

French Adaptation of the Diagnostic Adaptive Behavior Scale: An IRT-Based Approach

Claudio Straccia , Lorna von Rotz , Raphaele Tsao , Nathalie Touil , Agnès Lacroix , Vincent Desportes , Annick Comblain & Yannick Courbois

To cite this article: Claudio Straccia , Lorna von Rotz , Raphaele Tsao , Nathalie Touil , Agnès Lacroix , Vincent Desportes , Annick Comblain & Yannick Courbois (26 Aug 2026): French Adaptation of the Diagnostic Adaptive Behavior Scale: An IRT-Based Approach, American Journal on Intellectual and Developmental Disabilities, DOI: [10.1080/19447515.2026.2714920](https://doi.org/10.1080/19447515.2026.2714920)

To link to this article: <https://doi.org/10.1080/19447515.2026.2714920>



© 2026 The Author(s). Published with
license by Taylor & Francis Group, LLC.



Published online: 26 Aug 2026.



Submit your article to this journal [↗](#)



View related articles [↗](#)

French Adaptation of the Diagnostic Adaptive Behavior Scale: An IRT-Based Approach

Claudio Straccia^a, Lorna von Rotz^a, Raphaelae Tsao^b, Nathalie Touil^c, Agnès Lacroix^d, Vincent Desportes^c, Annick Comblain^e, and Yannick Courbois^f

^aUniversity of Fribourg, Fribourg, Switzerland; ^bUniversity of Aix-Marseille, Marseille, France; ^cUniversity Hospital Center of Lyon (CHU), Lyon, France; ^dUniversity of Rennes, Rennes, France; ^eUniversity of Liège, Liège, Belgium; ^fUniversity of Lille, Lille, France

ABSTRACT

The aim of this study was to evaluate the data collected using the French version of the Diagnostic Adaptive Behavior Scale (DABS-F) during the main field test of the adaptation project, from an approach based on the item response theory. Six hundred and eighty-six participants with typical development and 199 participants with intellectual disability were recruited across three European French-speaking countries (Belgium, France, and Switzerland). The data were analyzed replicating the analyses of the original developers of the DABS. Results showed that unidimensionality, link-items, discrimination, and difficulty parameters arising from the IRT analyses, as well as discriminant validity of the DABS-F, were all comparable with the results obtained using the original version of the instrument. These results were discussed in relation to the whole adaptation project of the DABS-F.

ARTICLE HISTORY

Received 15 April 2026
Accepted 19 July 2026

KEYWORDS

Transcultural adaptation;
DABS; adaptive behavior;
intellectual disability; IRT
analysis

Introduction

The construct of adaptive behavior (AB) refers to those skills that are learned and performed by individuals in order to function in everyday life and to meet society's expectations for people of the same age and culture (Luckasson et al., 2002; Schalock et al., 2010). It is made up of three domains. Conceptual skills (a) refer, for instance, to language, reading, writing, and memory, as well as to concepts of money and time. Social skills (b) include interpersonal skills, social responsibility, empathy, and the ability to follow rules and laws. Practical skills (c) involve, for example, activities of personal health and hygiene, safety, the ability to follow schedules, and money management.

The three diagnostic manuals of reference for the diagnosis of intellectual disability (ID; the American Association on Intellectual and Developmental Disabilities [AAIDD] manual [Schalock et al., 2021]; the *International Classification of Diseases*, 11th Revision [World Health Organization, 2019], and the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition [American Psychiatric Association, 2013]) all require the presence of significant limitations in (at least) one of the three domains of AB as a diagnostic criterion for this disorder. These limitations are defined as performances that are around two standard deviations, or more, below the population mean. Furthermore, severity levels of ID are now based on AB, rather than intellectual functioning, because the former is more directly related to the individual's support needs (American Psychiatric Association, 2013).

The importance of AB is not limited to ID. In fact, a number of other disorders are characterized by limitations in AB. Harrison and Oakland (2003), for instance, highlighted that deficits in AB are core features in disorders such as autism spectrum disorder, attention-deficit/hyperactive disorder, emotional and behavioral disorders, hearing and motor impairments, and communication disorders. It would appear that AB is also related to age-related decline, dementia, and Alzheimer's disease and that the assessment of AB can be very important for tracking disease progression or prognosis (Takata et al., 2013; Zigman et al., 2002).

CONTACT Claudio Straccia  Claudio.Straccia@unifr.ch  Department of Specialized Education, University of Fribourg, Rue Saint-Pierre Canisius 21, Fribourg 1700, Switzerland

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Due to the importance of AB, a large number of assessment tools have been developed over the last five decades. However, the vast majority of these instruments contain a number of disadvantages that complicate the use of these tools for the purpose of establishing a diagnosis of ID. Such disadvantages are evident in instruments that were developed to assess specific adaptive skills (e.g., toilet training) rather than the whole construct of AB, in the absence of general population norms, the inclusion of maladaptive behaviors, the focus on maximal rather than typical performance, and the inconsistency with the tripartite definition of AB (Tassé et al., 2012).

