Development and Standardization of the Diagnostic Adaptive Behavior Scale: Application of Item Response Theory to the Assessment of Adaptive Behavior

Marc J. Tassé, The Ohio State University
Robert L. Schalock, Hastings College
David Thissen, University of North Carolina at Chapel Hill
Giulia Balboni, University of Pisa (Italy)
Henry (Hank) Bersani, Jr.†, Western Oregon University
Sharon A. Borthwick-Duffy, University of California – Riverside
Scott Spreat, Woodland Center for Challenging Behaviors
Keith F. Widaman, University of California – Riverside
Dalun Zhang, Texas A&M University
Patricia Navas, University of Zaragoza (Spain)

Published in:

***American Journal on Intellectual and Developmental Disabilities, 121,* 79-94.**
DOI: 10.1352/1944-7558-121.2.79

Citation: Tassé, M. J., Schalock, R. L., Thissen, D., Balboni, G., Bersani, H. A., Borthwick-Duffy, S. A., Spreat, S., Widaman, K. F., Zhang, D., & Navas, P. (2016). Development and Standardization of the Diagnostic Adaptive Behavior Scale: Application of Item Response Theory to the Assessment of Adaptive Behavior. *American Journal on Intellectual and Developmental Disabilities,* 121, 79-94.

Authors' Note: The development and standardization of the *Diagnostic Adaptive Behavior Scale* (DABS) and the writing of this manuscript was supported in part with funding from the American Association on Intellectual and Developmental Disabilities and the U.S. Administration on Intellectual and Developmental Disabilities grant #90-DD-0691. The American Association on Intellectual and Developmental Disabilities owns the copyright on the DABS and will produce and distribute the DABS and its products. It should be noted that none of the DABS authors receive royalties from the sale of the DABS. The authors thank Suzanne Davis and Matthew Grover for their help with the recruitment and data collection process. We also express our gratitude to all the individuals who accepted to be assessed on the DABS as well as their parents/caregivers, teachers, and other professionals who participated as respondents and interviewers. Patricia Navas is now at University of Zaragoza and Keith F. Widaman is now at University of California – Riverside. Correspondence regarding this article should be addressed to Marc J. Tassé, The Ohio State University, Nisonger Center UCEDD, 1581 Dodd Drive, Columbus, OH USA 43210. E-mail: marc.tasse@osumc.edu.

Abstract

The Diagnostic Adaptive Behavior Scale (DABS) was developed using item response theory (IRT) methods and was constructed to provide the most precise and valid adaptive behavior information at or near the cutoff point of making a decision regarding a diagnosis of intellectual disability. The DABS initial item pool consisted of 260 items. Using IRT modeling and a nationally representative standardization sample, the item set was reduced to 75 items that provide the most precise adaptive behavior information at the cut-off area determining the presence or not of significant adaptive behavior deficits across conceptual, social and practical skills. The standardization of the DABS is described and discussed.

Keywords: adaptive behavior, intellectual disability, diagnosis, item response theory

Development and Standardization of the *Diagnostic Adaptive Behavior Scale*: Application of Item Response Theory to the Assessment of Adaptive Behavior

The American Association on Intellectual and Developmental Disabilities (AAIDD; Schalock et al., 2010), Diagnostic and Statistical Manual on Mental Disorders (DSM-5; American Psychiatric Association, 2013), and International Classification of Diseases (ICD-10, World Health Organization, 1992) all require the presence of significant deficits in adaptive behavior as a diagnostic criterion for intellectual disability. Schalock et al. (2010) operationally defined significant deficits in adaptive behavior as representing a performance that is approximately 2 standard deviations or more below the population mean. This statistical definition of "significant deficits" is well accepted and has also been used to define significant subaverage intellectual functioning in previous editions of the DSM and AAIDD manuals.

Deficits in adaptive behavior are also the defining feature of other developmental disabilities. For example, the US federal definition of developmental disabilities requires the presence of deficits in 3 of 7 adaptive skill areas (Developmental Disabilities Assistance and Bill of Rights Act, 2000). Harrison and Oakland (2003) also documented that several other disorders, including autism spectrum disorder, attention-deficit/hyperactive disorder, emotional and behavioral disorders, hearing and motor impairments, communication disorders, learning disabilities, and neurological disorders also present core features that are inclusive of deficits in adaptive behavior. Aging-related decline, dementia, and Alzheimer disease are other important areas that track changes in adaptive behavior as measures of disease progression and prognosis (Takata et al., 2013; Zigman, Schupf, Urv, Zigman, & Silverman, 2002).

Over the years, both AAIDD (Luckasson et al., 1992, 2002; Schalock et al., 2010) and DSM-5 (APA, 2013) diagnostic and classification systems have given the adaptive behavior criterion more importance in defining intellectual disability (see APA, 2013; Luckasson et al., 1992, 2002; Schalock et al., 2010). As further evidence of the important role of adaptive behavior, the DSM-5 abandoned its reliance on IQ scores to quantify the severity of intellectual disability. Instead, the DSM-5 is now using the level of deficits in adaptive functioning across conceptual, social, and practical skills as the specifier related to the severity of intellectual disability, which they classify as mild, moderate, severe, or profound (APA, 2013). This is a significant paradigm shift for the DSM. Hence, there is an increasing need for reliable and valid standardized measures of adaptive behavior.

