



ELSEVIER

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Italian Diagnostic Adaptive Behavior Scale: Reliability and diagnostic accuracy compared with the Vineland-II

Giulia Balboni ^{a,*}, Alice Bacherini ^a, Pasquale Anselmi ^b, Paola Brovedani ^c,
Serafino Buono ^d, Serena Micheletti ^{e,f}, Egidio Robusto ^b, Marc J. Tassé ^g

^a Department of Philosophy, Social Sciences and Education, University of Perugia, Piazza G. Ermini, 1, 06123 Perugia, Italy

^b Department of Philosophy, Sociology, Education and Applied Psychology, University of Padua, Via Venezia, 14, 35131 Padua, Italy

^c IRCCS Stella Maris Foundation, Viale del Tirreno, 331, Calambrone, 56128 Pisa, Italy

^d Oasi Research Institute-IRCCS, Via Conte Ruggero, 73, 94018 Troina, Italy

^e Unit of Child Neurology and Psychiatry, ASST Spedali Civili di Brescia, Piazzale Spedali Civili, 1, 25123 Brescia, Italy

^f Department of Clinical and Experimental Sciences, University of Brescia, Viale Europa, 11, 25123 Brescia, Italy

^g The Ohio State University Nisonger Center, 357G McCampbell Hall, 1581 Dodd Drive, Columbus, OH 43210, USA

ARTICLE INFO

No. of reviews completed is 2

Keywords:

Adaptive behavior
Diagnostic Adaptive Behavior Scale
Diagnosis
Sensitivity
Specificity
Adaptation
Vineland-II

ABSTRACT

Background: The Diagnostic Adaptive Behavior Scale (DABS) is a short scale with excellent properties to assess the conceptual, social, and practical adaptive behavior domains for the diagnosis of intellectual disability (ID) in individuals aged 4–21 years.

Aims: Investigate the test-retest and inter-respondent reliability of the Italian adaptation of the DABS, verify its diagnostic accuracy in identifying individuals with ID and excluding individuals with typical development (TD), and compare its psychometric properties to those of the Vineland-II.

Methods: Test-retest reliability: The same respondent completed the Italian DABS for the same assessed person at two separate times ($n = 71$). Inter-respondent reliability: Two respondents for the same assessed person completed the Italian DABS independently ($n = 57$). Diagnostic accuracy: The same respondent completed the Italian DABS and Vineland-II for the same assessed person ($n = 378$; 50 % ID, 50 % TD).

Results: Italian DABS test-retest and inter-respondent correlation coefficients were excellent. Italian DABS sensitivity was 86 % and specificity was 99 %, Italian DABS Areas Under the ROC Curves were excellent (or good, practical skill domain), and comparable to the results reported for the Vineland-II.

Conclusions: The Italian DABS is an excellent measure to evaluate the adaptive behavior for ID diagnosis; it is comparable to the Vineland-II but being shorter, the Italian DABS requires less time to administer.

What this paper adds?

The present study is innovative in presenting the Italian version of DABS and its psychometric properties. The DABS is focused on

* Corresponding author at: Department of Philosophy, Social Sciences and Education, University of Perugia, Piazza G. Ermini, 1, 06123 Perugia, Italy.

E-mail address: giulia.balboni@unipg.it (G. Balboni).

<https://doi.org/10.1016/j.ridd.2022.104185>

Received 25 September 2021; Received in revised form 21 December 2021; Accepted 27 January 2022

0891-4222/© 2022 Elsevier Ltd. All rights reserved.

assessing the three adaptive behavior domains (conceptual, practical, and social skills) to verify if an individual aged 4–21 years presents the significant limitations in adaptive behavior, a necessary criterion for the diagnosis of the intellectual disability (ID). Therefore, we estimated the Italian DABS ability to correctly identify individuals with ID and correctly exclude individuals with typical development (i.e., no ID). Moreover, we compared the Italian DABS diagnostic accuracy with that of the Vineland-II that were administered in a counterbalanced order with the same participants. We found that the Italian DABS has excellent diagnostic accuracy, comparable to that of the Vineland-II. However, the DABS is a shorter scale than the Vineland-II and its administration is easier and less time-consuming. We also found that the Italian DABS normative scores were stable (test-retest reliability) and with good concordance between respondents (inter-respondent reliability). Therefore, the Italian DABS seems to be an excellent scale to evaluate the adaptive behavior for the purpose of diagnosing ID. The DABS original version also has very strong psychometric properties (Balboni et al., 2014; Tassé, Schalock, Balboni et al., 2016; Tassé, Schalock, Thissen et al., 2016). With the present study, we have shown the utility of the DABS also outside the USA.

1. Introduction

Adaptive behavior refers to the skills that are learned throughout life and performed to meet the social and personal responsibilities expected of us and our same age peers from the community settings in which we live, learn, work, and recreate (Schalock et al., 2021). Based on the tripartite model of adaptive behavior, it refers to the three domains of conceptual, social, and practical skills (American Psychiatric Association [APA], 2013; Schalock et al., 2021; World Health Organization [WHO], 2021). Adaptive behavior assessment has long been an essential component for the diagnosis of intellectual disability (ID) (see APA, 1968; Heber, 1961). We have come a long way since the publication of the Vineland Social Maturity Scale (VSMS; Doll, 1936, 1953). The number of standardized scales of adaptive behavior has proliferated since the publication of the VSMS. Schalock (1999) estimated that there were more than 200 published adaptive behavior scales available. However, Tassé et al. (2012) pointed out that the vast majority of these scales were developed to assess discrete abilities of adaptive behavior (e.g., social skills, work, toilet training) and not the overall adaptive behavior construct. Other scales were normed on a sample that was not representative of the general population. Hence, few of these scales were actually considered to be comprehensive, robustly standardized, and normed on the general population to the point of being adequate to be used in the process of making a determination of ID.

Tassé and Balboni (2021) reviewed five standardized measures of adaptive behavior that they considered to be part of the current “gold standard” measures of adaptive behavior, that should be selected when assessing adaptive behavior for the purpose of ruling-in or ruling-out a diagnosis of ID: the Adaptive Behavior Assessment System – Third Edition (ABAS-3; Harrison and Oakland, 2015), the Adaptive Behavior Diagnostic Scale (ABDS; Pearson et al., 2016), the Scales of Independent Behavior, Revised (SIB-R; Bruininks et al., 1996), the Vineland Adaptive Behavior Scales Third Edition (Vineland-3; Sparrow et al., 2016), and the Diagnostic Adaptive Behavior Scale (DABS; Tassé et al., 2017). All these measures, however, have been developed and normed on a USA population. Since the assessment of adaptive behavior is considered a universal construct, much like intellectual functioning, and is also included in the WHO’s International Classification of Diseases (ICD; WHO, 2021) diagnostic criteria for diagnosing ID, it is essential to translate and adapt these measures for use in other populations.

Because recommended practice in the assessment of adaptive behavior includes ensuring cultural appropriateness of the skills assessed (Schalock et al., 2021), proper translation and adaptation are critical to ensuring the valid use of any adaptive behavior scale. Test adaptation is defined as transforming a test from use in one language and culture to another (International Test Commission [ITC], 2017). The ITC has continued to put forth clear methodological guidelines to be followed when adapting a test for use in a different language and culture. Translating a test is frequently the first component in the adaptation of a test. Tassé and Craig (1999) recommended using a committee-approach over the translation/back-translation approach to translating the item stems of a test. The committee members also consider potential cultural adaptations to item content to preserve the intent of the item stem. An important remaining step recommended by ITC for adapting a test includes checking the psychometric properties of the adaptation, including assessing its reliability and validity.

For the purposes of making a determination of ID, the assessment process includes three prongs: (a) assessing deficits in intellectual functioning, (b) assessing deficits in adaptive behavior, and (c) establishing that deficits in intellectual functioning and adaptive behavior, if present, originated during the developmental period (APA, 2013; Schalock et al., 2021; WHO, 2021). Furthermore, the DSM-5 and American Association on Intellectual and Developmental Disabilities (AAIDD) have both operationally defined “deficits in adaptive behavior” as the presence of significant limitations in one or more of the three adaptive behavior domains, i.e., conceptual, social, or practical skills, or on the overall adaptive behavior level. The significant limitations in adaptive behavior are operationalized in a normative score, i.e., a standard score, that is two standard deviations below the population mean. Adaptive behavior scales typically use normative scores with mean = 100 and standard deviation = 15; therefore, deficits in adaptive behavior are present with a normative score of approximately ≤ 70 . Hence, the reliability and validity of measures of adaptive behavior must be focused on the psychometric properties of these three domain scores.

For adaptive behavior scales, it is crucial to verify the diagnostic accuracy that may be defined as the ability to correctly identify individuals with a prior diagnosis of ID and correctly exclude individuals without a prior diagnosis of ID. The diagnostic accuracy is generally assessed by computing the sensitivity and specificity of the scale. Sensitivity (i.e., the true positive prevalence) represents the proportion of individuals with a diagnosis of ID who were correctly identified by the adaptive behavior scale (i.e., who obtained a normative score approximately ≤ 70 on one or more of the three adaptive behavior domains or on the overall adaptive behavior level). Test specificity (i.e., the true negative prevalence) represents the proportion of individuals without a diagnosis of ID who were correctly identified by the adaptive behavior scale (i.e., who did not have a domain or the overall adaptive behavior level normative

score approximately ≤ 70).

Adaptive behavior scales are typically administered to third-party respondents who provide ratings that are based on their direct observations of the assessed person's typical performance on all the adaptive behavior scale items. Thus, estimating the concordance between respondents is vital. In addition, it is important to assess score stability, using a test-retest procedure, as well as the scale's validity when evaluating the psychometric properties of any new or adapted test (American Educational Research Association et al., 2014).

The aim of this study was to assess the psychometric properties of the Italian adaptation of the DABS (Balboni et al., in preparation; Tassé et al., 2017), including estimating the test-retest reliability, inter-respondent reliability, and its sensitivity and specificity in the determination of ID compared to the Vineland-II. The DABS was selected because its primary goal is to diagnose ID in individuals 4–21 years of age. The DABS was specifically developed using the item response theory and the tripartite model of adaptive behavior (conceptual, social, and practical skills). The DABS consists of 75 items across conceptual, social, and practical skills (each domain consisting of 25 items) measuring the three adaptive behavior domains and yielding the Total Score as an overall adaptive behavior assessment. Three age forms of the DABS are available for individuals 4–8, 9–15, and 16–21 years of age. The DABS is administered via an interview between a professional and a respondent who knows the assessed person very well. The DABS has shown excellent reliability, validity, and diagnostic accuracy (for more information see: Balboni et al., 2014; Tassé, Schalock, Balboni et al., 2016; Tassé, Schalock, Thissen et al., 2016).

2. Study 1: test-retest and inter-respondent reliability of the Italian DABS

To investigate the test-retest reliability (i.e., the stability of the DABS scores over time) we had the same respondent complete the Italian DABS on the same assessed person's adaptive behavior at two separate time points. We investigated the inter-respondent reliability (i.e., the degree of concordance between two independent respondents' scores) by having two different respondents who knew well the same assessed person complete the Italian DABS independently.

2.1. Material and methods

2.1.1. Participants

The test-retest reliability was investigated for 71 individuals (52 % female, $n = 37$) between the ages of 4–21 years ($M [SD] = 13.41$ years [5.30]). All assessed participants had a typical development except for five individuals (7 %) who had mild ($n = 3$) or moderate ($n = 2$) ID. Participants with typical development were recruited using the snowball sampling technique, and participants with ID were recruited through the Italian Association of Individuals with Down Syndrome (AIPD). Participants came from a broad geographic area of Italy, and lived mostly in the central regions (51 %) and southern regions (48 %) of Italy.

The inter-respondent reliability was investigated for 56 individuals (55 % male, $n = 31$) between the ages of 4–21 years ($M [SD] = 13.48$ years [5.45]). All assessed participants had a typical development and were recruited using the snowball sampling technique. Participants recruited for the inter-respondent reliability study also came from a broad geographic area of Italy, and lived mostly in the central regions (39 %) and southern regions (52 %) of Italy.

Seventeen assessed participants took part in both studies (i.e., 24 % of participants of the test-retest reliability study and 30 % of participants of the inter-respondent reliability study). Table S1 of the Supplementary Material reports the number of participants by chronological age who participated in the study investigating the test-retest reliability and inter-respondent reliability of the DABS.

2.1.2. Instruments

The Italian translation and adaptation of the DABS (Balboni et al., in preparation) was used to measure the individual's adaptive behavior. It was developed following the procedures recommended by Tassé and Craig (1999) and the ITC Guidelines for Translating and Adapting Test (ITC, 2017). Additionally, similarly to the original version, the item calibration and standardization of the Italian DABS was conducted using a two-parameter item response theory model (Tassé et al., 2017; Tassé, Schalock, Thissen et al., 2016). As for the original DABS, the individual's performance on each adaptive skill item is assessed according to a four-point rating: 0 (rarely or never does it), 1 (does it with reminders or assistance but rarely or never independently), 2 (does it sometimes independently but sometimes needs reminders or assistance), and 3 (does it always or almost always independently, never or rarely needs reminders or assistance). There is also a NS (no score) rating that is permitted when the respondent has no direct knowledge of the individual's typical performance on a specific item/skill being assessed or the environment does not permit the assessed person the opportunity to perform the skill. However, the NS rating should be used infrequently when the respondent knows the assessed individual well. All the items on the DABS must receive a rating (for more information, see Tassé et al., 2017). The normative scores ($M = 100$; $SD = 15$) may be computed for each of the conceptual, social, and practical skills domain scores and for the Total Score, given the individual's raw score.

The Italian DABS showed strong psychometric properties (Balboni et al., in preparation). The internal consistency for the calibration group ($n = 799$) was "excellent" based on the criteria suggested by Cicchetti et al. (2011), with ordinal Cronbach alpha coefficients (Zumbo et al., 2007), and were computed independently for the three DABS age forms and separately for each adaptive behavior domain (i.e., the 25 items of each conceptual, social, and practical skills) and the Total Score (i.e., 75 items). The conceptual skills domain coefficients of .98 were found in all the three DABS age forms; for the social skills domain, coefficients of .96 were detected in 4–8 and 9–15 years old DABS age forms and of .97 in the 16–21 DABS age form; for the practical skills domain, coefficients of .96 were found in the 4–8 and 16–21 DABS age forms and of .95 in the 9–15 DABS age form; and for the Total Score, coefficients of

.99 were detected in the 9–15 and 16–21 DABS age forms and of .98 in the 4–8 DABS age form. The construct validity was ascertained using the standardization group ($n = 577$) verifying the presence of moderate to large intercorrelations (Cohen, 1988) between the Italian DABS domain scores (.40–.61). The convergent/divergent validity with the Italian version of the Vineland Adaptive Behavior Scales - Second Edition (Vineland-II; Sparrow et al., 2005; It. ad. Balboni et al., 2016) was ascertained using a group of 378 individuals with typical development ($n = 189$) or ID ($n = 189$). For each Italian DABS domain the correlation coefficients with the corresponding Italian Vineland-II domain were higher than the correlation coefficients with the other Italian Vineland-II domains. The discriminant validity was verified investigating, for the calibration group, that (a) individuals with ID obtained statistically significant lower Italian DABS scores than individuals with typical development, and (b) the Italian DABS scores increased through each year of participants' chronological age (i.e., 4–21 years old), reflecting the presence of the developmental trend of the adaptive behavior construct.

2.1.3. Procedure

For the test-retest reliability study, the Italian DABS was administered on two separate occasions (Time 1 and Time 2), with the same interviewer interviewing the same respondent approximately one month apart ($M = 29.70$; $SD = 3.76$; $Med = 30.42$; range = 20–38 days). The respondents were the mother (84 %, $n = 60$), father (13 %, $n = 9$), or a sibling (3 %, $n = 2$) of the assessed individual. The interviewers ($n = 6$, 100 % females) were undergraduate (17 %) or graduate (83 %) students in psychology (number of Italian DABS administrations for each interviewer: $M = 11.83$; $SD = 6.68$; $Med = 10$; range = 5–24). The data collection took place between December 2019 and November 2020. In March 2020, the Covid-19 pandemic caused the administration of the Italian DABS to shift from an in-person administration to an interview administration by telephone or video calls (45 % of interviews) using WhatsApp, Skype, or Microsoft Teams. Eleven percent of the test-retest data collection was conducted with a hybrid approach method with the Time 1 interviews being in-person while the Time 2 interviews being done via video interviews.

For the inter-respondent reliability study, the Italian DABS was administered by the same interviewer independently to two different respondents, who were the mother and the father of the assessed individual, generally on the same day ($M = 0.60$; $SD = 1.20$; $Med = 0$; range = 0–5 days) (counterbalanced order of administration, first mother, then father: 46 %). The interviewers ($n = 5$, 100 % females) were undergraduate (20 %) or graduate (80 %) students in psychology (number of Italian DABS administrations for each interviewer: $M = 11.20$; $SD = 7.79$; $Med = 10$; range = 2–21). The data collection took place between January and November 2020. Due to the Covid-19 pandemic, 71 % of interviews were done using WhatsApp, Skype, or Microsoft Teams.

2.1.4. Data analysis

The test-retest reliability and the inter-respondent reliability of the Italian DABS domain scores and Total Score were estimated using the Pearson r correlation coefficient and the intraclass correlation coefficient (ICC). Following Cicchetti et al. (2011), the magnitude of the Pearson r correlation coefficients was interpreted as small (.10–.29), medium (.30–.49), large (.50–.69), or very large ($\geq .70$). Similarly, the magnitude of the ICC was explained as poor (<.40), fair (.40–.59), good (.60–.74), or excellent ($\geq .75$) (Cicchetti et al., 2011).

Additionally, the distributions of the normative scores of the Italian DABS domain scores and Total Score obtained during the two administrations (i.e., test-retest and inter-respondent) were compared using Student's t -tests for dependent groups. 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 ($\geq .80$).

2.2. Results and discussion

As seen in Table 1, the Italian DABS test-retest correlation coefficients were “very large” for the Pearson r_s (range: .91–.95) and “excellent” for the ICCs (range: .95–.98). These results are similar to the test-retest reliability coefficients reported for the original version of the DABS, which had “very large” Pearson r correlation coefficients and “excellent” ICCs as well (Tassé, Schalock, Balboni et al., 2016). Additionally, the test-retest correlation coefficients found are similar to the “excellent” ICCs reported for the Vineland-II (e.g., ICC = .82–.94; Sparrow et al., 2005), and the “very large” Pearson r_p correlation coefficients detected for both the ABAS-II (e.g., $r_p = .70$ –.90; Harrison & Oakland, 2003) and Vineland-3 (e.g., $r_p = .80$ –.92; Sparrow et al., 2016).

Scores obtained at Time 2 were statistically significantly higher than those obtained at Time 1, for all the DABS domains and Total Score. This finding can most likely be attributed to the individual's learning and development. However, the effect sizes were “small”

Table 1

Test-Retest Reliability: Normative Scores Obtained at Time 1 and Time 2 Administrations Across Italian DABS Domains and Total Score, Pearson r Correlation Coefficients (r_p) and Intraclass Correlation Coefficients (ICC), t -tests for Dependent Groups, and Cohen's d Effect Sizes ($n = 71$).

DABS	T1 <i>M (SD)</i>	T2 <i>M (SD)</i>	r_p	ICC	Student's t -test	Cohen's d
Conceptual Skills	94.82 (18.92)	99.36 (19.90)	.91**	.95***	-4.70***	.24
Social Skills	98.02 (17.10)	99.79 (17.98)	.94**	.97***	-2.52*	.10
Practical Skills	96.55 (16.84)	99.59 (17.70)	.93**	.96***	-4.00***	.18
Total Score	95.65 (19.30)	99.28 (20.42)	.95**	.98***	-4.91***	.18

*** $p < .001$.

** $p < .01$.

* $p < .05$.

for the conceptual skills and “negligible” in all other cases. Therefore, it can be said that the stability of the Italian DABS scores was found to be excellent.

As seen in Table 2, the Italian DABS inter-respondent correlation coefficients ranged from “large” to “very large” for the Pearson r s (range: .64–.89) and were “excellent” for the ICCs (range: .78–.94). The highest inter-respondent concordance was observed for the Italian DABS practical skills, whereas the lowest for the Italian DABS conceptual skills. Interestingly, this trend in the inter-respondent concordance was slightly different than the one reported for the original version of the DABS, for which the highest level of concordance was found for the conceptual skills (ICC = .87) and the lowest for the social skills (ICC = .61) (Tassé, Schalock, Balboni et al., 2016). Overall, our results compared well with the coefficients of inter-respondent agreement found for other standardized adaptive behavior scales, such as the Vineland-II (e.g., ICC = .61–.83; Sparrow et al., 2005), Vineland-3 (e.g., r_p = .70–.81; Sparrow et al., 2016) and ABAS-II (e.g., r_p = .74–.95; Harrison & Oakland, 2003), which reported “good” to “excellent” ICCs, and “very large” Pearson r correlation coefficients. Additionally, there were no statistically significant differences in normative scores across the two administrations.

The inter-respondent reliability coefficients of the Italian DABS were not perfect but are more than acceptable. When measuring adaptive behavior, due to context-specific measurement of the construct, obtaining perfect agreement between respondents is quite improbable given that they may observe the individual in different contexts, have different expectations, the individual may behave differently in different contexts, and/or the respondent may interpret their adaptive behavior differently. For these reasons, it is recommended to interview multiple respondents, when possible, to obtain the most comprehensive assessment of a person’s adaptive behavior (Tassé et al., 2012, 2017).

The participants were mostly individuals with typical development and just a few with ID. However, there is no evidence to suggest that the stability of scores or concordance between respondents would be affected by the participant’s ability level. Indeed, similar results were found with the original version of the DABS where a majority of participants had ID or a closely related developmental disability (Tassé, Schalock, Balboni et al., 2016).

Tables S2 and S3 of the Supplementary Material present the results of the test-retest and inter-respondent reliability analyses for each of the three DABS age forms. In most cases, these results are similar to those presented in Tables 1 and 2 obtained for the whole group of participants 4–21 years old. However, the test-retest Pearson r correlation coefficient of the practical skills of the DABS 4–8 years old, and inter-respondent Pearson r and ICC coefficients of the DABS versions 4–8 and 9–15 years old were slightly lower than those found for the whole group of participants 4–21 years old. Moreover, the inter-respondent correlation coefficient for the conceptual skills of the DABS 9–15 years old form was the only one that was not statistically significant. These few differences might be attributable to the lower number of participants for each DABS age forms.

3. Study 2: diagnostic accuracy of the Italian DABS compared with the Italian Vineland-II

This study aimed to investigate the diagnostic accuracy of the Italian DABS in detecting individuals with ID and with typical development. Additionally, the diagnostic accuracy of the Italian DABS was compared to that of the Italian Vineland-II for the same participants (i.e., both the Italian DABS and the Italian Vineland-II were administered to the same participants).

3.1. Materials and methods

3.1.1. Participants

The participants were 378 individuals (51 % male, $n = 193$) who had typical development (50 %, $n = 189$) or ID (50 %, $n = 189$; 31 % mild and 19 % moderate ID). The participants’ chronological age ranged between 4 and 21 years (individuals with typical development: M [SD] = 13.06 years [5.18]; individuals with ID: M [SD] = 12.09 years [4.55]), and in both groups 51 % were male. Table S1 shows the number of participants recruited for each chronological age. There were no statistically significant differences in the frequency of males and females between the participants with typical development and ID ($\chi^2_{(1)} = .01$, $p = .918$). The individuals with typical development were slightly older than those with ID ($t_{(376)} = 1.94$, $p = .053$; Cohen’s $d = .20$).

Participants lived in 15 of Italy’s 20 regions and therefore represented a broad geographic area of Italy. Specifically, participants came from the northern (26 %), central (46 %), and southern parts of Italy (28 %). Participants with typical development were recruited using the snowball sampling technique. Those with ID were recruited in several Research Hospitals specialized in child

Table 2

Inter-Respondent Reliability: Normative Scores Obtained for the Two Administrations with Respondents 1 and 2 Across Italian DABS Domains and Total Score, Pearson r Correlation Coefficients (r_p) and Intraclass Correlation Coefficients (ICC), and t -tests for Dependent Groups ($n = 56$).

DABS	Respondent 1 M (SD)	Respondent 2 M (SD)	r_p	ICC	Student’s t -test
Conceptual Skills	102.10 (16.64)	100.74 (15.89)	.64**	.78***	.74
Social Skills	100.90 (14.91)	101.95 (19.91)	.87**	.91***	-.78
Practical Skills	100.41 (17.43)	99.23 (17.29)	.89**	.94***	1.09
Total Score	101.26 (16.99)	100.71 (19.04)	.87**	.93***	.44

*** $p < .001$.

** $p < .01$.

neurology and psychiatry, national/regional mental health services for children and adults, or associations for individuals with disabilities.

3.1.2. Instruments

The Italian adaptation of the DABS (Balboni et al., in preparation; see Paragraph 2.1.2.) and the Italian version of the Vineland-II (Sparrow et al., 2005; It. ad. Balboni et al., 2016) were used. The Vineland-II allows the assessment of the three domains communication, socialization, and daily living skills for individuals 0–90 years of age. These three domains are similar to the conceptual, social, and practical skills domains of the DABS. Additionally, a motor skills domain is also available on the Vineland-II for individuals zero to six years of age. The Vineland-II Adaptive Behavior Composite provides an assessment of the overall adaptive behavior level. Normative scores ($M = 100$; $SD = 15$) are computed for three adaptive behavior domains and the Adaptive Behavior Composite. The Italian adaptation of the Vineland-II (Balboni et al., 2016) was approved by the Pearson Editor and showed excellent psychometric properties and robust normative scores (see Balboni et al. (2017); Balboni et al., 2021; Balboni, Rebecchini et al. (2020); and Tassé et al. (2019) for recent use of the Italian adaptation of the Vineland-II).

3.1.3. Procedure

Data collection was conducted between March 2018 and December 2019. Both the Italian DABS and Vineland-II were administered to the same respondent by the same interviewer in a counterbalanced order (DABS - Vineland-II: 48 %), with an interval that ranged from 0 to 55 days ($M [SD] = 1.27 [6.87]$; $Med = 0$ days) between the two administrations.

The interviewers ($n = 26$; 92 % female, $n = 24$) were psychologists (46 %), other health care professionals (35 %, e.g., pediatric neurologists, speech and language pathologists), or graduate students in psychology (19 %) (number of Italian DABS/Vineland-II administrations for each interviewer: $M = 14.54$; $SD = 22.13$; $Med = 5$; range = 1–93). The respondents were the mother (85 %, $n = 322$), father (11 %, $n = 43$), or another individual (4 %, $n = 13$) who knew well the person being assessed.

3.1.4. Data analysis

The diagnostic accuracy of the Italian DABS and Vineland-II was defined as correctly identifying individuals with a prior diagnosis of ID and correctly excluding individuals with typical development. Diagnostic accuracy was estimated using three statistical methods.

First, normative scores obtained by individuals with typical development and with ID were compared at the domain level and at the overall adaptive behavior score, independently for the Italian DABS and Vineland-II, using Student's *t*-tests for independent groups. For any statistically significant differences, 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 ($\geq .80$).

Second, sensitivity, specificity, overall correct classification index (OCC; i.e., overall proportion of individuals with ID and typical development correctly identified by the scale), rate of false positives and false negatives, positive likelihood ratio (LR+; i.e., odd that an individual with ID is correctly detected by the scale) and negative likelihood ratio (LR-; i.e., odd that an individual with typical development is erroneously classified with ID by the scale) were computed independently for both instruments. In agreement with the second criteria for the diagnosis of ID (APA, 2013; Schalock et al., 2021; WHO, 2021), a normative score of 70 was used as the cutoff score (i.e., two standard deviations below the mean) for the presence of significant limitations in adaptive behavior. Therefore, for both Italian DABS and Vineland-II, a normative score ≤ 70 on at least one of the three adaptive behavior domains or the overall adaptive behavior score was used as the criterion for detecting ID.

Sensitivity, specificity, and OCC are statistically significant if their confidence intervals (CI) do not include the value of .50 (i.e., the classification made by the test scores is more accurate than a random classification). Values $\geq .70$ for sensitivity and $\geq .80$ for specificity represent the minimum standard for accuracy of diagnostic tests (Matthey & Petrovski, 2002), and higher values are preferred. Conversely, lower rates of false positives and false negatives are preferred. LR + e LR- are statistically significant if their CIs do not include the value of 1. Furukawa et al. (2002) recommended the following parameters for their clinical significance: LR+ > 1 or LR- < 1 = fail; LR+ > 2 or LR- $< .5$ = poor; LR+ > 5 or LR- $< .2$ = good; and LR+ > 10 or LR- $< .1$ = excellent.

In agreement with the current conceptualization of the model of adaptive behavior (Schalock et al., 2021; Tassé et al., 2012), motor skills are not a component of the adaptive behavior construct. Motor skills are not measured by the DABS, but they are assessed on the Vineland-II and contribute to producing the Vineland-II Adaptive Behavior Composite score for individuals 4–6 years old. Consequently, to improve the comparability of the two instruments, sensitivity, specificity, false positive and false negative rates, LR + and LR- of the Italian Vineland-II Adaptive Behavior Scores were also calculated after removing the contribution of the motor skills scale for individuals aged 4–6 years.

Third, receiver operating characteristic (ROC) curves analysis was run independently for the Italian DABS and Vineland-II. The ROC represents another way to measure sensitivity and specificity in identifying individuals with ID and with typical development based on their test scores. The ROC plots sensitivity versus 1 - specificity (i.e., false positive rate) for different possible test cutoff scores (Altman & Bland, 1994; McFall & Treat, 1999; Streiner, 2003). The ROC yields an area under the curve (AUC) representing the probability of the test correctly classifying individuals with ID and with typical development. Specifically, the AUC reflects the probability that a randomly selected person from the group of individuals with ID will have a normative score below the normative score of a randomly selected person from the group of individuals with typical development. AUC values range from 0 to 1 and they are statistically significant if their CIs do not include the value of .5. Swets (1998) suggested the following parameters to evaluate the AUC values: .5–.6 = fail; .6–.7 = poor; .7–.8 = fair; .8–.9 = good; and .9–1.0 = excellent. Moreover, independently for the Italian DABS and Vineland-II domains and the overall adaptive behavior score, the presence of an overlap between AUC CIs of the test domains was ascertained, reflecting the absence of statistically significant differences among the test scales in correctly classifying individuals with

typical development and ID.

Lastly, the procedure suggested by Hanley and McNeil (1983) was applied to compare the AUC of the pairs of Italian DABS and Vineland-II scales measuring a similar adaptive behavior domain (i.e., DABS conceptual skills and Vineland-II communication, DABS social skills and Vineland-II socialization, DABS practical skills and Vineland-II daily living skills) or the overall adaptive behavior (i.e., DABS Total Score and Vineland-II Adaptive Behavior Composite).

3.2. Results and discussion

As shown in Table 3, participants with ID had average domain and total scores significantly lower ($p < .001$) than the average domain and total scores of participants with typical development across all Italian DABS or Vineland-II adaptive behavior domains and the overall adaptive behavior score. The effect size was “large”, approximately 2.00 or greater, reflecting differences about equal to two standard deviations. Similar results were found for each of the three DABS age forms (see Table S4 in the Supplementary material).

Fig. 1 shows the normative scores on the Italian DABS and Vineland-II domains obtained by individuals with typical development and ID for each year of age. Individuals with ID always had lower scores than individuals with typical development. Moreover, the average domain scores of individuals with ID were below the cutoff score of 70 on the overall adaptive behavior score and conceptual skills domain across all chronological ages, and for the majority of ages on the social skills domain (DABS: 83 %; Vineland-II: 72 %) and practical skills domain (DABS: 50 %; Vineland-II: 83 %).

Table 4 reports sensitivity, specificity, OCC, false positive and false negative rates, LR + and LR- for the Italian DABS and Vineland-II. The excellent discriminant validity of the Italian DABS to correctly classify individuals with ID and with typical development is comparable to that of the Vineland-II. Indeed, for both scales, sensitivity, specificity, and OCC values were close to 1 and statistically significant, given that their CIs were higher than .50. Furthermore, LR + and LR- values were statistically significant, because they did not include the value of 1, and they represented an “excellent” clinical significance considering LR + and a “good” clinical significance considering LR- for both the Italian DABS and Vineland-II. These analyses across the three separate DABS age forms showed similar results (see Table S5 in the Supplementary Material).

The estimation of sensitivity, specificity, OCC, false positive and false negative rates, LR + and LR- of the Italian Vineland-II were recalculated after removing the motor skills domain from the computation of the Adaptive Behavior Composite score for individuals who were 4–6 years of age, and results were comparable to those reported in Table 4.

Table 5 shows the AUC of the Italian DABS and Vineland-II domains and the overall adaptive behavior score. For both instruments, none of the AUC CIs included the value of .5, indicating that an individual with ID has more than a 90 % of probability to obtain normative scores lower than those obtained by an individual with typical development. Based on Swets' (1998) criteria of clinical significance of the AUC, all Italian DABS and Vineland-II domains and overall adaptive behavior scores yielded “excellent” ability levels in estimating the classification of individuals with ID and with typical development, except the Italian DABS practical skills and Vineland-II daily living skills which yielded a “good” ability level. Furthermore, for both instruments, the overlap between the AUC CIs expressed the absence of statistically significant differences among the domains to correctly classify the individuals.

Table 5 also shows the comparisons between the AUC of the Italian DABS and Vineland-II domains measuring similar domains and the overall adaptive behavior score. The Italian DABS and Vineland-II showed comparable diagnostic accuracy, with the only exception being for the social skills domain, given that the Italian Vineland-II socialization had a statistically significant greater AUC than that of the Italian DABS social skills domain. However, when this comparison was replicated separately considering each Italian DABS version age group (4–8 years old, $n = 104$; 9–15 years old, $n = 169$; 16–21 years old, $n = 105$), this difference was no longer detectable in any of the three age groups.

Finally, Fig. 2 shows the ROC curves of the Italian DABS conceptual skills, social skills, practical skills, and DABS Total Score, and of the Italian Vineland-II communication, socialization, daily living skills, and Adaptive Behavior Composite.

The Italian DABS showed a diagnostic accuracy in detecting individuals with ID comparable to that of the original version of the DABS (Balboni et al., 2014). Furthermore, the comparison of the diagnostic accuracy of the Italian DABS with that of the Italian Vineland-II showed that these two instruments performed equally well in correctly classifying individuals with ID or with typical development. However, the Italian DABS has the added advantage of having fewer items (i.e., number of items = 75) and thus requiring less time to administer than the Italian Vineland-II (i.e., number of items = 383).

Table 3

Diagnostic Accuracy: Normative Scores of Individuals with Typical Development ($n = 189$) and Intellectual Disability ($n = 189$) on the Italian DABS and Vineland-II, Student's t -tests for Independent Groups and Cohen's d Effect Sizes.

	Typical Development	Intellectual Disability	Student's t -test	Cohen's d
DABS Conceptual Skills	102.36 (12.42)	56.77 (18.59)	28.03	2.89
Vineland-II Communication	97.72 (9.80)	54.52 (19.94)	26.74	2.76
DABS Social Skills	101.00 (12.58)	66.63 (19.29)	20.52	2.12
Vineland-II Socialization	103.43 (9.51)	63.26 (21.49)	23.50	2.42
DABS Practical Skills	99.53 (13.23)	70.06 (19.15)	17.41	1.80
Vineland-II Daily Living Skills	94.94 (12.84)	64.38 (21.28)	16.91	1.74
DABS Total Score	101.04 (11.86)	57.58 (19.65)	26.04	2.68
Vineland-II Adaptive Behavior Composite	98.67 (8.99)	53.89 (20.99)	26.96	2.78

Note. All Student's t -tests were statistically significant ($p < .001$).

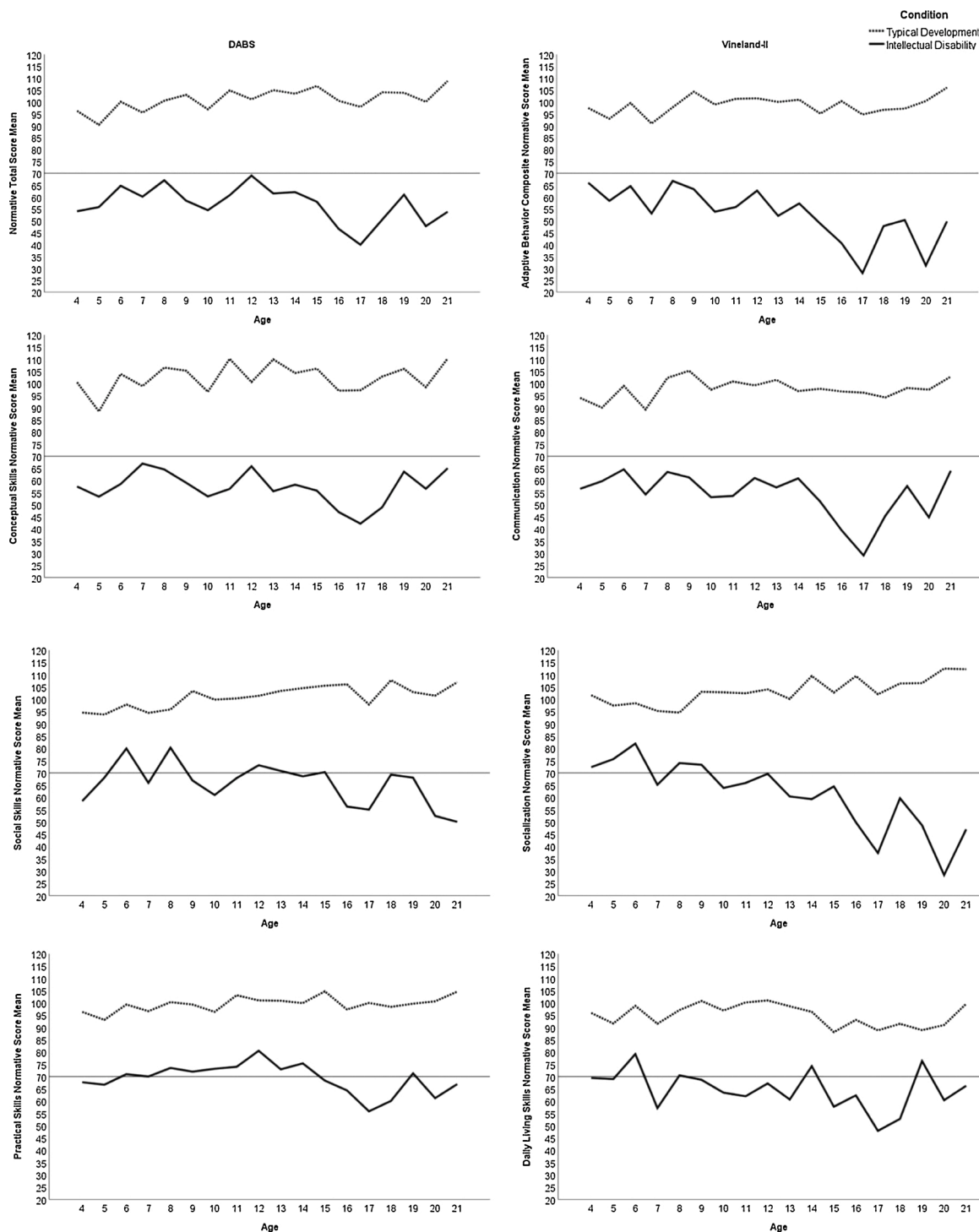


Fig. 1. Diagnostic Accuracy: Comparing Normative Scores of Individuals with Typical Development ($n = 189$) and with Intellectual Disability ($n = 189$) on the Italian DABS and Vineland-II Across Chronological Age (Years).

Therefore, the Italian DABS represents an excellent assessment tool in determining whether an individual 4–21 years old has significant limitations in the adaptive behavior for the purpose of making a diagnosis of ID. This observation is perhaps even more important considering that the Vineland-3 (Sparrow et al., 2016) has 34 % more items than the Vineland-II (i.e., 132 items more; Farmer, Adedipe et al., 2020), and the Vineland-3 factorial structure has not yet been empirically confirmed (Farmer, Floyd et al., 2020; Pandolfi and Magyar, 2021).

Table 4

Diagnostic Accuracy: Sensitivity, Specificity, Overall Correct Classification (OCC), False Negatives, False Positives, Positive Likelihood Ratio (LR+), and Negative Likelihood Ratio (LR-) (95 % Confidence Interval) of the Italian DABS and Vineland-II ($n = 378$).

	DABS Value (IC)	Vineland-II Value (IC)
Sensitivity	.86 (.81–.91)	.86 (.81–.91)
Specificity	.99 (.98–1)	.95 (.92–.98)
OCC	.93 (.90–.96)	.90 (.87–.93)
False negatives	.14 (.09–.19)	.14 (.09–.19)
False positives	.01 (.00–.02)	.05 (.02–.08)
LR+	81.50 (20.83–357.01)	18.00 (9.23–32.18)
LR-	0.14 (0.10–0.20)	0.15 (0.10–0.21)

Table 5

Diagnostic Accuracy: Area Under the Curve (AUC) (95 % Confidence Interval) of the Italian DABS and Vineland-II, and z-test for Comparisons ($n = 378$) Between the AUCs of the Corresponding Domains and Overall Scores.

	AUC (CI)	z
DABS Conceptual Skills	.968 (.949–.987)	
Vineland-II Communication	.974 (.960–.989)	-.62
DABS Social Skills	.932 (.905–.958)	
Vineland-II Socialization	.970 (.955–.984)	-3.30*
DABS Practical Skills	.895 (.862–.929)	
Vineland-II Daily Living Skills	.882 (.849–.915)	.74
DABS Total Score	.960 (.939–.981)	
Vineland-II Adaptive Behavior Composite	.976 (.962–.990)	-1.62

Note. Critical z score was equal to $|2.39|$ after the Bonferroni correction for multiple comparisons ($p < .05$).

* $p < .05$.

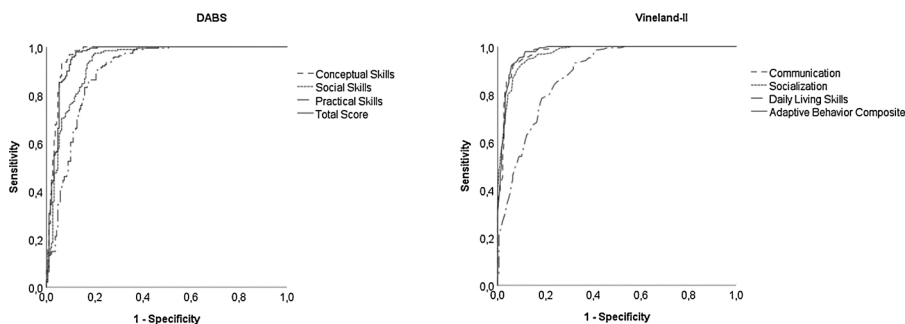


Fig. 2. Diagnostic Accuracy: ROC Curve of the Italian DABS and Vineland-II Domains ($n= 378$).

4. Conclusions

It is critical to have high quality standardized measures available across languages and cultures to provide reliable and valid tools to assist clinicians in making the determination of ID around the world. This article reports that the Italian DABS has excellent test-retest and inter-respondent reliability. Equally important, the Italian DABS is an excellent assisting tool for diagnosing ID. Indeed, the Italian DABS showed diagnostic sensitivity and specificity that are as strong as those of the Italian Vineland-II, but has few items and a less time-consuming administration.

The original version of the DABS has unique features and excellent psychometric properties. The present investigation contributes to show that the Italian DABS has comparably strong psychometric properties and can be useful for the diagnosis of ID. A Spanish version of the DABS is already available that has also reported excellent psychometric properties (Arias et al., 2013; Navas et al., 2012; Verdugo et al., 2021). A French version of the DABS is under development (von Rotz et al., 2021) as well as a Portuguese (Brazil) one (Schutz et al., 2021). Future studies will be necessary to investigate the invariance of the DABS across countries. Further investigations are needed to examine the factorial structure of the Italian DABS and its invariance across individuals with typical development and those with ID. The measurement invariance represents the gold standard approach to ensuring that the meaning of the construct being assessed by an instrument is consistent across groups and time (Little, 2013; Millsap, 2012; Newsom, 2015; van de Schoot et al., 2012, 2015).

CRedit authorship contribution statement

GB contributed to the conceptualization of the study, methodology, formal analysis, resources, data curation, writing original draft preparation and review and editing, supervision, project administration. AB contributed to the methodology of the study, formal analysis, investigation, data curation, writing original draft preparation and review and editing. PA contributed to the methodology of the study, software, formal analysis, data curation, writing review and editing. PB, SB, SM contributed to investigation, resources, data curation. ER contributed to conceptualization, methodology, writing review and editing of the study. MJT contributed to the conceptualization of the study, methodology, resources, writing review and editing. All authors read and approved the final manuscript.

Ethical standards

This study was performed in accordance with the ethical standards in the 2013 Fortaleza version of the Declaration of Helsinki. All parents and individuals over 18 years old provided written informed consent and their anonymity was guaranteed. Participants did not receive any form of incentive to participate in this study. The University Bioethics Committee of the University of Perugia approved the study procedure and all study documents (#12/2018, #36356/2020).

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work has been partially supported by grant-RC 1.22 "Studi pilota con approcci innovativi sulle premesse neuro-fisiopatologiche, la diagnosi e la riabilitazione dei disturbi neurologici del bambino" from the IRCCS Stella Maris Foundation (and the 5x 1000 voluntary contributions, Italian Ministry of Health).

The authors wish to thank the following hospitals, mental health services, and associations that helped with the recruitment of the participants with ID of this study: AEMOCON Emozioni di Conoscere Association, Bologna; Ambulatorio PRI-A, Casa del Giardiniere, Local Health Care Unit of Bologna, Bologna; Boggiano Pico Center, Opera Don Orione, Genova; Campi d'Arte Social Cooperative, Bologna; Don Gnocchi Foundation, Centro Santa Maria Al Castello, Pessano con Bornago (Milan); IRCCS Mondino Foundation, Pavia; IRCCS Stella Maris Foundation, Pisa; Italian Association of Individuals with Down Syndrome, Sections of Livorno and Pisa; Oasi Research Institute-IRCCS, Troina; Opera Don Guanella Rehabilitation Center, Casa San Giuseppe, Rome; Private Studio of Licensed Psychologists Bastianello, Vittorio Veneto (Treviso); Private Studio of Licensed Psychologists Centro Clinico Psiche, Genova; Regional Center for Language and Cognitive Disabilities, Local Health Care Unit of Bologna, Bologna; Serafico Institute of Assisi, Perugia; Unit of Adult Mental Health, Local Health Care Unit of North-West Tuscany, Pisa; Unit of Child Neurology and Psychiatry, ASST Spedali Civili di Brescia, Brescia; Unit of Neurodevelopmental Pediatrics, Local Health Care Unit of CN1, Cuneo and Mondovì; Unit of Neurodevelopmental Pediatrics, Local Health Care Unit of Umbria 1, Perugia.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ridd.2022.104185>.

References

- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *BMJ*, *309*(6947), Article e102. <https://doi.org/10.1136/bmj.309.6947.102>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author. <https://doi.org/10.1176/appi.books.9780890425596>
- Arias, B., Verdugo, M. A., Navas, P., & Gómez, L. E. (2013). Factor structure of the construct of adaptive behavior in children with and without intellectual disability. *International Journal of Clinical and Health Psychology*, *13*(2), 155–166. [https://doi.org/10.1016/S1697-2600\(13\)70019-X](https://doi.org/10.1016/S1697-2600(13)70019-X)
- Balboni, G., Bacherini, A., Anselmi, P., & Robusto, E. (in preparation). Diagnostic Adaptive Behavior Scale. Adattamento italiano. [Diagnostic Adaptive Behavior Scale. Italian Adaptation]. Hogrefe.
- Balboni, G., Bacherini, A., Rebecchini, G., Cagiano, R., Mancini, A., Tancredi, R., & Muratori, F. (2021). Individual and environmental factors affecting adaptive behavior of toddlers with autism spectrum disorder: Role of parents' socio-cultural level. *Journal of Autism and Developmental Disorders*, *51*(10), 3469–3482. <https://doi.org/10.1007/s10803-020-04803-x>
- Balboni, G., Belacchi, C., Bonichini, S., & Coscarelli, A. (2016). *Vineland Adaptive Behavior Scales. Survey interview form. Standardizzazione italiana (2nd ed.)* [Vineland Adaptive Behavior Scales. Survey interview form. Italian Standardization (2nd ed.)]. Giunti OS Organizzazioni Speciali.
- Balboni, G., Incognito, O., Belacchi, C., Bonichini, S., & Cubelli, R. (2017). Vineland-II adaptive behavior profile of children with attention-deficit/hyperactivity disorder or with specific learning disorders. *Research in Developmental Disabilities*, *61*, 55–65. <https://doi.org/10.1016/j.ridd.2016.12.003>

- 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(11), 2884–2893. <https://doi.org/10.1016/j.ridd.2014.07.032>
- Balboni, G., Rebecchini, G., Elisei, S., & Tassé, M. J. (2020). Factors affecting the relationship between adaptive behavior and challenging behaviors in individuals with intellectual disabilities and co-occurring disorders. *Research in Developmental Disabilities*, 104, Article 103718. <https://doi.org/10.1016/j.ridd.2020.103718>
- Bruininks, R. H., Woodcock, R., Weatherman, R., & Hill, B. (1996). *Scales of Independent Behavior—Revised*. Chicago, IL: Riverside.
- Cicchetti, D. V., Koenig, K., Klin, A., Volkmar, F. R., Paul, R., & Sparrow, S. (2011). From Bayes through marginal utility to effect sizes: A guide to understanding the clinical and statistical significance of the results of autism research findings. *Journal of Autism and Developmental Disorders*, 41, 168–174. <https://doi.org/10.1007/s10803-010-1035-1036>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Doll, E. A. (1936). Preliminary standardization of the Vineland Social Maturity Scale. *American Journal of Orthopsychiatry*, 6(2), 283–293. <https://doi.org/10.1111/j.1939-0025.1936.tb05235.x>
- Doll, E. A. (1953). *Measurement of social competence: A manual for the Vineland Social Maturity Scale*. Vineland Training School. <https://doi.org/10.1037/11349-000>
- Farmer, C., Adedipe, D., Bal, V. H., Chlebowski, C., & Thurm, A. (2020). Concordance of the Vineland Adaptive Behavior Scales, second and third editions. *Journal of Intellectual Disability Research*, 64(1), 18–26. <https://doi.org/10.1111/jir.12691>
- Farmer, R. L., Floyd, R. G., & McNicholas, P. J. (2020). Is the Vineland-3 Comprehensive Interview Form a multidimensional or unidimensional scale?: Structural analysis of subdomain scores across early childhood to adulthood. *Assessment*, 28(7), 1848–1864. <https://doi.org/10.1177/1073191120947804>
- Furukawa, T. A., Strauss, S., Bucher, H. C., & Guyatt, G. (2002). Diagnostic tests. In G. Guyatt, & D. Rennie (Eds.), *Users' guides to the medical literature* (pp. 121–140). AMA Press.
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839–843. <https://doi.org/10.1148/radiology.148.3.6878708>
- Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System, second edition: Manual*. Harcourt Assessment.
- Harrison, P. L., & Oakland, T. (2015). *Adaptive Behavior Assessment System, third edition (ABAS-3): Manual*. Harcourt Assessment.
- Heber, R. (1961). *A manual on terminology and classification in mental retardation* (rev. ed.). American Association on Mental Deficiency.
- International Test Commission. (2017). *The ITC guidelines on translation and adapting tests* (2nd ed.) <https://www.intestcom.org/>.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Matthey, S., & Petrovski, P. (2002). The Children's Depression Inventory: Error in cutoff scores for screening purposes. *Psychological Assessment*, 14(2), 146–149. <https://doi.org/10.1037//1040-3590.14.2.146>
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*, 50, 215–241. <https://doi.org/10.1146/annurev.psych.50.1.215>
- Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.
- Navas, P., Verdugo, M. A., Arias, B., & Gómez, L. E. (2012). Development of an instrument for diagnosing significant limitations in adaptive behavior in early childhood. *Research in Developmental Disabilities*, 33, 1551–1559. <https://doi.org/10.1016/j.ridd.2012.03.006>
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.
- Pandolfi, V., & Magyar, C. I. (2021). Vineland-3 structural validity and interpretability of domain scores: Implications for practitioners assessing adolescents with developmental conditions. *American Journal on Intellectual and Developmental Disabilities*, 126(3), 216–229. <https://doi.org/10.1352/1944-7558-126.3.216>. PMID: 33910242.
- Pearson, N. A., Patton, J. R., & Mruzek, D. W. (2016). *Adaptive Behavior Diagnostic Scale: Examiner's manual*. PRO-ED.
- Schalock, R. L. (1999). The measurement of adaptive behavior. In R. L. Schalock (Ed.), *Adaptive behavior and its measurement: Implications for the field of mental retardation* (pp. 99–101). American Association on Mental Retardation.
- Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and system of supports* (12th ed.). American Association on Intellectual and Developmental Disabilities.
- Schutz, A., Xavier, C. E., Marques, C., Pimentel, M., Hallberg, S. C. M., & Bandeira, D. R. (2021, June 30 - July 3). *Versão Brasileira da Diagnostic Adaptive Behavior Scale (DABS): evidências preliminares de validade [Brazilian Version of the Diagnostic Adaptive Behavior Scale (DABS): Preliminary Evidence of Validity]* [Poster presentation]. 10th Congresso Brasileiro de Avaliação Psicológica, Brazil.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Pearson Assessment.
- Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales* (3rd ed.). Pearson Assessment.
- Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, 81(3), 209–219. https://doi.org/10.1207/S15327752JPA8103_03
- Swets, J. A. (1998). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293. <https://doi.org/10.1126/science.3287615>
- Tassé, M. J., & Balboni, G. (2021). Theories and measurement of adaptive behavior. In L. M. Glidden (Ed.), *APA handbook of intellectual and developmental disabilities: Vol. 1. Foundations*. American Psychological Association. <https://doi.org/10.1037/0000194-016>.
- Tassé, M. J., & Craig, E. M. (1999). Critical issues in the cross-cultural assessment of adaptive behavior. In R. L. Schalock (Ed.), *Adaptive behavior and its measurement: Implications for the field of mental retardation* (pp. 161–184). American Association on Mental Retardation.
- Tassé, M. J., Balboni, G., Navas, P., Luckasson, R. A., Nygren, M. A., Belacchi, C., Bonichini, S., Reed, G., & Kogan, C. (2019). Developing behavioural indicators for intellectual functioning and adaptive behaviour for ICD-11 disorders of intellectual development. *Journal of Intellectual Disability Research*, 63(5), 386–407. <https://doi.org/10.1111/jir.12582>
- Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Jr., Borthwick-Duffy, S. A., Spreat, S., Thissen, D., 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. <https://doi.org/10.1352/1944-7558-117.4.291>
- Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Spreat, S., ... Zhang, D. (2017). *Diagnostic Adaptive Behavior Scale user's manual*. American Association on Intellectual and Developmental Disabilities.
- Tassé, M. J., Schalock, R. L., Balboni, G., Spreat, S., & Navas, P. (2016). Validity and reliability of the Diagnostic Adaptive Behaviour Scale. *Journal of Intellectual Disability Research*, 60, 80–88. <https://doi.org/10.1111/jir.12239>
- Tassé, M. J., Schalock, R. L., Thissen, D. M., Balboni, G., Bersani, H., Jr., 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. <https://doi.org/10.1352/1944-7558-121.2.79>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6, Article 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Verdugo, M. A., Arias, B., & Navas, P. (2021). Escala de Diagnóstico de Conducta Adaptativa (DABS). *Manual del usuario. Adaptación y validación del instrumento DABS: Diagnostic Adaptive Behavior Scale-Standardization Version (Tassé et al., 2017)*.
- von Rotz, L., Courbois, Y., Desportes, V., Lacroix, A., Reymond, M.-P., Tassé, M. J., Touil, N., Tsao, R., & Straccia, C. (2021, July 6–8). *The French version of the DABS: Translation process and preliminary results on content validity* [Poster presentation]. 6th Europe Congress of the International Association for the Scientific Study of Intellectual and Developmental Disabilities (IASSIDD), Amsterdam, The Netherlands.
- World Health Organization. (2021). *International statistical classification of diseases and related health problems (11th Revision)*. ICD-11 (who.int).
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29. <https://doi.org/10.22237/jmasm/1177992180>