In light of this, the AAIDD (Balboni et al., 2014; Tassé et al., 2017; Tassé et al., 2016) developed the Diagnostic Adaptive Behavior Scale (DABS), the first psychometrically sound instrument specifically designed to optimize discrimination around the zone defining significant limitations in AB in children and young people (ranging from 4–21 years of age). The DABS has been tailored according to the current tripartite definition of AB, it is a short and accurate assessment instrument (with 75 items, 25 items per domain, it is noticeably shorter than comparable instruments), it does not include maladaptive behaviors to focus directly on AB, and it expands items related to higher order social adaptive skills. It is a proxy report administered via a face-to-face interview between a professional (i.e., interviewer) and a respondent (e.g., parent, caregiver, teacher) and each item is coded on a 4-point scale ranging from 0 = *no—rarely or never does it* to 3 = *Yes—does it always or almost always independently—never or rarely needs reminders or assistance*. The instrument is available in three age range versions: 4 to 8, 9 to 15, and 16 to 21 years of age. The estimated administration time for the DABS interview is approximately 30 minutes. The DABS scoring is completed by means of a computerized scoring system that uses item response theory (IRT) algorithms to analyze the response patterns and computes a score for each domain: conceptual, practical, and social skills, as well as a composite score. These scores are then standardized on a distribution with a mean of 100 and standard deviation of 15. The cutoff defining significant limitations in adaptive skills is settled at two standard deviations below the mean.

The DABS has been adapted in several other countries, including Spain (Verdugo Alonso et al., 2014), Italy (Balboni et al., 2022, 2023), and Brazil (Hallberg et al., 2025; Schütz et al., 2025). All of these adaptations show psychometric properties consistent with the original version. The DABS has been recently adapted in a European French-speaking context (Belgium, France, and Switzerland) following the recommendations for transcultural adaptation methodology in the field of intellectual and developmental disabilities (IDD; Tassé & Craig, 1999) and showing good levels of content validity (Von Rotz et al., 2024).

The next stage in the development of the European French adaptation of the DABS (DABS-F) will be to examine its psychometric properties from an approach based on the IRT. In fact, one of the distinguishing features of the original version of the instrument is the fact that its development based on IRT models enables the instrument items to be calibrated to be reliable and discriminant at the AB levels that are relevant for the ID diagnosis (i.e., around the cutoff point for determining two standard deviations below the mean). Therefore, the aim of this study is to replicate the IRT analysis performed in the original version of the DABS in the DABS-F. Accordingly, the results were mostly discussed in reference to the results found by the original developers.

Method

Participants and Procedure

The target sample of this study was structured and stratified around three variables: age, biological sex, and presence of ID diagnosis. For the purposes of the first variable, we aimed to recruit at least 30 participants with typical development (TD) per year of age. This was necessary in order to have enough participants in every year of age covered by the DABS-F (4–21) because the smoothed means and standard deviations used for standardizing the DABS-F scores would be computed in each of these age groups. For biological sex, we simply aimed to have a balanced sample in terms of males and females. When it came to the third variable, an overrepresentation of individuals presenting a diagnosis of ID was required to ensure we could obtain the necessary amount of information on the levels of AB at the point of two standard deviations below the mean of the general population. This meant that at least 20% of the sample would have to be represented by participants with ID, or more specifically, it would be necessary to recruit about eight participants with ID

Table 1. Demographic Characteristics of the Sample

Demographics	Sample (<i>n</i> = 885)	
	<i>M</i>	<i>SSD</i>
Age	12.44	5.21
	<i>n</i>	Percentage
Sex		
Female	430	48.59
Male	455	51.41
Country		
Belgium	103	11.64
France	474	53.56
Switzerland	308	34.80
ID-related diagnosis		
None (TD)	686	77.51
Mild ID	139	15.71
Moderate ID	60	6.78
Respondent		
Mother	586	66.21
Father	126	14.24
Other relative	66	7.46
Professional	86	9.72
Other	21	2.37

Note. ID = Intellectual Disability, TD = Typical Development, *M* = Mean, *SD* = Standard Deviation.

per year of age, with a 2:1 proportion between mild and moderate severity levels of ID. The first criterion of inclusion in the sample of participants with ID was to have a medical diagnosis of this disorder. Therefore, we had initially planned to retrieve the severity levels from diagnosis records. However, the medical diagnosis records almost never included information on severity levels. Thus, we have translated in French the observable descriptions of mild and moderate ID provided by Tassé et al. (2019), and we have used those descriptions to classify participants' severity levels of ID. To facilitate the recruitment of this large number of participants, six research sites were involved in the study—one each in Belgium and Switzerland and four research sites in France, located in the north, south, east, and west of the country.

