The Diagnostic Adaptive Behavior Scale: Evaluating its diagnostic sensitivity and specificity

ARTICLE in RESEARCH IN DEVELOPMENTAL DISABILITIES · AUGUST 2014
Impact Factor: 4.41 · DOI: 10.1016/j.ridd.2014.07.032 · Source: PubMed

9 AUTHORS, INCLUDING:

Giulia Balboni
Università di Pisa
36 PUBLICATIONS 149 CITATIONS

Scott Spreat
Woods Services, Inc.
74 PUBLICATIONS 771 CITATIONS

Keith Widaman
University of California, Davis
174 PUBLICATIONS 9,896 CITATIONS

Patricia Navas
University of Zaragoza
19 PUBLICATIONS 44 CITATIONS

Available from: Giulia Balboni
Retrieved on: 31 July 2015
The Diagnostic Adaptive Behavior Scale: Evaluating its diagnostic sensitivity and specificity

Giulia Balboni a,*, Marc J. Tassé b, Robert L. Schalock c, Sharon A. Borthwick-Duffy d, Scott Spreat e, David Thissen f, Keith F. Widaman g, Dalun Zhang h, Patricia Navas b

a University of Pisa, Via Roma, 67, 56126, Pisa, Italy
b The Ohio State University, 1581 Dodd Dr., Columbus, OH 43210, USA
c Hastings College, P.O. Box 285, Chewelah, WA 99109-0285, USA
d University of California – Riverside, 20500 Yate Circle, Riverside, CA 92508-3126, USA
e Woodland Center for Challenging Behaviors, 9 Imlaystown-Hightstown Road, Allentown, NJ 08501-2011, USA
f University of North Carolina – Chapel Hill, CB #3270, Chapel Hill, NC 23599-3270, USA
g University of California – Davis, One Shields Avenue, Davis, CA 95616-8686, USA
h Texas A&M University, 4225 TAMU, College Station, TX 77843-4225, USA

1. Introduction

All current diagnostic and classification systems ([Diagnostic and Statistical Manual of Mental Disorders, DSM-5; American Association on Intellectual and Developmental Disabilities’ [AAIDD] Terminology & Classification Manual; and International Statistical Classification of Diseases and Related Health Problems, ICD-10] have three common criteria to rule-in or rule-out a diagnosis of Intellectual Disability (ID): (a) significant limitations in intellectual functioning, (b) significant limitations in adaptive behavior, and (c) an age of onset during the developmental period (American Psychiatric Association [APA] 2013; Schalock et al., 2010; World Health Organization, 1993).
Adaptive behavior has been formally included in defining the condition of ID for at least the past 50 years (see Heber, 1959, 1961; Schalock et al., 2010). As discussed in Tassé et al. (2012), we have today essentially the same definition of adaptive behavior as originally proposed by Heber (1959) where adaptive behavior was defined as the collection of conceptual, practical, and social adaptive skills. Both the AAIDD (Schalock et al., 2010) and the DSM-5 (APA, 2013) define adaptive behavior similarly. Moreover, there is agreement in defining significant limitations in adaptive behavior as performance that is approximately two standard deviations below the mean (i.e., represented by a standard score of approximately 70 or less) in at least one of the three adaptive behavior domains of conceptual, social, or practical adaptive skill or in a total score from an adaptive behavior scale (APA, 2013; Schalock et al., 2010).

With the advent of the intelligence tests movement, Intelligence Quotient (IQ) scores quickly became the predominant criterion for diagnosing persons as having ID as well as planning interventions and services for this population, overshadowing consideration of the person’s adaptive behavior skills and deficits. Perhaps due to debate over the precise structure of adaptive behavior (Bruininks, McGrew, & Maruyama, 1988; Harrison, 1989; Meyers, Nihira, & Zetlin, 1979; Widaman, Bornick-Duffy, & Little, 1991; Widaman & McGrew, 1996) and even its conceptualization (e.g., Greenspan & Granfield, 1992), the importance of the role of IQ scores increased. However, since 2002 (Luckasson et al., 2002), the relevance of the construct of adaptive behavior and the role it should play in a diagnosis of ID has clearly been reconfirmed.

Despite being an integral part of diagnosing intellectual disability, a relative paucity of research studies have investigated the accuracy with which standardized adaptive behavior measures accurately identify persons with ID as having significant limitations in adaptive behavior. A key issue in such studies in identifying those behaviors that best distinguish individuals with ID from those without ID. Although more than 200 adaptive behavior measures have previously been identified (Schalock, 1999), only four of them are normed on a representative U.S. sample of the general population, and few have been developed specifically for the purpose of ruling in/out a diagnosis of ID (Tassé et al., 2012). The four instruments normed on representative US samples are: (1) Vineland Adaptive Behavior Scales—2nd edition, VABS-II (Sparrow, Cicchetti, & Balla, 2005); (2) Adaptive Behavior Assessment System—II, ABAS-II (Harrison & Oakland, 2003); (3) Scales of Independent Behavior—Revised, SIB-R (Bruininks, Woodcock, Weatherman, & Hill, 1996); and (4) Adaptive Behavior Scale—School Version, ABS-S:2 (Lambert, Nihira, & Leland, 1993). Some of these (e.g., ABS & SIB-R) have not been re-normed in 2 decades.

When reviewing the psychometric properties of these measures, accuracy to correctly identify persons with ID from individuals without ID is a key aspect. The accuracy of differential diagnosis may be evaluated comparing the standard scores of assessed individuals with and without and ID diagnosis and computing diagnostic efficiency (or validity) statistics (e.g., Streiner, 2003) such as sensitivity and specificity. Sensitivity is defined as the proportion of true positives (i.e., person having ID) that are correctly identified as having ID by the test (Altman & Bland, 1994), reflecting how good a test is at correctly identifying people who have the condition the test is intended to measure (Loong, 2003). Specificity refers to the ability to correctly reject a diagnosis or identify true negatives, when the person is known not to have the said diagnosis (Altman & Bland, 1994).

Information regarding the sensitivity and specificity of an adaptive behavior assessment instrument is critical in establishing the valid use of the instrument. For example, in the case of the ABS-S:2, the authors stated simply that “mean scores for the groups with developmental disabilities are sufficiently below the average scores of the normal group” (Lambert et al., 1993, p. 50). However, the actual percentages of people correctly classified were not provided, so the precision of the ABS-S:2 to accurately measure significant limitations in adaptive behavior and correctly identify someone as having an ID is not known.

The SIB-R Comprehensive Manual (Bruininks et al., 1996) provided a more detailed description of sensitivity and specificity. In the manual, the authors stated that 76% of the individuals within the standardization sample were correctly classified into their original groups (51% mild ID; 74% moderate ID; and 82% non-ID). From a decision-making perspective, the fact that the lowest degree of accuracy (i.e., 51%) of the SIB-R is for the group with mild ID is not surprising, as this group is the closest to the actual diagnostic cut-off score. However, if only about half of persons diagnosed with mild ID have scores in a range that would support diagnosis, the SIB-R may not be sufficiently sensitive to identify correctly persons with mild forms of ID. Furthermore, although the SIB-R covers the age span (from infancy and up), sensitivity and specificity data were not provided for separate age groups.

For the VABS-II (Sparrow et al., 2005), the percentages of people correctly classified ranged from 71% to 100% for those aged 6–18 years old (71% mild ID; 87% moderate ID; and 100% severe and profound ID), and from 97% to 100% for individuals aged 19–86 (97% mild ID and 100% for individuals with moderate and with severe and profound ID). Data regarding the sensitivity of the VABS-II were also available by domain (i.e., communication, daily living skills, socialization, and motor skills). However, the proportion of people who were correctly excluded from the diagnosis of ID (i.e., specificity) was not reported.

The ABAS-II (Harrison & Oakland, 2003) offers data on both sensitivity and specificity across four infant-preschool samples, five school-age samples, and one adult sample. The results for the infant-preschool samples indicate that 58–73% of those with ID (sensitivity), and 0–5% without ID (specificity) scored at least 2 standard deviations below the mean on the General Adaptive Composite (GAC) score, whereas 77–86% children diagnosed with ID and 3–19% of the matched control group scored at least 2 standard deviations below the mean on one or more adaptive domains or the GAC. Across the five school-age samples and the adult sample 50–87% of those with ID and 0–17% of the matched control group scored at least 2 standard deviations below the mean on the GAC, and 62–100% with ID and 5–19% without ID scored at least 2 standard deviations below the mean on one or more adaptive domains or the GAC.
Although the four standardized measures of adaptive behavior briefly reviewed above are appropriate instruments for making a determination of ID, all were generally developed using Classical Test Theory (CTT) models (Tasse et al., 2012). In distinction, the Diagnostic Adaptive Behavior Scale (DABS; Tasse et al., in preparation) used an Item Response Theory (IRT) model. Unlike CTT, IRT models provide guidance in test development and allow scale developers to tailor the efficiency of the instrument for a specific ability level of the trait that is measured. For example, given the intended use of an adaptive behavior to provide accurate and precise scores near the cutoff point for diagnosis (i.e., 2 SD below the mean), an IRT approach to scale construction can build up precision at a given point on the standard score scale, which should yield more sensitive measures for diagnostic decisions (Hays, Morales, & Reise, 2000). Moreover, IRT models provide a better estimate of the individual’s true score (Santor & Ramsay, 1998), which is described as a function of the trait (e.g., adaptive behavior) being measured and test item parameters (Thissen, Nelson, Rosa, & McLeod, 2001), such as each item’s level of difficulty and discrimination strength.

2. Study purpose

The purpose of this study was to evaluate the DABS with regard to its sensitivity and specificity. The DABS is an adaptive behavior assessment instrument that was developed using IRT (Graded Response Model for ordinal polytomous data [Samejima, 1969]), focusing its item pool on the items that provide the most precise information for the purpose of making a diagnosis of ID. IRT models were used to estimate the person’s adaptive functioning and item parameters to identify a final set of items that provide the most information around the cutoff point for determining significant limitations in adaptive behavior (i.e., approximately 2 standard deviations below the population mean) from 4 to 21 years old (Tasse et al., in preparation).

The goals of the present paper include:

(1) comparing the DABS standard score of assessed individuals with and without an ID diagnosis and determining sensitivity and specificity of the DABS to correctly identify persons with an ID diagnosis from individuals who do not have an ID diagnosis; and

(2) evaluating the sensitivity and specificity across age groups: 4–21 years old.

3. Method

3.1. Participants

The participants comprised individuals assessed on the DABS during the standardization phase. The sample was composed of 1058 persons, aged between 4 and 21 years old (M = 11.1, SD = 4.9), and 51% were male. Individuals in the standardization sample came from 46 American states, with none from Alaska, Arkansas, North Dakota, or West Virginia, and 71.5% were white.

As can be seen in Table 1, 125 individuals (12%) had a formal diagnosis of mental retardation/ID or developmental delay and were classified as “participants with an ID-related diagnosis” or the “ID group.” The remaining 933 (88%) were reported as not having any significant deficits in intellectual functioning and adaptive behavior. Of these individuals, 20% had another verified condition that was different from an ID-related diagnosis (e.g., ADHD, learning disability, etc.). These individuals were classified as “participants without an ID-related diagnosis” or the “non-ID group.”

The participants were divided into three age groups: 4–8, 9–15, and 16–21 years old (Table 2). In each age group, about 50% were male, and 8%, 9%, and 21%, respectively, had a formal ID-related diagnosis. No significant age or gender differences

<table>
<thead>
<tr>
<th>Condition</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID-related diagnosis</td>
<td>125</td>
<td>12</td>
</tr>
<tr>
<td>Non-ID</td>
<td>933</td>
<td>88</td>
</tr>
<tr>
<td>Typically developing</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>Other verified conditions</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Prevalence (non-cumulative)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Autism Spectrum Disorder</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Learning disability</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Language impairment</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Emotional disturbance</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Hearing impairment</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Visual impairment</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Other health impairment</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
between the ID and non-ID groups were found, with two exceptions: (a) a gender difference (more male participants) in the ID versus the non-ID participants in the 9–15 year old group, $\chi^2(1) = 6.92, p < .01$; and (b) an age difference in the 16–21 year old group, in which participants with ID were slightly older than their peers without ID, $t(236) = 1.99, p < .05$.

3.2. Instrument

The DABS was developed over a 6-year period with the aim of making available an adaptive behavior test that is focused on providing diagnostic information for ID. ID was defined as a condition that is: “... characterized by significant limitations in both intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before age 18” (Schalock et al., 2010, p. 1). The DABS reflects both the theoretical and empirical work done over the last two decades, which has supported the 3-factor structure of adaptive behavior as defined in Schalock et al. (2010) as: “the collection of conceptual, social, and practical skills that have been learned by people in order to function in their everyday lives” (Schalock et al., 2010, p. 15) (Arias, Verdugo, Navas, & Gómez, 2013; Navas, Verdugo, Arias, & Gómez, 2012).

The DABS administration is done via a face-to-face interview between a professional and a respondent who knows the assessed person very well. The individual’s performance of each adaptive skill is assessed according to a four-point rating scale from: 0 (rarely or never does it) to 3 (does it always or almost always independently) and all the items on the DABS must receive a rating. The interviewer of the DABS is described as a professional (e.g., psychologist, teacher, case manager, social worker) or other professional who has completed at least a Bachelor’s degree, has had direct work experience with people with an intellectual or closely related developmental disability, and who has had previous individual assessment experience. The DABS should be completed with as many respondents as the interviewer deems necessary to provide the most valid and complete assessment of the assessed person’s adaptive behavior. Respondents should inform about individual’s typical performance during daily routines and changing circumstances.

As described further in Tasse et al. (submitted for publication), IRT analyses yielded a DABS final version with a total of 75 items for each of the three age groups (4–8; 9–15; 16–21 years old), consisting of 25 items per domain (conceptual, social, and practical skills). Domain scores for each of the three domains as well as an Overall Adaptive Behavior Standard Score (OABSS) are reported on a standardized scale with a mean of 100 and standard deviation of 15. Therefore, the “2 SDs below the mean” cut-off score for the diagnosis of ID is equal to a standard score of approximately 70. However, the evaluator should also consider the variability of the individual’s observed score caused by several potential sources of measurement errors. The average SEM of the DABS is equal to approximately ±3 points for each domain standard score (i.e., conceptual, social, and practical skills) and ±2 points for the OABSS. Using a 95% confidence interval, the ID diagnosis cut-off represents domain standard score that might go up to 76 (70 + 2 SEM) and 74 (70 + 2 SEM) for the OABSS.

3.3. Procedure

Data collection for the DABS standardization was conducted between 2008 and 2011. Participants were recruited nationally through several waves using electronic media, listservs, mailings, and web postings from various professional societies and national disability groups, including: AAIDD, Association of University Centers on Disabilities (AUCD), National Association of School Psychologists (NASP), and APA – Division 33, to name a few.

DABS accuracy in the measurement of significant limitations in adaptive behavior was investigated using: (1) comparison of the standard score of assessed individuals with and without an ID-related diagnosis on each domain (i.e., conceptual, social, and practical skills) and the OABSS; and (2) estimation of sensitivity and specificity in correctly identifying assessed individuals with and without an ID-related diagnosis based on their DABS scores (i.e., a standard score on one of the three domains or the OABSS that was at or below 76 or 74, respectively).
3.4. Analyses

Comparisons of standard scores were performed using t-tests. When statistically significant differences were found, Cohen’s d was computed (Cohen, 1988), and the effect sizes were interpreted according to the following criteria: negligible (<.20), small (.20–.49), medium (.50–.79), and large (.80+) effect size.

Test sensitivity was calculated as the proportion of individuals with an ID-related diagnosis who were correctly identified by the DABS (i.e., who obtained a standard score on one of the three domains that was at or below 76 or an OABSS that was at or below 74). Test specificity was calculated as the proportion of individuals without an ID-related diagnosis who did not have a domain standard score at or below 76 or an OABSS at or below 74.

One can combine the sensitivity and specificity into a single statistic called Overall Correct Classification (OCC). The OCC is defined as the proportion of all individuals with and without an ID-related diagnosis who were correctly identified and detected by the DABS, respectively (e.g., McFall & Treat, 1999; Streiner, 2003).

If expressed in percentage form, sensitivity, specificity, and OCC are determined to be statistically significant if their CIs do not include the value of 50% (i.e., if they are more accurate than a random classification). Matthey and Petrovski (2002) suggested that sensitivity coefficients ≥70% and specificity coefficients ≥80% are considered appropriate benchmarks to be attained by diagnostic tests. Higher values of sensitivity, specificity, and OCC are always desirable.

Another way to measure sensitivity and specificity in discriminating participants with and without ID based on their DABS scores was estimated using Receiver Operating Characteristic (ROC) curves analyses. The ROC is computed by plotting the true positive rate (sensitivity) against the false positive rate (1 – true negatives or 1 – specificity) for different possible cut points on the DABS (e.g., Altman & Bland, 1994; McFall & Treat, 1999; Streiner, 2003). The Area Under the Curve (AUC) is equal to the probability of a test giving a correct diagnosis (i.e., correctly identifying individuals with and without ID). In other words, the AUC reflects the probability that a randomly selected person from the ID group will have a standard score below the standard score of a randomly selected person from the non-ID group. AUC values range from 0 to 1 and are statistically significant if their CIs do not include the value of .5. Swets (1988) suggested the following parameters when interpreting AUC values: .5–.6 = fail; .6–.7 = poor; .7–.8 = fair; .8–.9 = good; and .9–1.0 = excellent. For each age group, comparisons within the AUCs of the standard scores for the DABS Overall Adaptive Behavior and the conceptual, social, and practical domains were confirmed by analyzing the presence or absence of overlaps between the corresponding CIs.

4. Results

4.1. Comparisons of DABS standard score obtained by participants with and without an ID-related diagnosis

As shown in Table 3, the mean OABSS for those within the ID group aged 4–21 years old was significantly below (p < .001) the mean OABSS of the assessed individuals within the non-ID group aged 4–21 years old (with effect sizes greater than 2.20, meaning that the difference is larger than two standard deviations). As expected (see Fig. 1), all the individuals with an ID-related diagnosis, with the exception of those who were 10 years old, had a mean OABSS below the ID diagnosis cut-off standard score of 74. The mean OABSS was approximately 100 for those without an ID-related diagnosis, regardless of their age range. Standard deviations (see Table 3) were approximately 15 for the non-ID groups (i.e., 4–8; 9–15; 16–21 years old).

Table 3
Comparisons of the standard score obtained by the participants of each of the three age groups with and without ID-Related diagnosis on the domains and on the Overall Adaptive Behavior scale: standard score Mean (SD), t-test and Cohen’s d values.

<table>
<thead>
<tr>
<th>Age groups</th>
<th>ID (n = 32)</th>
<th>Non-ID (n = 356)</th>
<th>t value (Cohen’s d)</th>
<th>ID (n = 42)</th>
<th>Non-ID (n = 390)</th>
<th>t value (Cohen’s d)</th>
<th>ID (n = 51)</th>
<th>Non-ID (n = 187)</th>
<th>t value (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age groups</td>
<td>Conceptual</td>
<td>Social</td>
<td>Practical</td>
<td>OABSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–8</td>
<td>64.22</td>
<td>73.27</td>
<td>72.72</td>
<td>66.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(22.39)</td>
<td>(15.27)</td>
<td>(19.23)</td>
<td>(17.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9–15</td>
<td>64.88</td>
<td>72.10</td>
<td>72.16</td>
<td>65.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.69)</td>
<td>(16.88)</td>
<td>(10.39)</td>
<td>(19.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15.04)</td>
<td>(15.09)</td>
<td>(15.17)</td>
<td>(15.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16–21</td>
<td>54.70</td>
<td>63.73</td>
<td>60.88</td>
<td>53.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.44)</td>
<td>(15.99)</td>
<td>(15.27)</td>
<td>(16.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15.51)</td>
<td>(15.27)</td>
<td>(15.45)</td>
<td>(15.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. All the differences are statistically significant (p < .001); OABSS, Overall Adaptive Behavior standard score.
participants with and without an ID-related diagnosis. This figure shows that, generally, the standard scores for individuals with an ID-related diagnosis fell under the ID cut-off score of 76. The only exceptions noted were for 4, 5, 9, and 10 years old participants for the Social skills domains and for the 4, 10 and 15 years old participants for the Practical skills domain, for which the mean standard score was higher than the cut-off point. However, for these age groups the mean OABSS was generally below the ID cut-off point.

4.2. Sensitivity and specificity of the DABS

False negatives (i.e., participants with an ID-related diagnosis who obtained DABS standard scores above the ID diagnosis cut-off point) and false positives (i.e., participants without an ID-related diagnosis who obtained DABS standard scores below the ID diagnosis cut-off point), sensitivity, specificity, and OCC are reported in Table 4. The DABS yielded sensitivity coefficients ranging from 81% to 98%, specificity coefficients ranging from 89% to 91%, and OCC coefficients ranging from 89% to 92%. Their CIs did not include the value of 50% (i.e., they were statistically significant) for any age group. Sensitivity and specificity were higher than the clinical significance value of 70% (except one value of 68%) and 80%, respectively.

![Fig. 1. DABS Overall Adaptive Behavior standard score of participants with and without an ID-related diagnosis (ID diagnosis cut-off standard score is 74 standard score).](image1)

![Fig. 2. DABS conceptual, social and practical skills standard score of participants with and without an ID-related diagnosis (ID diagnosis cut-off standard score is 76 standard score).](image2)