The availability of psychometrically sound standardized tests to assess adaptive behavior has increased dramatically over the past 50 years. According to Schalock (1999), a review of the available instruments indicated the existence of more than 200 adaptive behavior instruments. Not all instruments are of equal psychometric robustness nor were they all developed for use in making a diagnosis of intellectual disability. Tassé and his colleagues (2012) determined that four standardized adaptive behavior instruments possessed characteristics and psychometric properties sufficient to be used in the clinical determination of intellectual disability. These four instruments are: *Adaptive Behavior Assessment System, 2nd edition* (ABAS-2; Harrison & Oakland, 2003), *Adaptive Behavior Scale – School, 2nd edition* (ABS-S: 2; Lambert, Nihira, & Leland, 1993), *Scales of Independent Behavior-Revised* (SIB-R; Bruininks, Woodcock, Weatherman, & Hill, 1996), and *Vineland Adaptive Behavior Scale, 2nd edition* (VABS-2; Sparrow, Cicchetti, & Balla, 2005). These four adaptive behavior instruments are based on the measurement of specific adaptive skills that reflect a multidimensional conceptual and measurement model of adaptive behavior. However, as discussed by Tassé et al. (2012), all these instruments were designed to measure the full range of individual differences in adaptive behavior without regard to capturing those skill levels around the cutoff point for determining significant limitations in adaptive behavior.

AAIDD (Luckasson et al., 2002; Schalock et al., 2010) operationally defined adaptive behavior as skills that are learned and performed to meet society's expectations of individuals of the same age and cultural group across three domains: **conceptual skills** (language, reading and writing, and money, time and number concepts), **social skills** (interpersonal skills, social responsibility, self-esteem, social problem solving, following rules and obeying laws, avoiding being victimized, and wariness), and **practical skills** (personal hygiene, safety, healthcare, travel/transportation, following schedule, home-living skills, and occupational skills). The concept of adaptive behavior and its definition has been a part of the AAIDD definition of intellectual disability for the past 5 decades. The AAIDD tripartite definition of adaptive behavior consisting of conceptual, social and practical skills served as the basis for the development of the Diagnostic Adaptive Behavior Scale (DABS) and its standardization is described in this article[1].

In subsequent sections, we describe the two main phases of data collection and analyses that led to the final DABS. Data collection began using item tryout forms that included all potential items across the entire age range; interviewer feedback suggested that many items appeared inappropriate for some ages, so Phase 1 analyses of early data were carried out to divide the items into subsets appropriate for three age ranges, for individuals 4-8 years old, 9-15 years old, and 16-21 years old. The remainder of the scale-development data was collected with the items divided into forms for those three age groups. In the Phase 2 analyses, the items were calibrated to support item selection and creation of the final forms. The items were then recalibrated using only the data from the ultimately-selected items, and the scoring rules were established.

Phase 1: Preliminary Analyses and the Development of Age-Specific Forms

Participants

The sample of assessed persons consisted of 474 individuals recruited from across the country, including participants from 36 US states. Table 1 presents the demographics of the assessed individuals in this initial calibration sample. These participants ranged in age from 4 to 21 years with a mean of 12.0 years and standard deviation of 4.9 years. Fifty-two percent of the assessed individuals were males. All DABS were administered via an interview between an interviewer and a respondent. No DABS data was collected using self-report, per DABS administration procedures. The respondents were selected on the basis of their knowledge and familiarity with the assessed person. More than 80% of the respondents interviewed with the DABS were parents and other family members of the assessed individual. All respondents had to have known the assessed individual for at least three months. The interviewers were professionals who had at least a bachelor degree and some training or experience with individual assessment. Interviewers included teachers, school psychologists, social workers, graduate students, and a variety of other professionals. A total of 96 interviewers participated in assessing the 474 individuals on the [Phase 1 version] DABS. The number of DABS interviews done by individual interviewers ranged from 1 to 20, with the mean number of DABS interviews being close to 5. Complete demographic information regarding the interviewers is found in Tassé et al. (in press).

<Table 1>

[1] The reader interested in learning more about the initial field-testing of items and the use of Q-sort process and expert consultation in the earlier phases of scale development should consult the DABS: User's Manual (Tassé et al., in press).

Scale

The DABS is an individually administered standardized test of adaptive behavior developed based on the tripartite model of adaptive behavior including: conceptual, social, and practical adaptive skills. The DABS is unique in that it was specifically designed and developed to ensure congruence with the AAIDD (Schalock et al., 2010), and now DSM-5 (APA, 2013) definition of adaptive behavior.