Table 1 presents the main sociodemographic characteristics of the sample. The mean age of the sample was around 12 years old (*SD* = 5.21). More importantly, we successfully recruited at least 30 participants with TD per year of age, with the lowest frequency (*n* = 33) at 9 years, and the highest (*n* = 46) at 15 years. Participants with ID represented 22.49% (*n* = 199) of the whole sample (*n* = 885), with between eight and 16 participants with ID in each age group. As planned, we recruited more participants with a mild severity level of ID than a moderate severity level, in order to obtain information around the cutoff point defining significant limitations in AB. The ratio between the frequencies of the two severity levels was 2.3:1. As shown in Table 1, the vast majority of respondents were the mothers of the assessed individuals, followed by fathers in a proportion very close to 4:1. Other relatives included grandparents, (adult) siblings, and uncles. Professional respondents fell into three main categories: specialized teachers, teachers, and support staff. Finally, other respondents (e.g., mothers- or fathers-in-law), participated in the study in a small number of cases.

All respondents signed an informed consent form that explained the procedure and the goal of the project and stated that they were free to withdraw from the procedure at any time, without having to provide any explanation. Three types of informed consent were generated for the assessed individuals (for children, adolescents, and adults). Children were given information in a simplified document and were not asked to sign the consent. On the other hand, both adolescents and adults were asked to sign the consent, giving us their permission to interview the designated respondent. This procedure has been approved by the ethical committee authority (Cantonal Ethics Commission on Human Research, submission number 2020–00617).

Instrument

AB was assessed using the DABS-F. This instrument was developed through a rigorous translation and adaptation process using the committee translation method and then subjected to an initial content validity

assessment (von Rotz et al., 2023). The clarity and relevance of the items and the rating scale were reviewed by a panel of 34 researchers and practitioners, experts in the field of IDD. The analyses showed that all the experts judged the instrument relevant for their practice and that the rating scale, as well as the vast majority of items, were considered clear (>96%) and relevant (>94%), demonstrating a good level of content validity. The administration procedure for this study allowed for reformulations and examples with respect to the few items considered less clear/relevant.

The administration of the DABS-F was completed by a sociodemographic questionnaire that included questions about the administrator, the respondent, and the assessed individual. This questionnaire was developed on the basis of the one used by the original developers (Tassé et al., 2017) primarily in order to be able to describe the sample in a comparable way.

Data Analysis

As mentioned above, one of the main features of the DABS development has been the use of an IRT-based approach. The main aim of this study is to replicate the analyses performed by the original developers (Tassé et al., 2016). This includes the application of the graded response model (GRM, Samejima, 1969, 1997), which is an extension of IRT models for polytomous items. Specifically, we used the two GRM parameters that provide discriminant ability for each item and difficulty thresholds for each item response modality (from 0 to 3). The GRMs were performed separately for the three DABS domains (conceptual, social, practical). The assessment of these models and parameters (calibration) was conducted over the whole sample (calibration sample). These parameters were then used to compute the expected a posteriori (EAP; for the detailed scoring procedure, see Tassé et al., 2017) score of AB for each individual. As for the original version of the DABS, the scale of those scores was standardized using the 9 to 15 age group as a reference. The EAP score was then transformed into a standardized score with mean = 100 and standard deviation = 15, by subtracting the smoothed mean of the reference age group and then dividing by the smoothed standard deviation of the same group. These means and standard deviations were computed using the locally estimated scatterplot smoothing (LOESS) procedure (Cleveland, 1979) on a subsample (normative sample) in which we reduced the number of participants with ID in order to obtain a proportion of individuals with ID comparable to that of the general population (1%–2%). Specifically, the normative sample was composed of all participants with TD and one randomly selected participant with ID for each age group.

Before computing the GRMs on the DABS-F domains, it was necessary to complete two processes; namely, (a) the verification of the unidimensionality of the data, and (b) the analysis of the invariance of the link (or anchor) items. (a) Unidimensionality is an assumption in (unidimensional) IRT models. This refers to the fact that the items in a scale assess the same latent factor. We examined the unidimensionality of each DABS-F domain using three approaches: explained common variance (ECV), hierarchical ω (ω_h), and a recently developed index (unidim, u, Revelle & Condon, 2025). (b) Investigating the GRMs of a DABS-F domain requires the items common to two contiguous age versions of the scale to present the same parameters (parameter invariance). We investigated whether these common items actually presented the same parameter values using differential item functioning (DIF) analyses. The next paragraphs present a detailed explanation of those analyses.

In order to examine the unidimensionality indices mentioned above, we ran bifactor models using the Schmid–Leiman transformation (Schmid & Leiman, 1957) for each version of each DABS-F domain (3x domains, 3x versions) generating nine separate bifactor models. These models presented a structure characterized by one general factor common to all scale items, and several group (or specific) factors. Following the development of the original version of the DABS, each item can be attributed to one domain (conceptual, social, or practical), and also to a more specific factor related to this domain. For instance, the items of the conceptual domain for the 4 to 8 years old version obviously assess the main general factor (i.e., conceptual skills), but they can also be attributed to several specific domains, such as reading, language, self-direction, and time skills. In this example, we tested a structure composed by one general factor and four specific factors, following the original structure of the instrument. These analyses were conducted using the software R version 4.5.2 (R Core Team, 2025) and the psych package (Revelle, 2020).

In this context, ECV refers to the ratio between the variance explained by the general factor, and the sum of this same variance and the one explained by all specific factors. ECV values can range from 0 to 1. The higher the ECV value, the more the general factor shows a dominance in the factor structure and consequently supports a unidimensional factor structure. ω_h represents the proportion of general factor variance to total test variance; the higher the ω_h value, the more the variance of the model is explained by the general factor. We complete the unidimensionality evaluation by adding another, newer index that appears to be particularly robust across different numbers of specific factors and relatively small sample size ($n > 200$), which are two key features of our analyses. This u index is defined as the product of the congeneric model fit and the τ -equivalence and indicates the extent to which the items conform to a unidimensional measurement structure (for a discussion of the three indices, see Revelle & Condon, 2025). The three indices can be interpreted in the same manner, with higher values providing evidence of unidimensionality. We lack clear cutoffs for interpretation; however, values higher than .70 are generally read as an indication of substantial unidimensionality (e.g., Schroeders & Achaa-Amankwaa, 2026).

As mentioned, the second requirement prior to running the GRMs on the three DABS-F domains consisted of verifying the link item invariance. The Conceptual domain of the DABS-F presents nine items shared between the 4 to 8 years old and the 9 to 15 years old versions (e.g., “Relates experiences in narrative form (i.e., tells stories)”), as well as nine items common to the 9 to 15 years old and the 16 to 21 years old versions (e.g., “Uses ruler or tape measure to measure length”). The Social domain shares 11 items between the 4 to 8 years old and the 9 to 15 years old versions (e.g., “Accepts compliments”), and 10 items between the 9 to 15 years old and the 16 to 21 years old versions (e.g., “Maintains mutually satisfying relationships”). Finally, the Practical scale includes 18 items common to the 4 to 8 years old and the 9 to 15 years old versions (e.g., “Adjusts the water temperature at the faucet”), and seven items common to the 9 to 15 years old and the 16 to 21 years old versions (e.g., “Uses caution around electrical outlets”). Because of this, we conducted six DIF analyses to examine the parameter invariance of these link items, using IRTPRO software version 6.0 (Cai et al., 2011). The parameters were interpreted as significantly different (nonlink items) when the p of the comparison was lower than the critical value obtained using the Benjamini–Hochberg correction (critical BH, Benjamini & Hochberg, 1995). Conversely, p higher than critical BH indicated parameter invariance (link items) across different versions.

Results

Unidimensionality

The results of the analysis on unidimensionality showed that factor loadings on the general factor of all items were all positive and all above .50. For the vast majority of the items, these loadings were higher for the general factor than for the specific factor. This represents a first indication of substantial unidimensionality in the analyzed scales. Table 2 presents the unidimensional indices for the nine models. All models showed ω_h values above .70, and above .80 for eight models out of nine. These values were even higher for u , with all models presenting values above .80, and above .90 for eight models out of nine. ECV values were a little less

Table 2. Unidimensionality Indices of All Versions and Domains of the DABS-F

DABS-F domains	ω_h	u	ECV
Conceptual			
4–8	0.89	0.96	0.78
9–15	0.91	0.98	0.82
16–21	0.89	0.98	0.81
Social			
4–8	0.80	0.96	0.68
9–15	0.82	0.92	0.63
16–21	0.84	0.97	0.70
Practical			
4–8	0.89	0.95	0.82
9–15	0.72	0.83	0.61
16–21	0.84	0.93	0.80

Note. ω_h = hierarchical omega, u = unidim index, ECV = Explained Common Variance for the general factor.

strong than the other two indices; however, all were higher than .60 and were higher than .70 for the majority of the models. ECV appears to be more affected by the number of specific factors and small sample sizes, and this can probably explain these slightly lower values (Revelle & Condon, 2025). These results strongly support the assumption of unidimensionality in the DABS-F scales, particularly in view of the fact that perfect unidimensionality is more a theoretical ideal than a practical reality, and consequently this assumption should aim to demonstrate substantial rather than strict unidimensionality (for a discussion, see Ip, 2010).

Link Items

Table 3 presents the summary of the unique, link, and nonlink items resulting from the DIF analyses. Overall, the DABS-F shows that the vast majority of items common to two versions are invariant (i.e. their parameters are equivalent in the two versions of the scale). This was the case for 54 items out of 64. This result is very similar to the results obtained by the authors of the original version. For example, for the Conceptual domain, the only difference was that one item shared between the 9 to 15 and the 16 to 21 version is nonlink in this study, but was found to be invariant in the original version. The item in question is, “Makes a list (e.g. things to do, things to buy).” The Social domain presented a similar picture, while the Practical domain showed a number of nonlink items higher than the number found for the original version. The original version of the DABS reported only link items for this domain. The DIF analyses on the DABS-F revealed three nonlink items between the 4 to 8 and the 9 to 15 years old versions (e.g., “Adjusts the water temperature at the faucet”), and one between the 9 to 15 and the 16 to 21 years old versions (“Shows an understanding of when to dial 911 or equivalent emergency number to get emergency assistance (e.g. fire)”). Despite these marginal differences, these results confirmed that the majority of common items presented substantial invariance across versions. This means that the DABS-F seems to have a fair number of link items presenting a good level of discrimination (see Weber et al., 2026, for a discussion) enabling the contiguous versions of the three domain scales to be connected (or anchored).

Calibration

Discrimination and difficulty parameters for the three domain scales (Conceptual, Social, Practical) were estimated using GRMs utilizing IRTPRO software version 6.0 (Cai et al., 2011). We ran three GRMs using full information maximum likelihood as the estimation method, one model per DABS-F domain. For each model, we required the parameters of all link items to be equal across all three DABS-F domains and required the variance and mean of the group corresponding to the 9 to 15 years old version to be equal to 1 and 0, respectively. The results showed a good fit for all three models with the Steiger–Lind root mean square error of approximation (RMSEA) = .04 for conceptual skills, RMSEA = .06 for social skills, and RMSEA = .05 for practical skills. This index is considered to exhibit a good level of fit when its value is equal to, or lower than .06 (Hu & Bentler, 1999).

The discrimination parameters found with the DABS-F appear to be very similar to those found in the original version. In fact, the Conceptual domain of the DABS-F, for instance, showed discrimination parameters ranging from 1.27 to 5.37 for the 4 to 8 years old version, from 1.64 to 4.94 for the 9 to 15 years old version, and from 1.26 to 4.94 for the 16 to 21 years old version. The three versions of the original DABS presented the following ranges: 1.48 to 4.88, 2.00 to 4.88,

Table 3. Summary of Unique, Linked, and Nonlinked Items Across DABS-F Domains and Versions

Domain	Versions				
	4–8	4–8 and 9–15	9–15	9–15 and 16–21	16–21
Conceptual skills	16 unique	5 link 4 nonlink	7 unique	8 link 1 nonlink	16 unique
Social skills	14 unique	10 link 1 nonlink	5 unique	10 link	16 unique
Practical skills	7 unique	15 link 3 nonlink	0 unique	6 link 1 nonlink	18 unique

Table 4. EAP Means and Standard Deviations for the Scales of the DABS-F (and for the Original Version)

Versions	Conceptual skills		Social skills		Practical skills	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
4–8	–1.13 (–1.26)	0.93 (0.94)	–0.70 (–0.45)	1.18 (0.90)	–1.16 (–0.75)	1.04 (0.89)
9–15	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	1 (1)
16–21	0.33 (0.37)	0.94 (1.26)	0.24 (0.29)	1.06 (1.21)	0.04 (0.13)	0.97 (0.94)

and 1.40 to 5.95, respectively. In the 4 to 8 years old version, the most discriminant item of the original version was the second most discriminant in the DABS-F. Similarly, the least discriminant item of the original version was the second least discriminant item in the French adaptation. These two item couples presented very similar difficulty thresholds. The most discriminant couple showed $b_1 = -1.85$, $b_2 = -1.59$, and $b_3 = -1.41$, for the DABS; and $b_1 = -1.34$, $b_2 = -1.15$, and $b_3 = -0.84$, for the DABS-F. These parameters are similar, but we can consider that the item in the French version appears to be slightly more difficult relative to the DABS-F sample than the original version was relative to the original sample. The less discriminant couple showed $b_1 = -3.51$, $b_2 = -2.55$, and $b_3 = -1.27$, for the DABS; and $b_1 = -3.68$, $b_2 = -2.62$, and $b_3 = -1.54$, for the DABS-F. The difficulty level in the two versions appears to be similar for this item. Furthermore, all DABS-F items were monotonic (i.e., the difficulty parameters were ordered in the expected) and showed at least two difficulty thresholds (out of three) to be negative; that is, below the mean on the assessed ability level. The vast majority of the items actually showed three negative thresholds. This result confirmed that the DABS-F items measure the AB level below the mean and that they are discriminant around the cutoff point for determining significant limitations in AB. Therefore, considered overall, the calibration of the DABS-F seems to be consistent with that provided by the original developers.

The scales' EAP means and standard deviations also appear to be comparable with those found in the original version (Table 4). The only difference showing a higher mean for the DABS-F was found in the 4 to 8 years old version of conceptual skills. In all other cases, the original version shows higher means, with very small differences for the 16 to 21 years old versions, and slightly higher differences for the 4 to 8 years old versions. However, overall, these results seem to be consistent with those presented by the original developers.

Discriminant Validity

AB is defined as a collection of skills that tend to increase with age in the developmental period and then reach a plateau in adulthood (Tassé et al., 2012; Tassé et al., 2016). Therefore, we expected to find this tendency in the data that we collected using the DABS-F. Figure 1 presents the LOESS means of the EAP scores of the three DABS-F domains across age groups, and separately for participants with ID and with TD. LOESS curves were fitted to individual-level data, and 95% confidence intervals were estimated using bootstrap resampling (500 iterations). The three graphs were generated using the ggplot2 package (Wickham, 2016). First, we can see that the means of all domains actually increase with age. This increase seems to be more pronounced during early childhood and more stable when approaching adulthood, confirming expectations. Second, both participants with ID and those with TD appear to follow the same trajectory. Third, the curves of the two groups are parallel but clearly distinct, with the curve representing participants with ID being significantly lower than the curve for participants with TD. This latter interpretation is confirmed by statistically testing the difference between the standardized scores of the two groups on the three domains. The results indicated significant differences between the two groups with extremely high effect sizes (Cohen's d) for all domains. Specifically, independent t tests showed the following results: for conceptual skills, $t(883) = 34.09$, $p < .001$, $d = 2.75$; for social skills, $t(883) = 25.97$, $p < .001$, $d = 2.09$; and for practical skills, $t(883) = 24.45$, $p < .001$, $d = 1.97$. These results tend to confirm that the DABS-F is sensitive to expected differences in terms of age and the presence of ID, providing evidence of the discriminant validity of the instrument.

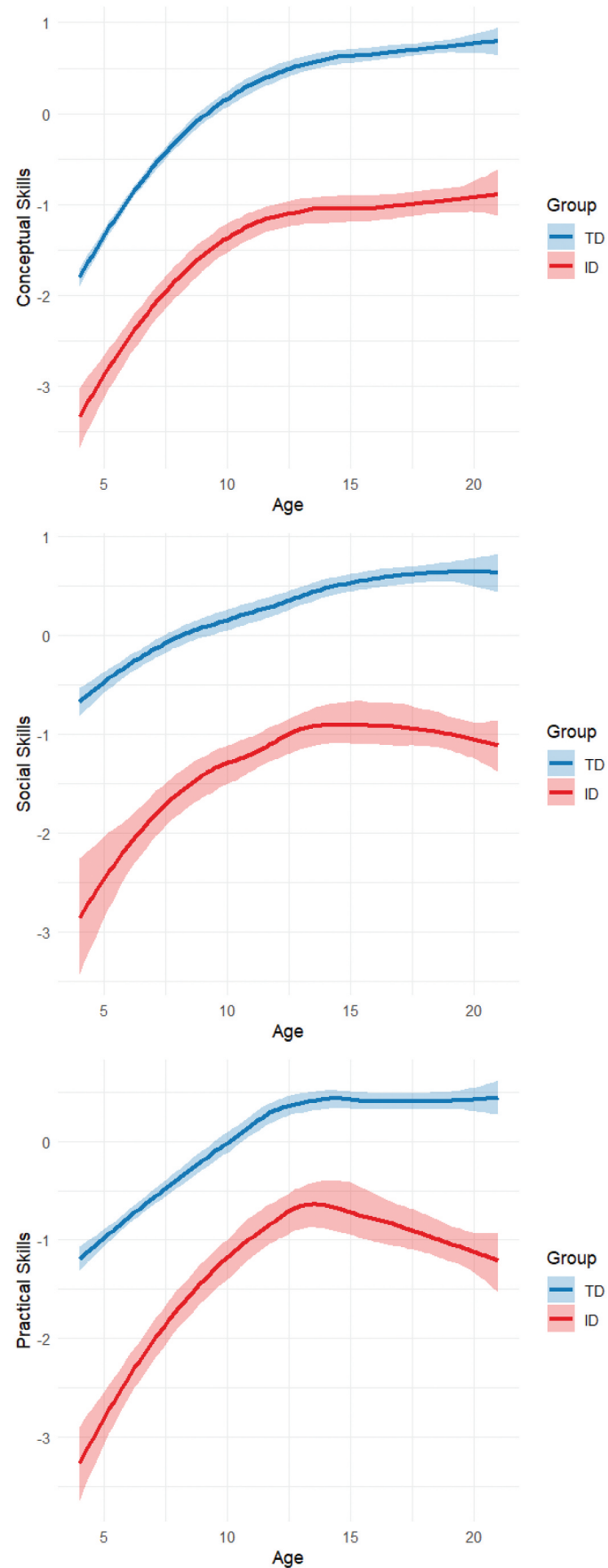


Figure 1. LOESS EAP means and bootstrapped 95% confidence intervals across age for Conceptual, Social, and Practical skills domains.

Discussion

The first stage in the validation process of the DABS-F was to adapt the original instrument to European French-speaking culture and provide evidence of the content validity of the French adaptation. The results obtained in this first stage (Von Rotz et al., 2024) allowed us to begin the main field test. The second stage in the validation process, presented in this study, consisted in administering the DABS-F to a large number of participants, building an overall sample stratified by age, presence of ID, and biological sex. The aim of the sampling procedure was to have at least 30 participants per year of age, with an overrepresentation in the proportion of participants with ID, and avoiding large differences in biological sex representation. The overrepresentation in the proportion of participants with ID was based on the need to obtain information in those levels of AB below the mean of the general population that are around the cutoff point for determining significant limitations in AB. The successful recruitment of this sample allowed us to start working on the core of the second stage of the validation process: the IRT calibration of the DABS-F. The aim of this study was to provide an IRT calibration of the French version of the DABS and compare it with the original version of the instrument. In fact, as stated by the authors of other French adaptations of measurement tools used in this field (e.g., Straccia et al., 2013) the goal of instrument adaptation is to demonstrate equivalence with the original version, rather than examining psychometric properties with reference to general thresholds or gold standards. This was particularly true with respect to this stage in the validation of the DABS-F, given that the use of IRT models was a key feature in the development of the original American version.

The results obtained in this study appear to provide evidence of this cross-cultural equivalence. First, the DABS-F domain scales appeared to measure a substantial unidimensional construct. This result is more important than a simple verification of a statistical assumption. In fact, the constructs that we assessed, Conceptual, Social, and Practical skills, are large and complex entities. It is easy to understand that the number of social skills, for example, that one person should master “in order to function in everyday life and to meet society’s expectations for people of the same age and culture” is enormous. The fact that we are not always aware of having and mastering some of those skills does not imply that their capture in an assessment instrument would be simpler. Given this complexity, the unidimensional levels found in this study, comparable with the original version, as well as other adapted versions (e.g., the Brazilian and Italian versions), are more than satisfactory and provided an initial confirmation of the internal consistency of the DABS-F. The second main result concerned the adjustment of the GRMs and the parameters arising from these analyses. The good levels of adjustment of the models of all domain scales should also be interpreted in relation to the complexity of the structure of those models. As far as this aspect is concerned, we do not have the fit indices for the original DABS. However, if we compare the present results with those obtained in the Brazilian adaptation (Hallberg et al., 2025), we observe that the fit indices that we found are equivalent or even indicate a slightly better fit. This comparison should, however, be viewed with caution, because the models tested in the two studies were not identical and were computed using different software. Finally, as mentioned in the Results section, the parameters of the DABS-F presented similar levels of discrimination and difficulty to those of the original version. More qualitatively, when comparing the more and the less discriminant items of the two versions, similar results were found, showing consistency at the items’ content level.

Obviously, this study also has certain limitations. One of these limitations refers to the lack of precision in the information that we collected concerning the severity levels of ID. Although the criteria that we used were published recently by recognized experts in assessment in the field of IDD (Tassé et al., 2019), we cannot be completely sure that these behavioral indicators have been interpreted and applied in an identical manner across the different research sites and across the number of social–educational institutions that were kind enough to collaborate on this project. However, the lack of information on the ID severity levels is not specific to this study; it should instead be seen as a representation of the reality in the countries in which the project took place. We do not have empirical studies examining ID diagnostic practices in European French-speaking countries at our disposal, but a report evaluating the state of the art in the field of IDD provided by French-speaking experts highlighted the fact that, for example, the evaluation of AB was not yet part of the current practice in ID diagnosis in French-speaking countries in Europe (Inserm, 2016). Despite this limitation, the IRT calibration process developed AB scores that seem to be able to discriminate between expected different levels, depending on age and presence of ID.

Given the nature of the DABS, the analyses using an IRT approach represented an essential step in the validation of the French adaptation. However, this does not mean that an IRT approach is hierarchically or epistemically superior to a classical test theory approach. These two approaches are complementary. Therefore, the third stage in the DABS-F validation process will be to examine the validity and reliability of the instrument. The original DABS provided strong evidence in support of convergent validity, internal consistency, and interrespondent and test–retest reliability (Tassé et al., 2016). We will continue to apply the same methodology in order to investigate these psychometric properties and to add more evidence of the validity of the European French version of the DABS, which has already been partially confirmed.

Acknowledgments

We thank all those who took part in the study. We are also grateful to the AAIDD and to the authors of the original version of the DABS for their permission to develop a French adaptation of the instrument. Special gratitude to professors Giulia Balboni and Pasquale Anselmi for their support and collaboration.

Author Contributions

CRediT: **Claudio Straccia:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing; **Lorna von Rotz:** Data curation, Investigation, Resources, Writing – review & editing; **Raphael Tsao:** Investigation, Project administration, Resources, Writing – review & editing; **Nathalie Touil:** Investigation, Project administration, Resources, Writing – review & editing; **Agnès Lacroix:** Investigation, Project administration, Resources, Writing – review & editing; **Vincent Desportes:** Investigation, Project administration, Resources, Writing – review & editing; **Annick Comblain:** Investigation, Project administration, Resources, Writing – review & editing; **Yannick Courbois:** Conceptualization, Investigation, Project administration, Resources, Writing – review & editing.

Funding

This study was financed by the Swiss National Science Foundation (Project 100019_188865, PI: Claudio Straccia). The salary of the second author of the study was in part financed by the Stiftung Heilpädagogisches Zentrum of Freiburg that we also thank.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Balboni, G., Bacherini, A., Anselmi, P., Brovedani, P., Buono, S., Micheletti, S., Robusto, E., & Tassé, M. J. (2022). Italian diagnostic adaptive behavior scale: Reliability and diagnostic accuracy compared with the vineland-II. *Research in Developmental Disabilities*, 123, 104185. <https://doi.org/10.1016/j.ridd.2022.104185>
- Balboni, G., Bacherini, A., Anselmi, P., & Robusto, E. (2023). *DABS diagnostic adaptive behavior scale. Validazione e standardizzazione italiana*. Hogrefe Editore.
- Balboni, G., Tassé, M. J., Schalock, R. L., Borthwick-Duffy, S. A., Spreat, S., Thissen, D. M., 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>
- 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 (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTpro for Windows*. Scientific Software International. <https://vpgcentral.com/>
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Hallberg, S. C. M., Schütz, A., de Mendonça Filho, E. J., Tassé, M. J., & Bandeira, D. R. (2025). Brazilian version of the DABS: Adaptation process and validity evidence based on the internal structure. *Research in Developmental Disabilities*, 163, 105041. <https://doi.org/10.1016/j.ridd.2025.105041>

- Harrison, P., & Oakland, T. (2003). *Adaptive behavior assessment system (ABAS-II)*. The Psychological Corporation.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Inserm. (2016). *Déficiences intellectuelles*. EDP Sciences.
- Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *The British Journal of Mathematical and Statistical Psychology*, 63(Pt 2), 395–416. <https://doi.org/10.1348/000711009x466835>
- Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H., Coulter, D. L., Craig, E. M. P., Reeve, A., Schalock, R. L., Snell, M. E., Spitalnik, D. M., & Spreat, S. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). American Association on Mental Retardation.
- R Core Team. (2025). *R: A language and environment for Statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2020). *Psych: Procedures for Psychological, psychometric, and personality Research*. Northwestern University. <https://doi.org/10.32614/cran.package.psych>
- Revelle, W., & Condon, D. (2025). Unidim: An index of scale homogeneity and unidimensionality. *Psychological Methods*, Advance online publication. <https://doi.org/10.1037/met0000729>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(S1), 1–97. <https://doi.org/10.1007/bf03372160>
- Samejima, F. (1997). Graded response model. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). Springer.
- Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H., Coulter, D. L., Craig, E. M., Gomez, S. C., Lachapelle, Y., Luckasson, R., & Reeve, A. (2010). *Intellectual disability: Definition, classification, and systems of supports*. ERIC.
- Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and systems of supports* (12th ed.). American Association on Intellectual and Developmental Disabilities.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61. <https://doi.org/10.1007/bf02289209>
- Schroeders, U., & Achaa-Amankwaa, P. (2026). Developing NOVA—A next-generation open vocabulary assessment. *European Journal of Psychological Assessment*, No Pagination Specified-No Pagination Specified. https://doi.org/10.31234/osf.io/vhakw_v2
- