have generally followed the categories defined by the AAIDD (Tassé et al., 2012), including communication, community use, functional academics, independent living and daily living skills, health and safety, and leisure. The items in each category address skills associated with that aspect of daily life. However, a previous content analysis of the VABS-II (Gleason and Coster, 2012) suggested that social and communication impairments, in particular, might influence ratings on items, and therefore summary scores, in a variety of subdomains not explicitly seeking to measure communication, including the VABS-II Daily Activities subdomain. Alternative approaches to defining domains, in which items measure skills that depend primarily on a single underlying ability (e.g., motor coordination; communication) might improve our ability to find patterns and relationships between underlying symptomology, functional skills, and life outcomes.

The World Health Organization (WHO) has adopted an alternative framework for describing function in daily life. The International Classification of Functioning, Disability, and Health (ICF) (World Health Organization, 2001) and the Children and Youth version (ICF-CY) (World Health Organization, 2008) were designed to provide a universal language for describing health, disability, and function across populations. The ICF-CY distinguishes several components of functioning and disability. “Body structures and functions” include foundational psychological and intellectual functions such as adaptability, attention and orientation, and range and regulation of emotion, which are typically impaired in children and youth with ASD. Impairments in body structures and functions may lead to difficulties with ‘Activity’, which is defined as the execution of a task or skill. Difficulties in activities are represented as functional limitations and can be described in terms of capacity (ability to perform in a standardized environment) or performance (typical behavior in one’s own usual environment). The broader ICF component of ‘Participation’ describes the person’s involvement in life situations and reflects the person’s engagement in culturally relevant social roles and settings. This framework distinguishes underlying impairment from the person’s application of their abilities, as demonstrated by functional skills and participation, and provides a guiding structure to conceptualize measurement of each component. Nine chapters define the domains of activity and participation: Learning and applying knowledge; general tasks and demands; communication; mobility; self-care; domestic life; interpersonal interactions and relationships; major life areas; and community, social, and civic life. The ICF framework also incorporates the influence of contextual factors (personal and environmental) on function. This framework has increasingly been utilized as a conceptual framework for outcomes measurement in health and disability research for a variety of clinical populations, including youth with developmental disabilities and autism (Bonanni, et al., 2009; Castro, et al. 2013; Poon, 2011).

The ICF’s emphasis on universal versus condition-specific description of health and function, along with growing adoption of modern test development approaches, has stimulated research into the extent to which instruments used to measure function can be applied across clinical populations. One focus of this research has been whether the underlying structure of function – i.e., the relations among different dimensions of performance in a given domain or the relative level of ability required to perform a particular set of tasks remains the same across populations, a property known as measurement invariance (Alguren, et al., 2011; Kim, et al., 2013). If not, then these differences pose a potentially significant threat to valid interpretation of scores since scores from the same instrument may represent different underlying patterns of function in different groups. While measurement invariance has been investigated for the most frequently used measures of adaptive behavior, these investigations have been limited to subgroups defined by age, gender, and socio-economic status (Sparrow et al., 2005; Harrison, and Oakland, 2003). Comparable investigations of measurement invariance of these instruments across clinical populations are not yet common (although see Frazier, et al. 2014 for a recent example).

This paper reports results from such an investigation, in which we evaluate the structural validity of a new instrument developed within an ICF framework, the Pediatric Evaluation of Disability Inventory-Computer Adaptive Test (PEDI-CAT) for children and youth with

symptoms of ASD. The PEDI-CAT offers a potential alternative method for describing functional profiles for children and youth with ASDs. However, given that it is based on a different conceptual framework than other measures of adaptive behavior, empirical evidence is needed to verify that the assessment functions as intended in youth with ASDs. It is possible that the unique profile of strengths and limitations that characterize ASDs may be associated with systematic differences that could threaten the validity of assessment scores.

Specific research questions were:

1. Do items in each PEDI-CAT domain co here to represent a unidimensional construct in this population, suggesting the performance of youth with ASDs fits the conceptual model of the instrument?
2. Do all items in each domain demonstrate acceptable fit along a unidimensional continuum, indicating that items reflect a common underlying construct in youth with ASDs?
3. Do any items show differential item functioning between the original PEDI-CAT sample and this population, indicating measurement variance in this clinical population?
4. Are scores generated by the CAT administration similar to scores generated from the full item set?

Methods

Participants

A convenience sample of parents of children and youth ages 3–21 was recruited through local and national service, support, and advocacy groups for children with ASDs via flyers, list-serv emails, and website announcements. Eligibility was determined via email or over the phone by positive responses to the following questions: 1) Has your child been diagnosed with Autism Spectrum Disorder, including Autism, Asperger's syndrome, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS)? 2) Do you and your child currently live in the United States of America? 3) Is your child between the ages of 3 years, 0 months and 21 years, 11 months? 4) Are you the child's legal guardian (even if your child is above the age of 18)? The research team used two additional screening questions to reduce the risk of enrolling respondents purposefully misrepresenting their eligibility for participation, a common risk in internet research (Kramer, et al., 2014).

Instruments

The original PEDI (Haley et al. 1992) is a parent- or clinician-report measure that has been used extensively in rehabilitation and special education programs as well as in clinical research with children with a variety of disabilities including cerebral palsy (Hinderer et al., 1996; Dolva, et al., 2004), Down syndrome (Dolva et al., 2004), osteogenesis imperfecta (Engelbert, et al. 1997), and acquired brain injury (Kothari, et al. 2003). It has been translated into multiple languages and studies have demonstrated its reliability, validity and

responsiveness with various populations (Haley, et al. 2011). Recently, the PEDI was revised to increase the age range covered (up through age 20), revise and update the content and rating scales to make them maximally applicable across the diverse population of children and youth with disabilities, and to develop a computer-based assessment. The ICF-CY (World Health Organization 2008) served as a guiding conceptual framework when developing the PEDI-CAT domains and corresponding rating scales.

The PEDI-CAT assesses four domains of function: Daily Activities, Social/Cognitive, Mobility, and Responsibility. The Daily Activities, Social/Cognitive, and Mobility domains of the PEDI-CAT were designed to measure the performance aspect of the ICF construct of 'activity', i.e., how the child or youth performs the activities in his or her usual daily environment. Items in the Daily Activities domain assess a child's daily living skills such as eating, dressing, and grooming activities. The Daily Activities domain also includes items related to household maintenance and the operation of electronic devices. The Social/Cognitive domain assesses a child's ability to interact with others in a community and to manage functional cognitive activities such as counting out change. Social/Cognitive items address communication, interaction, safety, behavior, play, attention, and problem-solving. The Mobility domain examines the child's ability to perform a variety of functional movements including walking, getting in and out of a chair, climbing stairs, or running. The Daily Activities and Mobility items also include line drawings depicting the activity assessed in each item. In order to maximize item fit to a unidimensional measurement model, items were written as much as feasible to reflect the application of a single primary underlying ability: full body movement in Mobility; use of arms and hands in Daily Activities; cognition for Social/Cognitive. Item descriptions were worded to minimize reference to particular modes of performance and thus credit the capabilities of children and youth with disabilities who may use alternative methods to accomplish daily tasks successfully.

The child or youth's performance of items in the PEDI-CAT Daily Activities, Social/Cognitive, and Mobility domains is rated using a four-point Likert scale measuring the extent to which he or she has difficulty performing each activity: "Unable," "Hard," "A little hard," and "Easy" (see Supplementary Table S1 in Appendix). Parents identify the most appropriate rating by considering their child's typical performance while using usual supports such as alternative communication devices.

The Responsibility domain assesses the extent to which a young person is managing life tasks that are important for the transition to adulthood and independent living (e.g. fixing a meal, planning and following a weekly schedule). Thus, taking responsibility for managing life tasks supports participation in life situations and reflects the "participation" construct within the ICF. This domain contains content assessing health management and literacy, citizenship, safety, and community mobility and, like the activities dimension, reflects the person's current typical behavior in his or her usual environment. A five-point rating scale is used to assess the shift of responsibility for a life task from parents taking all responsibility, to shared responsibility, to the young person taking all responsibility. A 5-point scale was used because the shift of responsibility has a meaningful mid-point where parent and child equally share responsibility for managing a task. The ratings for the Responsibility items do

not require the young person to perform each life task independently, but instead reflect overall independence in managing the task. Independence from this perspective may include requesting specific assistance or resources as needed or directing others (e.g., personal assistant) in order to accomplish the task.

Each domain of the PEDI-CAT can be completed separately. Scores computed include norm-based T-scores (mean of 50, standard deviation of 10), criterion-referenced scores (reported on a 20–80 scale), and percentile ranges. Criterion scores are the preferred scores to detect change over time in a context of overall delay because they capture changes in performance along the overall continuum of function represented by the items.

The PEDI-CAT was standardized with a stratified nationally representative (US) sample of 2205 children and youth without disabilities and 703 children and youth with heterogeneous disabilities ages 0–21, including 108 with ASDs (Haley, et al., 2011). Item parameters obtained from the standardization sample were then used to construct the computer-adaptive test (CAT). Results from a prospective field study of the CAT showed excellent re-test reliability ($r = .95$ for all domains) and differentiation between groups with and without disabilities (Dumas, et al, 2012). Parents took an average of 12 minutes to complete all four PEDI-CAT domains.

Prior to the current study we conducted an extensive qualitative evaluation of the appropriateness of the content and rating scales of the PEDI-CAT for children and youth with ASD (Kramer, et al., 2012). First, a series of focus groups and cognitive interviews were conducted with professionals and parents of children with ASDs to identify whether aspects of the original PEDI-CAT might need modification. We focused on the Daily Activities, Social/Cognitive, and Responsibility domains, as they are most likely to be uniquely impacted by characteristics associated with autism. The findings were used to expand the item pools and modify some aspects of instructions to give directions for rating certain behavior patterns common in ASDs (e.g., inconsistent performance; see details in Supplementary Table S1 in Appendix). The Social/Cognitive domain was expanded with four new items, five items were reinstated after being dropped from the original PEDI-CAT, and 15 items were supplemented with additional directions to guide rating decisions for children with ASDs (see Supplementary Table S2). The Responsibility domain was expanded with six new items, one reinstated item, and two items supplemented with additional directions. The Daily Activities domain was expanded by reinstating eight items dropped from the original PEDI-CAT. This modified PEDI-CAT, which includes the modified instructions and minor item revisions and additions, was used in the present study and for clarity will be referred to as the PEDI-CAT (ASD). One of the advantages of IRT-based scales, as discussed below, is that new items can be added to an existing domain and situated along the same measurement continuum based on findings from the IRT analysis. These additions do not alter the meaning of criterion scores obtained using the original set of items, rather they provide additional items that can be used to estimate the person's location on the functional continuum (Haley, et al., 2009).

The CAT approach to assessment

A computer adaptive test (CAT) (Wainer, et al., 2000) employs a simple form of artificial intelligence that selects questions directly tailored to the person's estimated ability level, shortens or lengthens the test to achieve the desired precision, scores everyone on a standard metric so that results can be compared, and displays results instantly. CAT applications require a large set of items in any one domain, and those items should consistently scale along a unidimensional continuum from low to high functional proficiency.

The item response theory (IRT) measurement approach is used to obtain empirical evidence about the relationships of items within a domain. Item response theory approaches require that items meet two important assumptions: unidimensionality and local independence. Confirmation that these assumptions are met (as determined through IRT analysis) means that items from one domain tap a single common dimension. That is, the performance on different items is not related except through the shared underlying ability or trait (i.e., the continuum of function in that domain of the PEDI-CAT). The data obtained from the research sample are used to locate items along the underlying dimension, thus each item and rating are associated with a specific position on that continuum and a corresponding score. Items that are located below a particular score are likely to be achieved by a respondent with that score, and items that appear above that score represent performance that most likely is still developing. Thus the summary score is not strictly dependent on which items were administered, but rather represents the best estimate of the person's ability (location on the continuum of function) based on his or her performance on the items completed. This unique approach to conceptualizing the measurement of abilities was first applied to large-scale educational assessments and in recent years has been adopted as the method of choice in medical and rehabilitation fields to develop more effective and efficient outcome measures (Cella, et al., 2007; Velozo, et al., 2012). One important feature of IRT-derived scales is that the resulting measurement model is expected to be sample invariant. That is, the item parameters derived from the calibration sample are expected to remain invariant across other samples from the same population. This is a critical difference from measures developed from a Classical Test Theory (CTT) approach, in which the psychometric properties of the instrument are tied to (and reflect) the distribution of the trait or characteristic being measured in the sample used to standardize the instrument (Streiner and Norman, 2008).

Administration of the CAT begins with a global item that is selected a priori on the basis of the range of performance or ability it covers; thus all respondents answer the same first question. Based on the response to the first item, a score and standard error are estimated, then the computer algorithm selects the next optimal item and a response is recorded. With administration of the next item, the score is re-estimated. The computer algorithm determines whether a pre-determined stopping rule has been satisfied based on magnitude of the error estimate. If satisfied, the assessment of that domain ends. If not satisfied, new items are administered in an iterative fashion until the stopping rule is satisfied and the final score is obtained (See Figure 1).

Previous studies, including studies of the PEDI-CAT, suggest that after 15 items a well-designed CAT can provide scores that are as precise as scores obtained using a domain's full item set (Dumas, et al., 2012). CATs have several logistical benefits over paper and pencil

assessments. Reducing the number of items each respondent must complete reduces administration time and respondent burden. CATs can be administered via the internet, at an independent computer station, or on an electronic tablet and scores are generated instantly, thus reducing the personnel resources usually associated with assessment administration. The computer-based administration provides additional benefits such as the ability to include pop-up directions and pictures that help respondents understand the question; such features help ensure a variety of respondents interpret the assessment in a consistent manner.

CATs are well suited to meet the challenges of measuring a population with a very broad range of functional abilities such as children and youth with ASDs. The CAT process of selecting the most relevant items from a pool of possible items enables researchers and clinicians to administer the same assessment to respondents with a range of skills or impairment severity. Essentially, each respondent completes a unique set of items that are targeted to their abilities; however their resulting scores are located on the same overall scale and are highly precise. The location of the score along the continuum of items allows researchers and clinicians to quickly and accurately identify the expected performance on each item of a person with that ability level.

Procedures

Human ethics approval was received from the University Review Board before research activities were conducted. All data collection was completed online using a secure website hosted by a private company. After eligibility was determined, parents received a link and password to access the website. Parents provided informed consent online before proceeding to the online survey. They first completed a demographic questionnaire to gather descriptive information about the children and their family context and the Social Communication Questionnaire (SCQ)-Current. The SCQ Current and Lifetime were used to describe current and past symptom severity. The SCQ (Rutter et al., 2003) is a 40-item parent report used to assess the presence of specific autistic behaviors. Higher scores indicate more severe behaviors. The SCQ has established psychometric properties for persons ages 2–40 years, and a series of studies suggests that the SCQ has acceptable sensitivity and specificity to screen for autism based on a cut-off score of 15 (Berument, et al., 1999; Chandler, et al., 2007; Charman, et al., 2007; Witwer and Lecavalier, 2007). Since our focus was to examine how the PEDI-CAT performs with individuals ranging in functional ability and symptoms, we did not exclude respondents based on SCQ scores.

Parents next answered all items in the Social/Cognitive (68 items), Daily Activities (76 items), and Responsibility (58 items) domains. Parents could exit the survey at any time and return within a 14-day period. After completing the PEDI-CAT items, parents had the option of completing the SCQ-Lifetime.

Data were downloaded from the secure website into SPSS. Duplicate respondents and respondents who did not provide data beyond initial consent were removed from the database.

Analytical procedures

An item response theory (IRT) approach was used to obtain item parameter estimates for each of the domain item pools for youth with ASD. We used a graded response IRT model to obtain item parameters that reflect both item difficulty (location along the underlying continuum) and discrimination (the extent to which each item is sensitive to differences in respondent abilities). These item parameters provide the information for the computer algorithm driving the CAT. In a graded response model analysis, data must meet model assumptions of unidimensionality and fit. The unidimensionality of each domain was evaluated using Confirmatory Factor Analysis and several indexes of fit. Comparative Fit Index (CFI) and Tucker Lewis index (TLI) values range from 0 to 1. Values of 0.90 or higher indicate acceptable fit and values above 0.95 indicate good fit. The Root Mean Square Error of Approximation (RMSEA) was also examined; values less than 0.08 indicate acceptable fit and less than 0.05 indicate good fit. Item parameters, estimated using PARSCALE, were examined for fit to the construct. Likelihood ratio chi-square statistics were used to examine item fit parameters across the distribution of the construct; a p-value less than 0.05 indicted item misfit (Haley et al. 2011)

Next, we used logistic regression to examine differential item function (DIF) to determine whether the item parameters derived from the ASD sample were significantly different from those derived from the standardization sample. DIF occurs when the difficulty of one item relative to another changes for respondents with different characteristics, although the item still fits along the underlying continuum. If DIF is present, the valid interpretation of scores is threatened due to shifts in locations of the items along the continuum. (See Figure 2). The dependent variable in the regression was the response to an item, and the independent variables were participants' total score, group membership (ASD sample or standardization) and an interaction term between the total score and group membership. The standardization sample was a subset of 525 youth without disabilities whose caregivers answered the full set of PEDI items in a particular domain during the original standardization.

The analytic approach was to successively add total score, group membership, and interaction term into the model in three steps, and the procedure was repeated for each item. The test statistic was the $-2\log$ likelihood difference between models, which is distributed as a chi-square with two degrees of freedom, and the effect size was the R^2 change between models. The following criteria were set for DIF analysis: if the likelihood difference test was statistically significant and the R^2 change was greater than .035 for one item, that item exhibited DIF. If substantial DIF was found between samples, item parameters from the ASD sample would need to be adjusted and then linked with those in the existing PEDI-CAT so that domain criterion scores obtained from the two assessment versions remained comparable (T-scores reflect only the model derived from the normative sample and are not adjusted) (Haley, et al., 2009, Embretson and Reise, 2000).

Simulations were also conducted to examine the accuracy and precision of the resulting PEDI-CAT (ASD) software compared to administration of all items in the domain. Simulations were completed to obtain estimated CAT scores for each respondent for each domain. As items were selected for administration in the simulation, responses were taken

from the actual data. For these simulations we established specific stop rules of 5, 10, and 15 items for each domain. This procedure produced one simulated record of responses for each respondent for a 5-, 10-, and 15-item CAT version. Results were evaluated based on strength of correlations between full item set scores and scores obtained from the three CATs. We also examined the percent of subjects with individual score reliability at two levels of confidence: $r > .90$ and $r > .95$.

Results

Children and youth in the sample ($n = 365$) ranged in age from 3 years to 21 years ($M = 11.9$ yrs, $sd = 4.67$ yrs). The majority of respondents were mothers (93.7%) reporting about their male children (83.3%). Parents reported their child's diagnosis as Autism (52.1%), Asperger's (25.5%), and Pervasive Developmental Disorder (22.2%). Twenty-one percent (21.9%) of parents indicated their child had a current or past diagnosis of intellectual disability. Table 1 contains additional demographic information about this sample.

The number of responses available for each domain varied: Social Cognitive ($n = 365$), Daily Activities ($n = 359$), and Responsibility ($n = 356$). Two respondents who were later determined to not meet inclusion criteria were included in the Social/Cognitive analysis and one was included in Daily Activities and Responsibility analyses; analyses remain as reported as IRT methods are robust to such small variations.

Evaluation of Unidimensionality

All three PEDI-CAT (ASD) domains showed good evidence of unidimensionality based on the Confirmatory Factor Analysis. The CFI, TLI, and RMSEA all indicated good fit to a unidimensional model for the Daily Activities and Responsibility domains, and acceptable fit for the Social/Cognitive domain (Table 2).

The domain items also showed good fit to their respective underlying continuum. Only four items across the three domains had poor fit (see supplementary Table S3 in the Appendix). In the Social/Cognitive domain, all revised items and all new and reinstated items except one had acceptable fit in this sample. In the Daily Activities and Responsibility domains all new, revised, and reinstated items had acceptable fit. We used a two-step approach to determine whether any of the misfitting items should be dropped: first, we examined item performance in the standardization sample and second, we examined the extent to which each item assessed a task of unique difficulty on the underlying continuum. Based on this analysis, only one item from the Social/Cognitive scale was removed.

Evaluation of Measurement Invariance

Four Responsibility items, 4 Daily Activities items, and 32 Social/Cognitive items demonstrated differential item functioning (DIF) between the sample without disabilities and the sample with ASDs. One Daily Activity item with poor fit also had DIF. Two Daily Activity items and one Responsibility item with DIF pertained to gender specific self-care activities. Items from the Social/Cognitive domain with DIF represented all four content areas addressed within the domain (communication, interaction, everyday cognition, and

self-management), however the area of everyday cognition had the highest percentage of items with DIF (64%).

Given the very limited DIF in the Daily Activities and Responsibility domains, no modifications to the original CAT parameters were deemed necessary. However the large number of items with DIF in the Social/Cognitive domain required an equating, or linking, procedure to be performed to ensure that criterion scores for youth with ASD would remain a valid representation of youths' functional performance given the unique item difficulty/ordering, and thus remain comparable to the original PEDI-CAT Social/Cognitive criterion scores. We followed recommended procedures for doing so (Embretson and Reise, 2000). First, we obtained item parameter estimates using the responses from the ASD sample. These unique item parameters reflect the relative difficulty of the items to one another along the underlying unidimensional continuum for youth with ASDs. Next, we linked the ASD item parameters to the original PEDI-CAT scale using a statistical transformation (Embretson and Reise, 2000). The result of this transformation is that the ASD items produce criterion scores that are comparable to the PEDI-CAT criterion scores. We used a similar equating technique for several items with missing parameter estimates for the highest rating scale thresholds: one Social/Cognitive item, three Daily Activities items, and two Responsibility items. This was a conservative approach that assumed DIF in the absence of the ability to empirically evaluate DIF.

Real data simulations were then conducted by utilizing the original or linked item estimates as described above in the CAT software. Results showed high correlations ($ICC > 0.95$) between the scores obtained from the full item set and scores from all three CAT options (Table 3). These results are highly similar to results from the original PEDI-CAT simulations (Haley, et al., 2011). Examination of individual score reliability indicated that substantially more subjects had individual score reliabilities at the designated levels with the 10-item and 15-item CATs compared to the 5-item CAT (Table 4).

Discussion

This study provides evidence that criterion scores of children and youth with ASD obtained from administering the PEDI-CAT (ASD) can be compared to those obtained from other groups assessed with the standard PEDI-CAT. That is, domain criterion scores on the PEDI-CAT and PEDI-CAT (ASD) represent the same degree of function in that domain. Good unidimensionality for all three domain item sets, as indicated by CFA results and item fit analyses, suggests that the structure of these scales is appropriate to assess children and youth with a range of autism symptoms who are heterogeneous in age and function. The simulation results indicate that the 15-item PEDI-CAT (ASD) yields a criterion score that is statistically equivalent to administering the full item set for each domain. These results provide evidence that the PEDI-CAT (ASD) can provide an efficient measure of the functional performance of children and youth with ASD ages 3–21.

The items in the Daily Activities and Responsibility domains showed little DIF. These findings suggest that the order of skill acquisition and taking on responsibility for life tasks, as measured by the PEDI-CAT items, is not substantially different for children with

heterogeneous ASDs. Results from a sample of 108 children and youth with ASD included in the development of the original PEDI-CAT found that they performed significantly lower than children without disabilities at the reference ages of 10 and 15 years, but not at 5 years (Kao, et al., 2012). Those findings are consistent with other research reporting that children with ASDs may acquire skills required to complete daily functional tasks at a slower rate than children without disabilities (Fisch, et al., 2002; Liss, et al., 2001). The lack of DIF in these domains in the present study's larger sample provides additional support for the idea that children with ASDs may present with delays but learn to perform increasingly complex daily activities and manage life responsibilities in approximately the same sequence as children and youth without disabilities. These findings provide additional information regarding the development of functional skills in this population and demonstrate the potential usefulness of the PEDI-CAT (ASD) for pursuing related lines of inquiry.

Conversely, the high number of items with DIF from the Social/Cognitive domain suggests that certain skills are uniquely easier or harder for children and youth with ASD to achieve compared to youth without disabilities at the same ability level. This finding was partially anticipated given the communication, interaction, and meta-cognitive difficulties that are associated with autism. DIF can arise from several factors, for example associated impairments that present unique challenges in certain types of activities. Further research is needed to determine what factors may account for the differences found in this study. It is important to note that although the difficulty parameters for the items with DIF suggest a unique influence, these items nevertheless still fit along the same underlying domain continuum, as indicated by item fit and CFA results. In other words, the items could all be located on the same continuum however their location on the continuum – i.e. their relative order of difficulty – was consistently different from that of the standardization sample.

These results should be considered in the context of several limitations. For feasibility reasons, in order to obtain the large sample size required for IRT-based item calibration, we used a convenience sample and relied on parent report measures of cognitive level and symptoms of ASD. Although participants were diverse in geographic location, they were mostly mothers, predominately Caucasian, and with moderate to high family incomes. However the resulting item parameters were compared to those from the nationally representative normative sample and were not substantially different for two of the three domains examined and for just under half of the items in the Social/Cognitive domain. Nevertheless, future studies should examine whether items would function in a similar manner for children with ASDs with other cultural backgrounds and family contexts and to examine the sensitivity and precision of the PEDI-CAT with children and youth with ASDs with more severe impairments.

This sample also included children and youth with SCQ Current scores that were below the published cut score of 15 and we did not have an independent verification of ASD diagnosis. However the present study was designed to norm the instrument or create a diagnostic profile, but rather to test the validity of the IRT measurement model of the original PEDI-CAT for the population with children and youth with symptoms of ASD. For that purpose, heterogeneity in age, a broad range in function, and potential variation in patterns of function is desirable so that results are broadly applicable. If the profiles of subgroups within

our sample were indeed substantially different from one another or the sample included false positives for ASD, this variability would have worked *against* the likelihood of obtaining positive results regarding unidimensionality, including both CFA and item fit, as well as DIF. The IRT assumption of sample independence allows us to apply our findings to youth with similar range of symptomologies as captured by the SCQ.

The present paper examined an important question related to the construct validity of the PEDI-CAT for use with the ASD population. For instruments developed using Modern Test Theory approaches such as IRT, this analysis is a required first step before proceeding to traditional psychometric analyses (Veloze, et al., 2012). A subsequent study investigating the properties of the PEDI-CAT (ASD) examines whether the high re-test reliability of the original PEDI-CAT is replicated for the PEDI-CAT (ASD) and the relation of PEDI-CAT (ASD) scores to a well-established measure of adaptive behavior (VABS-II). These results will be reported in a separate paper, as well as parents' responses to the novel CAT format.

Conclusion

This study provides initial support for the structural validity of the PEDI-CAT (ASD) for children and youth with ASDs. The criterion scores generated for the Daily Activities, Responsibility and equated (linked) Social/Cognitive domains take into account commonalities and differences in item parameters between children and youth with ASDs and children without disabilities. As a result, criterion scores account for the unique characteristics of children with autism symptoms while remaining comparable to scores generated from the original PEDI-CAT. The findings support the potential use of the PEDI-CAT (ASD) in practice and research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported through a grant from NIH/NICHD R21HD065281 to Boston University. Data featured in this manuscript are available through the NIH-supported National Database for Autism Research (NDAR). Collection ID 1880: http://ndar.nih.gov/data_from_labs.html?id=1880&showSingle=true.

References Cited

- Alguren B, Bostan C, Christensson L, Fridlund B, Cieza A. A multidisciplinary cross-cultural measurement of functioning after stroke: Rasch analysis of the brief ICF Core Set for stroke. *Topics in Stroke Rehabilitation*. 2011; 18(Suppl. 1):573–586. [PubMed: 22120027]
- Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. *British Journal of Psychiatry*. 1999; 175:444–451. [PubMed: 10789276]
- Bonanni P, Gobbo A, Nappi S, Moret O, Nogarol A, Santin M, Randazzo G, Martinuzzi A. Functioning and disability in patients with Angelman syndrome: utility of the International Classification of functioning disability and health, children and youth adaptation framework. *Disability and Rehabilitation*. 2009; 31(Suppl 1):S121–S127. [PubMed: 19968524]
- Carter AS, Volkmar FR, Sparrow SS, Wang JJ, Lord C, Dawson G, Fombonne E, Loveland K, Mesibov G, Schopler E. The Vineland Adaptive Behavior Scales: supplementary norms for

- individuals with autism. *Journal of Autism and Developmental Disorders*. 1998; 28(4):287–302. [PubMed: 9711485]
- Castro S, Ferreira T, Dababnah S, Pinto AI. Linking autism measures with the ICF-CY: functionality beyond the borders of diagnosis and interrater agreement issues. *Developmental Neurorehabilitation*. 2013; 16(5):321–331. [PubMed: 23869843]
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*. 2007; 45(5 Suppl 1):S3–S11. [PubMed: 17443116]
- Chandler S, Charman T, Baird G, Simonoff E, Loucas T, Meldrum D, Scott M, Pickles A. Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46(10):1324–1332. [PubMed: 17885574]
- Charman T, Baird G, Simonoff E, Loucas T, Chandler S, Meldrum D, Pickles A. Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry*. 2007; 191:554–549. [PubMed: 18055961]
- Dolva AS, Coster W, Lilja M. Functional performance in children with Down syndrome. *American Journal of Occupational Therapy*. 2004; 58(6):621–629. [PubMed: 15568546]
- Dumas HM, Fragala-Pinkham MA, Haley SM, Ni P, Coster W, Kramer JM, Kao YC, Moed R, Ludlow LH. Computer adaptive test performance in children with and without disabilities: prospective field study of the PEDI-CAT. *Disability and Rehabilitation*. 2012; 34(5):393–401. [PubMed: 21988750]
- Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah NJ: Lawrence Erlbaum Associates; 2000.
- Engelbert RHH, Custers JWH, van der Net J, van der Graaf Y, Beemer FA, Helders PJM. Functional Outcome in Osteogenesis Imperfecta: Disability Profiles Using the PEDI. *Pediatric Physical Therapy*. 1997; 9(1):18–22.
- Esbensen AJ, Bishop S, Seltzer MM, Greenberg JS, Taylor JL. Comparisons between individuals with autism spectrum disorders and individuals with Down syndrome in adulthood. *American Journal of Intellectual and Developmental Disabilities*. 2010; 115(4):277–290.
- Farley MA, McMahon WM, Fombonne E, Jenson WR, Miller J, Gardner M, Block H, Pingree CB, Ritvo ER, Ritvo RA, Coon H. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Research*. 2009; 2(2):109–118. [PubMed: 19455645]
- Fisch GS, Simensen RJ, Schroer RJ. Longitudinal changes in cognitive and adaptive behavior scores in children and adolescents with the fragile $\times$ mutation or autism. *Journal of Autism and Developmental Disorders*. 2002; 32(2):107–114. [PubMed: 12058838]
- Frazier TW, Ratliff KR, Gruber C, Zhang Y, Law PA, Constantino JN. Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the Social Responsiveness Scale-2. *Autism*. 2014; 18(1):31–44. [PubMed: 24019124]
- Gabriels RL, Ivers BJ, Hill DE, Agnew JA, McNeill J. Stability of adaptive behaviors in middle-school children with autism spectrum disorders. *Research in Autism Spectrum Disorders*. 2007; 1:291–303.
- Gleason K, Coster W. An ICF-CY-based content analysis of the Vineland Adaptive Behavior Scales-II. *Journal of Intellectual and Developmental Disability*. 2012; 37(4):285–293. [PubMed: 22989139]
- Haley SM, Coster WJ, Dumas HM, Fragala-Pinkham MA, Kramer J, Ni P, Tian F, Kao YC, Moed R, Ludlow LH. Accuracy and precision of the Pediatric Evaluation of Disability Inventory computer-adaptive tests (PEDI-CAT). *Developmental Medicine and Child Neurology*. 2011; 53(12):1100–1106. [PubMed: 22077695]
- Haley, SM.; Coster, WJ.; Ludlow, LH.; Haltiwanger, JT.; Andrellos, PA. *Pediatric Evaluation of Disability Inventory: Development, Standardization and Administration Manual*. Boston, MA: Trustees of Boston University; 1992.

- Haley SM, Ni P, Jette AM, Tao W, Moed R, Meyers D, Ludlow LH. Replenishing a computerized adaptive test of patient-reported daily activity functioning. *Quality of Life Research*. 2009; 18(4): 461–471. [PubMed: 19288222]
- Harrison, P.; Oakland, T. *Adaptive Behavior Assessment System, 2nd Edition (ABAS-II)*. San Antonio TX: Harcourt Assessment; 2003.
- Hinderer SR, Gupta S. Functional outcome measures to assess interventions for spasticity. *Archives of Physical Medicine and Rehabilitation*. 1996; 77(10):1083–1089. [PubMed: 8857891]
- Howlin P, Moss P, Savage S, Rutter M. Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2013; 52(6):572–581.e1. [PubMed: 23702446]
- Kao YC, Kramer JM, Liljenquist K, Tian F, Coster WJ. Comparing the functional performance of children and youths with autism, developmental disabilities, and no disability using the revised Pediatric Evaluation of Disability Inventory item banks. *American Journal of Occupational Therapy*. 2012; 66(5):607–616. [PubMed: 22917127]
- Kim J, Chung H, Amtmann D, Revicki DA, Cook KF. Measurement invariance of the PROMIS pain interference item bank across community and clinical samples. *Quality of Life Research*. 2013; 22(3):501–507. [PubMed: 22552604]
- Kothari DH, Haley SM, Gill-Body KM, Dumas HM. Measuring functional change in children with acquired brain injury (ABI): Comparison of generic and ABI-specific scales using the Pediatric Evaluation of Disability Inventory (PEDI). *Physical Therapy*. 2003; 83(9):776–785. [PubMed: 12940765]
- Kramer JM, Coster WJ, Kao YC, Snow A, Orsmond GI. A new approach to the measurement of adaptive behavior: development of the PEDI-CAT for children and youth with autism spectrum disorders. *Physical & Occupational Therapy in Pediatrics*. 2012; 32(1):34–47. [PubMed: 21846290]
- Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, Moed R, Lachowitz MT. Strategies to address the challenges of participant misrepresentation in web-based research. *International Journal of Methods in Psychiatric Research*. 2014; 23:120–129. [PubMed: 24431134]
- Liss M, Harel B, Fein D, Allen D, Dunn M, Feinstein C, Morris R, Waterhouse, Rapin I. Predictors and correlates of adaptive functioning in children with developmental disorders. *Journal of Autism and Developmental Disorders*. 2001; 31(2):219–230. [PubMed: 11450820]
- Newman, L.; Wagner, M.; Cameto, R.; Knokey, A. *The Post-High School Outcomes of Youth With Disabilities up to 4 Years After High School. A Report From the National longitudinal Transition Study-2 (NLTS2)*. Menlo Park, CA: SRI International; 2009. (NCSE 2009-3017)
- Perry A, Flanagan HE, Dunn Geier J, Freeman NL. Brief report: the Vineland Adaptive Behavior Scales in young children with autism spectrum disorders at different cognitive levels. *Journal of Autism and Developmental Disorders*. 2009; 39(7):1066–1078. [PubMed: 19234777]
- Poon KK. The activities and participation of adolescents with autism spectrum disorders in Singapore: findings from an ICF-based instrument. *Journal of Intellectual Disability Research*. 2011; 55(8): 790–800. [PubMed: 21375640]
- Rutter, M.; Bailey, A.; Lord, C. *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services; 2003.
- Schalock, R.; Borthwick-Duffy, S.; Bradley, V.; Buntinx, W.; Coulter, D.; Craig, E.; Gomez, S.; Lachapelle, Y.; Luckasson, R.; Reeve, A.; Shogren, K.; Snell, M.; Spreat, S.; Tassé, M.; Thompson, J.; Verdugo-Alonso, M.; Wehmeyer, M.; Yeager, M. *Intellectual Disability: Diagnosis, Classification, and Systems of Supports*. 11th ed.. Washington DC: American Association on Intellectual and Developmental Disabilities; 2010.
- Sparrow, SS.; Cicchetti, D.; Balla, D. *Vineland Adaptive Behavior Scales, Second ed. (Vineland-II)*. Circle Pines MN: American Guidance Service; 2005.
- Stone WL, Ousley OY, Hepburn SL, Hogan KL, Brown CS. Patterns of adaptive behavior in very young children with autism. *American Journal on Mental Retardation*. 1999; 104(2):187–199. [PubMed: 10207581]
- Streiner, DL.; Norman, GR. *Health Measurement Scales*. 4th ed.. NY: Oxford University Press; 2008.

- Tassé MJ, Schalock RL, Balboni G, Bersani H, Borthwick-Duffy SA, Spreat S, Thissen D, Widaman KF, Zhang D. The construct of adaptive behavior: its conceptualization, measurement, and use in the field of intellectual disability. *American Journal on Intellectual and Developmental Disabilities*. 2012; 117(4):291–303. [PubMed: 22809075]
- Taylor JL, Seltzer MM. Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood. *Journal of Autism and Developmental Disorders*. 2011; 41(5):566–574. [PubMed: 20640591]
- Veloza CA, Seel RT, Magasi S, Heinemann AW, Romero S. Improving measurement methods in rehabilitation: core concepts and recommendations for scale development. *Archives of Physical Medicine and Rehabilitation*. 2012; 93(8 Suppl):S154–S163. [PubMed: 22840881]
- Wainer, H.; Dorans, NJ.; Eignor, D.; Flaugher, R.; Green, BF.; Mislevy, RJ.; Steinberg, L.; Thissen, D. *Computerized Adaptive Testing: A Primer*. 2nd ed.. Mahway NJ: Lawrence Erlbaum Associates; 2000.
- Witwer AN, Lecavalier L. Autism screening tools: an evaluation of the Social Communication Questionnaire and the Developmental Behaviour Checklist-Autism Screening Algorithm. *Journal of Intellectual and Developmental Disabilities*. 2007; 32(3):179–187.
- World Health Organization. *International Classification of Functioning, Disability and Health - Children and Youth Version (ICF-CY)*. Geneva: WHO; 2008.
- World Health Organization. *The International Classification of Functioning, Disability, and Health (ICF)*. Geneva: WHO; 2001.

Figure 1. PEDI-CAT item selection process

The figure illustrates the sequence of steps implemented by the CAT algorithm to select which items to present and how many items to administer.

Figure 2. An Illustration of Differential Item Function (DIF)

This figure illustrates the location of a set of four items along the underlying (unidimensional) continuum of ability for two different samples. The location of 3 of the items is the same in both samples however the location of item B is not. The different location of item B indicates that item B is much more difficult (requires more underlying ability) for persons in sample 2. The dotted line identifies the score (level of ability) of a hypothetical person. According to the IRT measurement model for sample 1, a person with this score would be predicted to have achieved both skill A and skill B. However, this prediction would not be accurate for a person with that score in sample 2.

Table 1Sample Demographics ($N = 365$).

Demographics		<i>n</i> (%)
Race		
	Caucasian	312 (85.5%)
	Multi-racial	26 (7.1%)
	African-American/Black	5 (1.4%)
	Asian	7 (1.9%)
	Other	14 (4.1%)
Ethnicity		
	Hispanic/ Latino (a)	35 (9.6%)
	Not Hispanic	330 (90.4%)
Family Income *		
	Less than \$30,000	46 (12.6%)
	\$30,000–\$60,000	89 (24.4%)
	\$60,000–\$100,000	95 (26%)
	Above \$100,000	102 (27.9%)
Communication Method		
	Sentences	272 (74.5%)
	Single words	43 (11.8%)
	Sign language, gestures, augmentative, or facilitated communication	50 (13.7%)
SCQ Current Score		
	0–14	116 (31.8%)
	15–21	131 (35.9%)
	Greater than 22	118 (32.3%)
SCQ Lifetime Score **		
	0–14	31 (10.3%)
	15–21	72 (24%)
	Greater than 22	197 (65.7%)
Reported Full Scale IQ level ***		
	Average or above	127 (37.8%)
	Borderline/mild	93 (23.9%)
	Moderate/severe	30 (7.7%)

* 33 respondents (9%) preferred not to report their income

** optional: 300 respondents completed the SCQ Lifetime

*** 27% of respondents did not provide this information

Table 2

Confirmatory Factor Analysis of PEDI-CAT-ASD Domains

Domain	Number of Items	CFI	TLI	RMSEA
Daily Activities	76	0.98	0.98	0.04
Social/Cognitive	68	0.93	0.93	0.07
Responsibility	58	0.97	0.97	0.05

CFI = Comparative Fit Index; TLI= Tucker Lewis index; RMSEA = Root Mean Square Error of Approximation

Table 3

Agreement (Intraclass Correlation Coefficient, 95%CI) between scores from 5, 10, and 15-item CAT and full item bank for PEDI-CAT-ASD domains

	Daily Activities	Responsibility	Social Cognition
5-item CAT	0.95(0.94,0.96)	0.95(0.94,0.96)	0.95(0.94,0.96)
10-item CAT	0.97(0.96,0.98)	0.98(0.98,0.99)	0.98(0.97,0.98)
15-item CAT	0.98(0.99,0.99)	0.99(0.99,0.99)	0.98(0.98,0.98)

Table 4
Percent of subjects with score reliability >.90 and >.95 for PEDJ-CAT-ASD domains

	Daily Activities		Social/Cognitive		Responsibility	
	% r >.90	% r >.95	% r >.90	% r >.95	% r >.90	% r >.95
5-item CAT	84	24	77	19	75	38
10-item CAT	95	84	96	77	91	71
15-item CAT	96	91	98	85	95	84
Full item bank	98	95	99	97	96	92