The DABS was developed to assess the adaptive behavior of persons between the ages of 4 and 21 years, inclusive. The DABS version used in Phase 1 consisted of one protocol containing a total of 260 items across the three adaptive behavior domains: Conceptual Skill (Ni = 94), Social Skills (Ni = 86), and Conceptual Skills (Ni = 80). The DABS was constructed to provide precise information around the point where an individual presents significant limitations in adaptive behavior and to distinguish between persons who do and do not have intellectual disability. The DABS was developed to be administered via interview with a respondent who knows well the person whose adaptive behavior is being assessed. The *respondents* have to be adults and are generally selected from the assessed person's family, friends, teachers, employers, or direct support staff (if appropriate). The individual being assessed cannot serve as a respondent for himself/herself. The respondent should know the individual very well and have had the opportunity to observe her/him on a daily or weekly basis, preferably in a variety of settings, and over an extended period of time. The respondent assesses the individual's behavior based on direct observation and knowledge of the adaptive behavior of the assessed individual. The respondent rates the assessed person's typical performance on each behavioral item using the following ratings: 0 = No-rarely or never does it; 1 = Yes-does it with reminders or assistance but rarely or never independently; 2 = Yes-does it sometimes independently—but sometimes needs reminders or assistance; 3 = Yes-does it always or almost always independently—never or rarely needs reminders. Also available, as an exceptional rating, is the "No Score". The "No Score" should be used only if (a) the assessed person has a lack of opportunity due to cultural, gender, and/or geographic/regional factors; a lack of opportunity due to environmental constraints; or (b) the respondent has no direct knowledge of individual's typical performance. The "No Score" rating is treated as missing information when scoring the DABS and an item rated as "No Score" does not factor into the individual's standard score but the number of "No Score" ratings may impact the overall reliability index or standard error of measurement for that individual's DABS scores.

The DABS is administered via a semi-structured interview process between a respondent (described above) and an interviewer. The *interviewer* is generally a professional who has completed at least a bachelor's degree, has several years of direct work experience with individuals between the ages of 4-21 years who have an intellectual disability or related developmental disability, and has some previous assessment experience. Under certain circumstances, individuals who do not have minimally a bachelor's degree but have extensive professional experience and knowledge of behavioral assessment or psychological testing may be acceptable interviewers.

Data Collection

Interviewers were the principle point of recruitment for this study. We recruited potential interviewers by contacting various professional organizations (e.g., National Association of School Psychologists, American Psychological Association, American Association on Intellectual and Developmental Disabilities, etc.) and using email communications, webpage postings, and other electronic media. Recruitment targeted professionals who were comfortable with assessments and had access to individuals between the ages of 4 and 21 years with and without disabilities. These recruited interviewers were then instructed to find individuals from the general population to be assessed and find respective respondents to be interviewed to administer the DABS. The interviewers communicated with a research assistant working with the senior author and provided an estimate of the number of DABS protocols they would need. The research assistant sent these protocols and consent forms with self-addressed and post-paid return envelopes. We also recruited a small sample of individuals in this same age group who had a known diagnosis of intellectual disability to establish some basic psychometrics properties of the scale (described below). The Ohio State University Institutional Review Boards reviewed and approved the methodology involving human subjects.

Item Response Theory

As a tool for item analysis, item response theory makes explicit the fact that adaptive skills scales are intended to measure individual differences on an unobserved, or latent, dimension (e.g., conceptual, social, and practical skills), about which the item responses are observable indicators. Item response theory is based on functions T_i that trace the probability of a particular response to an item (e.g., item *i)* as a function of the latent variable being measured, represented by theta (θ). The logistic version of the Samejima's (1969, 1997) Graded Response Model was used to calibrate the DABS items and is used in scoring the DABS. For the DABS, θ represents the underlying level of conceptual, social, or practical skills, depending on the selection of items.

For any DABS item *i* with four ordered response alternatives *u* (i.e., potential respondent ratings on an item), where $u = k$, for $k = 0, 1, 2, 3$, the Graded Response Model is:

$$T_i\ (u_i = k \mid \theta) = \frac{1}{1+\exp[-a_i(\theta - b_{i,k})]} - \frac{1}{1+\exp[-a_i(\theta - b_{i,k+1})]}\ , \qquad \text{A.1}$$
$$= T_i^*(k) - T_i^*(k+1)\ .$$

To complete the model-definition, we note that the probability that a person scores 0 or higher on a given item is 1.0, or $T^*_i(0) = 1$ and the probability that a person would score higher than 3 is zero, or $T^*_i(4) = 0$. The parameter a_i is the slope of the binary logistics $T^*_i(k)$, which are the trace lines describing the probability that a response is in category *k* or higher, for each value of θ. In other words, the slope describes the ability of a specific item to differentiate among people at different levels along the trait continuum (i.e., discrimination parameter). The parameters denoted $b_{i,k}$ are the threshold locations and represent response category difficulty. The value of $b_{i,k}$ is the point on the θ-axis at which the probability passes 50% that the response is in category *k* or higher.

All analyses were carried out using IRTPRO 2.1 for Windows (Cai, Thissen, & du Toit, 2011). Information provided by the parameter estimates (i.e., discrimination and difficulty) can be used to select those items that are appropriately matched to the measured trait, concentrating their information low on the scale, near the cut score falling 2 standard deviations below the population mean on the dimension.

Initial item analyses were conducted to check the appropriateness of the graded model, and to refine and reduce the test protocol from one large 260-item instrument for the entire 4-21 year age range DABS to 3 smaller age-related forms: 4-8 years old, 9-15 years old, and 16-21 years old.

First Preliminary Analyses

The Nominal Response Model, (Bock, 1972; Thissen, Cai, & Bock, 2010) was used as a preliminary check that the four scored response categories empirically fall in the expected order. The Nominal Response Model indicated that response alternatives for items comprising the conceptual skills, social skills, and practical skills domains of the DABS had scoring function values very close to 0, 1, 2, and 3, supporting the use of a simpler model for graded data (i.e., Graded Response Model) for the final versions.

Item response theory (IRT) analyses using bifactor models (Gibbons, Bock, Hedeker, Weiss, Segawa, et al., 2007; Gibbons & Hedeker, 1992; Cai, 2010; Cai, Yang, & Hansen, 2011) were carried out to check for unidimensionality of the DABS conceptual, social, and practical adaptive behavior domains separately. The bifactor models indicated that substantial statistical multidimensionality was present within each of the sub-domain (conceptual, social, and practical) as indicated by the explained common variance or the amount of variance attributable to a general factor. Values ranged from .91 to .67. These results suggested, as will be discussed later, one of the final criteria for the selection of items within each DABS age form.

We calibrated the items with the Graded Response Model, which yields results like those shown graphically in Figure 1. Figure 1 illustrates items "*Puts on clothing"* and "*Gets a prescription filled,"* both within the practical skills subscale ("*Activities of Daily Living*" and "*Health Care*" subdomains, respectively). The top panel illustrates the trace lines for each response category (i.e., 0, 1, 2, & 3) for those two items. The center panel illustrates the expected scores as a function of θ (i.e., adaptive behavior), and the bottom panel reflects the item information curves (IIC) indicating the range over θ for which the scale is most informative or precise (numbers near the x-axis indicate average scores at different ages).

<Figure 1>

A parameter that was used to inform the assignment of items to age groups was b_3, the estimated level of the latent variable for a probability of 0.5 of responding in the highest category. The item "*Puts on clothing"* had a b_3 value of -1.04 (b_3 values for all the items in the "*Activities of Daily Living*" sub-domain ranged from .32 to -1.31), and the value for "*Gets a prescription filled"* is located higher on the scale. These values of b_3 can be "age-interpreted" by comparing them to the average scores (on the same θ axis) for each age (4, 5, 6, …), and assigning each item to the age with average closest to b_3.

We also computed the percentages of missing and "no score" for each item in three age groups: 4-8, 9-15, and 16-21. Then we sorted the items and divided them into three groups: 1) Items with relatively large percentages of missing data (more than 8%) or "no score" for the 4-8 and 9-15 groups; those became the (tentative) "16-21" items. 2) Items with relatively large percentages of missing data or "no score" for the 4-8 age group (only); those became the (tentative) "9-15" items. The rest of the items became the (tentative) "4-8" items.

Combining the categorization of the items based on "missingness" and the IRT calibration, we sorted the items within age categories by the age-interpreted value of b_3, and looked at the content of items near the age-group boundaries, and shifted a few items into age groups for which they had greater face validity. This produced three revised lists of items—one for age 4-8, a second for age 9-15, and a third for 16-21.

We did not administer three distinct, non-overlapping sets of items for children in the three age groups for the item tryout, for two reasons: 1) Our preliminary age-categorization of an item could have been wrong, and we could have decided to shift some; to do that we would have needed data from another age group. 2) We needed to "link" the three age groups with common items—so we need some items that both 8- and 9-year olds have taken, and some (other) items that both 15- and 16-year olds have taken.

The reduced-length item tryout forms were:

- The 4-8 year old children see the "4-8" items (only): 54 CS items, 49 SS items, and 29 PS items for a total of 132.
- The 9-15 year old children see the "4-8" items (to link, and to measure low in the skill range), and the "9-15" items: 81 CS items, 70 SS items, and 41 PS items for a total of 192.
- The 16-21 year old children see the "9-15" items (to link, and to measure low in the skill range), and the "16-21" items: 40 CS items, 37 SS items, and 51 PS items for a total of 128.

Second Preliminary Analyses

To confirm that we would still obtain suitable data for the final calibration if we administered only one of the three reduced forms for respondents in each age group, we re-analyzed the preliminary data, under the pretense that these three age specific item tryout forms had been used to collect the data, by making-missing all of the responses that weren't already missing for items that would not have been administered using the reduced forms. IRT calibration of the items with the graded model yielded comparable parameter estimates to those we had obtained using all of the data.

The remainder of the item tryout / calibration data were collected with the reduced, age-specific forms.

Phase 2: Overall Analyses, Item Selection, and Final Calibration and Scoring

Participants

The total calibration sample consisted of 1,058 assessed individuals aged from 4 to 21 years old from 46 states across the US were assessed on the DABS. Table 1 presents the demographics of the assessed individuals in the total calibration sample. This sample of assessed persons on the DABS has also been previously described in Balboni et al. (2014). The mean chronological age of the assessed individuals in the total calibration sample was 11.1 years with a standard deviation of 4.9 years. Gender distribution was evenly distributed with 50% of the sample being male. The respondents and interviewers were selected on the same basis as previously described for Phase 1. Again, the total sample of respondents was predominantly parents, with 78% being either the assessed individual's mother or father. A total of 171 interviewers participated in assessing the 1,058 individuals on the DABS. The modal number of DABS interviews completed by interviewers was 2 and the mean was 6. Complete demographic information regarding these participants is presented in Tassé et al. (in press).

Scale

The three-age forms of the DABS described above were used in Phase 2. The administration procedure is described above and remained unchanged. The only difference was, depending on the assessed individual's chronological age, she/he was administered the appropriate age-form of the DABS.

Data Collection

The recruitment procedure described for Phase 1 was continued in Phase 2. Data collection continued using the three age-forms: DABS 4-8 years old (Ni = 132), DABS 9-15 years old (Ni = 192), and DABS 16-21 years old (Ni = 128).

Analyses for Item Selection

We conducted IRT analyses separately on the three age forms, investigating differential item functioning (DIF) between age groups, estimating the parameters of bifactor models to provide item-level information about multidimensionality to use in item selection, and then, within each of the three adaptive skill domains, and item calibration using the entire sample of 1,058.

Six DIF analyses were conducted to determine the invariance of item parameters for the linking items between each adjacent pair of age groups. Using Wald tests (Woods, Cai, and Wang, 2013) and the Benjamini-Hochberg procedures for multiple comparisons (Benjamini & Hochberg, 1995; Williams, Jones, & Tukey, 1999; Thissen, Steinberg, & Kuang, 2002), a small number of items were found to exhibit DIF (18 items out of 164); those items were flagged, so that if they were ultimately selected for use on the scales, they would be treated as different items (non-link items) between age groups.

Factor loading estimates for each item from the bifactor models were compiled as information to be used in item selection, so that preference could be given to more locally independent items within each domain.

To obtain estimates the items' parameters (i.e., discrimination and difficulty), three unidimensional item response theory calibrations were performed, one each for the conceptual skills, social skills, and practical skills domains, including the items for all three age-groups simultaneously. The three age-group forms were linked using overlapping item sets as discussed above. Such linkage permits direct comparisons between item parameter estimates from different samples (Bjorner, Chang, Thissen, & Reeve, 2007). As explained by Orlando, Sherbourne, and Thissen (2000), once items parameters are estimated, linking methods allow the researcher to compare scores for different groups of people (i.e., individuals aged 4-8, 9-15, and 16-21 in this case) who did not answer exactly the same questions on a given domain. Linking items that appear on the scales for two age groups have the same parameters for both instances; "non-link" items (that exhibited DIF between age groups) have different parameters for the two age groups—it is as though those were different items.

The units of the item response theory variable θ (i.e., conceptual skills, social skills, and practical skills) were established by standardizing the normal population distribution for the scales for ages 9-15. The means and standard deviations in that metric for the other two age groups were also estimated, and those values are shown in Table 2.

<Table 2>

After calibration, four IRT-related criteria were used for the process of item selection/retention for each of the three DABS age forms with the final goal of developing the shortest possible and most precise measure (i.e., a total of 75 items, 25 items for each of conceptual skills, social skills, and practical skills domains):

1. Select items that make up a relatively unidimensional set within each of the three adaptive behavior domains. This criterion tried to avoid the substantial statistical multidimensionality detected before, so unidimensional item response theory models (Graded Response Model) could be used. Unidimensionality was operationalized as a relatively high value of the Explained Common Variance (ECV) statistic, computed using the parameters of the bifactor models.
2. Retain a certain representation of items within the sub-domains. In the process of making the domains unidimensional, we also wanted to retain, to the extent possible, a representation of items among the sub-domains.
3. Select items that have lower missing data (e.g., “No Score” ratings) proportions (i.e., less than 8%).
4. Select the items that have relatively high information function (i.e., reliability). Items were selected within the constraints of the preceding three criteria if they were informative, especially in the lower range of adaptive behavior where the cutoff score falls (i.e., approximately 2 SD below the population mean). To accomplish this, we computed the within-age means and standard deviations of the conceptual skills, social skills, and practical skills domains for the “Non ID” portion (N = 937) of the total sample. These data are presented in Table 3. We used IRTPRO to compute the *expected a posteriori* (EAP) scores (the expected latent score for each individual given his/her response pattern and the estimated model parameters), and then computed the means and SDs of those EAPs within each age group. These distributions provided preliminary estimates of the locations of cut scores.

<Table 3>

These four criteria, however, did not lead entirely to a mechanical item selection, because another prescription in IRT-related scale development is that, within the item selection process described in reference to the four criteria summarized above, the process needs to retain sufficient “linking items” to connect the scales across age groups. The final items were selected based on the preceding four criteria in addition to a consideration of including within each age form items that were: (a) “unique” in that the item appeared only on the respective age form; (b) “linked” from one age group to the next; and (c) “non-linked” in that they are the same questions, but with different item parameters for one age group than the other. A total of 75 items per age-group form (i.e., 25 items per adaptive behavior domain) were selected using these criteria. A summary of these unique, linked, and non-linked items is presented in Table 4.

<Table 4>

Final Calibration: Item Parameters Used For Scoring

The process of item-selection described above effectively re-defines what is being measured. Thus, it was necessary to recalibrate the scales using only the data from the selected items. Three major blocks of item response theory analyses were used to perform this recalibration function: (a)DIF analyses to check the (desired) invariance of item parameters across forms (i.e., age groups); (b) bifactor analyses to evaluate unidimensionality using ECV as an index; and (c) recalibration of the items to yield item parameters to be used for scoring.

The same six DIF analyses were described in the preceding section were repeated to determine the invariance of item parameters for the 5-18 linking items identified in Table 4. No significant DIF was found for any of the linking items. Because we found the desired invariance, we used the link sets as shown in Table 4.

Bifactor analyses were again used to check for unidimensionality of the final conceptual skills, social skills, and practical skills domains. As shown in Table 5, 85-94% of the common variance across the nine domains (i.e., 3 age forms, each with 3 adaptive behavior domains) was explained by the general factor. This confirms the essential unidimensionality (Stout, 1990) of the DABS adaptive behavior domains. Although some statistically significant multidimensionality was found within domains attributable to variation among responses to items within sub-domains (e.g. language, reading, self-direction, time, etc.), these sources of variance are small relative to the general factor within each conceptual skills, social skills, and practical adaptive behavior domains.

<Table 5>

The final recalibration of items to yield the item parameters to be used for scoring involved calibration of the selected items (i.e., total of 75 items per age form) as three cross-age scales (conceptual skills, social skills, and practical skills). In each of the three calibrations, the reference group for the item response theory scale is an assumed normal distribution with a mean of 0 and a standard deviation of 1 for the age 9-15 sample; means and standard deviations for the 4-8 and 16-21 samples were estimated using the linking items shown in Table 4. This yields item parameters that can be used to compute item response theory scale scores. All parameter estimates can be found in the DABS User's Manual (Tassé et al., in press).

After calibration of the items, item response theory metrics were formed for each of conceptual, social, and practical adaptive behavior domains, each representing a developmental continuum from age 4 to 21 years. Re-expressing the item response theory scale scores on a scale having within-age mean = 100 and within-age standard deviation = 15 requires computation of estimates of the within-age mean and standard deviation of the EAPs. Because the within-age sample sizes were small (all less than 100), we computed smoothed estimates of those means and standard deviations for the conceptual, social, and practical adaptive skills (Table 6) along with the corresponding smoothed values for the overall DABS standard score that is computed as the numerical average of the conceptual skills, social skills, and practical skills item response theory scores for each participant. Given the (smoothed) means and standard deviations in Table 6, the item response theory scale scores can be standardized within-age by subtracting the (smoothed) mean and dividing by the (smoothed) standard deviation. Then those within-age standardized scores are multiplied by 15 and added to 100 to yield the scores on the reporting scale (e.g., if (Person_Age <= 4) CSscore = 100+15*((EAP(q_{CS})-(-1.96))/0.44). The relation between the item response theory metrics and the reporting metrics is described in more detail in the DABS User's Manual (Tassé et al., in press) along with specific examples on how the final DABS scores are computed. In practice, these scores will be computed using software, because this process is very computationally intensive.

<Table 6>

Figures 2, 3, 4, and 5, show the loess-smoothed age-group means for the Non-ID sample in the upper curve and the two loess-smoothed within-age standard deviations below that mean in the lower curve. Individuals with an ID-related diagnosis, no verified condition, or a non-ID related diagnosis are also identified in these figures. Two general trends are apparent in these three figures. First, one sees an increase in conceptual, social, and practical adaptive skills (and total adaptive behavior skills) with age, and second, those with no verified condition have higher EAP [θ] scores. Reflecting this second trend, the sensitivity and specificity of the DABS have been evaluated and reported recently by Balboni et al. (2014). In that study, DABS sensitivity was reported to range from 81% to 98% and specificity to range from 89% to 91%, indicating strong sensitivity and specificity.

<Figures 2, 3, 4, 5>

Discussion

The DABS was developed and standardized using item response theory methods to select the most appropriate and informative items to assess conceptual, social, and practical skills. In contrast to classical test theory, item response theory provides a clear theoretical model tying individual differences on an underlying construct to the probability of item responses. Unlike classical test theory, under which psychometric characteristics of a test are group-dependent, item response theory postulates that item characteristics are specific to the item and considered invariant across groups (Lord, 1980).

One advantage of item response theory is its ability to provide scores for different individuals on a common metric without necessitating the administration of the same number of items to each individual or even any of the same items to different individuals, assuming these items are drawn from a common item bank. More importantly, item response theory allows one to measure individual levels of performance reliably across the continuum of adaptive skills and ages, with special attention given to providing precise information around the cutoff point for determining significant limitations in adaptive behavior—one of the three essential components of a diagnosis of intellectual disability.

Using item response theory as the approach to test development, the DABS was developed over the course of a six-year process. This paper describes the item parametrization based on a sample drawn from the general population from across the United State and the revisions of the DABS from one 260-item protocol to three age-forms ranging from 128, 132, and 192 items, respectively, to a final DABS version with three age forms containing each 75 items.

The final DABS instrument is a 75-item scale providing adaptive behavior information at or near the cutoff point for making a decision regarding a diagnosis of intellectual disability for individuals 4 to 21 years of age. The DABS provides standard scores for conceptual, social, and practical adaptive behavior skills on a metric with a mean of 100 and standard deviation of 15. It also provides an overall DABS standard score on the same standard metric. Although the three adaptive behavior domains are demonstrably unidimensional, an overall adaptive behavior score is nonetheless provided since many clinicians may want to know and review a composite or overall adaptive behavior estimate of the assessed individual.

Initial analyses indicated that the DABS can accurately identify individuals known to have an intellectual disability (Balboni et al., 2014). These findings are promising and may provide clinicians and researchers in intellectual disability with a scale that has fewer items and is therefore easier to administer without sacrificing sensitivity and specificity. Since the DABS was specifically developed as a robust measure to be used for the purpose of ruling-in/ruling-out a diagnosis of intellectual disability its strength is in measuring an individual's level of adaptive behavior at or near what is defined as "significant deficits" and thus contains fewer items to assess average or above average levels of adaptive behavior. It should also be noted that the utility of the DABS as a measure for planning intervention goals or its sensitivity to be used as a measurement of change in adaptive behavior are not known.

Future research is needed to further establish the psychometric properties of the DABS, including its reliability and validity. Validity studies should compare scores obtained on the DABS with other already established standardized measures of adaptive behavior. Because the DABS is a respondent-based measure, as are other measures of adaptive behavior, these psychometric studies should estimate the scale's inter-respondent concordance as well. The DABS remains a promising, new, standardized adaptive behavior scale that provides individual standard scores for conceptual, social, and practical adaptive behavior domains as well as an overall DABS standard score. This new instrument, with a total of 75 items, provides a comprehensive adaptive behavior assessment to assist with making a diagnosis of intellectual disability.

References

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Balboni, G., Tassé, M. J., Schalock, R. L., Borthwick-Duffy, S. A., Spreat, S., Thissen, D., Widaman, K. F., Zhang, D., & Navas, P. (2014). The Diagnostic Adaptive Behavior Scale: Evaluating its diagnostic sensitivity and specificity. *Research in Developmental Disabilities, 35,* 2884-2893.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289-300.

Bjorner, J. B., Chang, C. H., Thissen, T., Reeve, B. R. (2007). Developing tailored instruments: item banking and computerized adaptive assessment. *Quality of Life Research, 16,* 95-108.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika*, *37*, 29-51.

Bruininks, R. H., Woodcock, R., Weatherman, R., & Hill, B. (1996). *Scales of Independent Behavior—Revised*. Chicago, IL: Riverside Publishing.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*, 581-612.

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

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, *16*, 221-248.

Developmental Disabilities Assistance and Bill of Rights Act. (2000). *Public Law 106–402*. 106th US Congress.

Doll, E. A. (1953). *Measurement of social competence: A manual for the Vineland Social Maturity Scale*. Circle Pines, MN: American Guidance Service, Inc.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4-19.

Gibbons, R. D. & Hedeker, D. (1992). Full-information item bifactor analysis. *Psychometrika*, *57*, 423-436.

Harrison, P. L. & Oakland, T. (2003). *Adaptive Behavior Assessment System Second Edition: Manual*. San Antonio, TX: Harcourt Assessment, Inc.

Lambert, N., Nihira, K., & Leland, H. (1993). *Adaptive Behavior Scale—School, Second Edition*. Austin, TX: PRO-ED.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillside, NJ: Erlbaum.

Luckasson, R., Coulter, D. L., Polloway, E. A., Reiss, S., Schalock, R. L., Snell, M. E., Spitalnick, D. M., & Stark, J. A. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.). Washington, DC: American Association on Mental Retardation.

Luckasson, R., Borthwick-Duffy, S. A., Buntinx, W. H.E., Coulter, D. L., Craig, E. M., Reeve, A., Schalock, R. L., Snell, M. E., Spitalnick, D. M., Spreat, S., & Tassé, M. J. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.

Orlando, M., Sherbourne D. C., & Thissen, D. (2000). Summed score linking using item response theory: application to depression measurement. *Psychological Assessment, 12*, 354–359.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph, No. 17, 34*, Part 2.

Samejima, F. (1997). Graded response model. In W. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.

Schalock, R. L. (1999). The merging of adaptive behavior and intelligence: Implications for the field of mental retardation. In R. L. Schalock (Ed.), *Adaptive behavior and its measurement: Implications for the field of mental retardation* (pp. 43-59). Washington, DC: American Association on Mental Retardation.

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J. Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Gomez, S., Lachapelle, Y., Luckasson, R., Reeve, A., Shogren, K. A., Snell, M. E., Spreat, S., Tassé, M. J., Thompson, J. R., Verdugo Alonso, M. A., Wehmeyer, M. L., & Yeager, M. H. (2010). *Intellectual disability: Diagnosis, classification, and systems of supports* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland-II: Vineland Adaptive Behavior Scales* (2nd ed.). Minneapolis, MN: Pearson Assessments.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293-325.

Takata, Y., Ansai, T., Soh, I., Nakamichi, I., Akifusa, S., Goto, K. Yoshida, A., Fujii, H., Fujiasawa, R., & Sonoki, K. (2013). High-level activities of daily living and disease-specific mortality during a 12-year follow-up of an octogenarian population. *Clinical Interventions in Aging, 8,* 721-728.

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Spreat, S., Thissen, D. T., Widaman, K. F., & Zhang, D. (in press). *Diagnostic Adaptive Behavior Scale: User's Manual.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Spreat, S., Thissen, D. T., Widaman, K. F., & Zhang, D. (2012). The Construct of Adaptive Behavior: Its Conceptualization, Measurement, and Use in the Field of Intellectual Disability. *American Journal on Intellectual and Developmental Disabilities, 117,* 291-303.

Thissen, D., Cai, L., & Bock, R.D. (2010). The nominal categories item response model. In M.L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 43-75). New York, NY: Routledge.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27, 77-83.

Williams, V., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics, 24*, 42-69.

Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-Improved Wald Test for DIF Testing With Multiple Groups Evaluation and Comparison to Two-Group IRT. *Educational and Psychological Measurement*, *73*, 532-547.

World Health Organization. (1992). *ICD-10 Classifications of mental and behavioural disorder: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

Zigman, W. B., Schupf, N., Urv, T., Zigman, A., & Silverman, W. (2002). Incidence and temporal patterns of adaptive behavior change in adults with mental retardation. *American Journal on Mental Retardation, 107,* 161-174.

Table 1.

Demographics of initial calibration and total calibration samples of assessed individuals.

Demographics	Initial Calibration Sample (N=474)	Total Calibration Sample* (N=1,058)
Age	M = 12.0 yrs SD = 4.9 yrs	M = 11.1 yrs SD = 4.9 yrs
Gender	male = 52% female = 48%	male = 50% female = 50%
Race		
Caucasian	73%	72%
African American	15%	17%
Asian	2%	3%
American Indian/Alaska Native	<1%	<1%
Native Hawaiian/Pacific Islander	1%	<1%
Other/Mixed	2%	2%
Not reported	7%	5%
Ethnicity		
Hispanic/Latino	8%	10%
Known Diagnoses		
Non-ID	74%	88%
ID-related Dx	26%	12%
ADHD	11%	8%
ASD	11%	6%
Learning disability	9%	6%
Language impairment	15%	8%
Emotional disturbance	7%	4%
Hearing impairment	3%	2%
Visual impairment	5%	1%
Physical disability	4%	2%
Brain injury	<1%	<1%
Other health impairment	7%	4%
Other impairments	6%	4%

*The final calibration sample included the individuals from the initial calibration sample.

Table 2.

Theta Means and Standard Deviations for the Age Groups, IRT Metric.

Age	Conceptual Skills		Social Skills		Practical Skills	
Group	M	SD	M	SD	M	SD
4-8	-1.26	0.94	-0.45	0.90	-.75	0.89
9-15	0	1	0	1	0	1
16-21	0.37	1.26	0.29	1.21	0.13	0.94

Table 3.

Estimated Means and SDs by age for the non-ID subset of the sample.

	Conceptual Skills		Social Skills		Practical Skills		
Age	M	SD	M	SD	M	SD	*N*
4	-.81	.31	.74	.63	-1.28	.60	78
5	-.47	.41	-.31	.72	-.74	.66	59
6	-.12	.56	-.18	.72	-.61	.75	75
7	.09	.50	-.13	.65	-.50	.64	81
8	.55	.74	.20	.84	-.01	.76	63
9	.64	.51	.10	.82	.00	.88	64
10	.52	.52	-.12	.80	-.30	.79	75
11	.78	.53	.20	.86	.24	.94	63
12	.87	.67	.07	.87	-.03	.80	58
13	1.04	.57	.44	.70	.33	.75	44
14	1.04	.59	.52	.83	.14	.73	41
15	1.28	.71	.62	.89	.67	.81	48
16	1.13	.41	.57	.66	.19	.65	39
17	1.30	.66	.69	.94	.51	.83	45
18	1.23	.60	.82	.84	.63	.87	27
19	1.29	.64	.70	.75	.55	.68	31
20	1.64	.58	1.12	.90	.92	.57	28
21	1.59	.52	1.16	.67	.84	.72	18

Table 4.

Summary of Unique, Linked, and Non-Linked Items on the DABS.

	Form					
	4-8	4-8 & 9-15	9-15	9-15 & 16-21	16-21	Totals
CS	25 items 16 unique	5 link 4 non-link	25 items 7 unique	9 link	25 items 16 unique	57
SS	25 items 14 unique	11 link	25 items 5 unique	9 link 1 non-link	25 items 16 unique	55
PS	25 items 7 unique	18 link	25 items 0 unique	7 link	25 items 18 unique	50

Note: CS = Conceptual Skills; SS = Social Skills; PS = Practical Skills.

Table 5.

Explained Common Variance (ECV) for the DABS Scales.

Adaptive Skill Domain	Age range		
	4-8	9-15	16-21
Conceptual	.85	.88	.94
Social	.88	.87	.90
Practical	.90	.90	.90

Table 6.

LOESS-Smoothed Means and Standard Deviations by Age, IRT Metric.

	Conceptual Skills		Social Skills		Practical Skills		Total	
Age	Mean	SD	Mean	SD	Mean	SD	Mean	SD
4	-1.96	0.44	-0.71	0.68	-1.12	0.62	-1.27	0.51
5	-1.50	0.51	-0.49	0.72	-0.84	0.65	-0.94	0.55
6	-1.10	0.57	-0.31	0.75	-0.60	0.68	-0.67	0.59
7	-0.75	0.63	-0.18	0.78	-0.39	0.71	-0.44	0.62
8	-0.46	0.67	-0.08	0.80	-0.22	0.74	-0.25	0.65
9	-0.23	0.71	-0.03	0.83	-0.10	0.76	-0.12	0.68
10	-0.06	0.74	0.00	0.86	0.00	0.79	-0.02	0.70
11	0.10	0.75	0.05	0.85	0.07	0.79	0.08	0.70
12	0.24	0.75	0.12	0.83	0.14	0.78	0.17	0.69
13	0.40	0.72	0.25	0.81	0.22	0.72	0.29	0.66
14	0.54	0.70	0.38	0.78	0.30	0.67	0.41	0.63
15	0.63	0.70	0.49	0.80	0.35	0.65	0.49	0.63
16	0.70	0.72	0.56	0.82	0.39	0.64	0.55	0.64
17	0.77	0.74	0.63	0.84	0.43	0.63	0.61	0.65
18	0.85	0.76	0.71	0.84	0.48	0.61	0.68	0.64
19	0.93	0.77	0.81	0.83	0.53	0.60	0.76	0.63
20	1.03	0.77	0.91	0.81	0.59	0.58	0.84	0.61
21	1.12	0.77	1.02	0.79	0.64	0.56	0.93	0.59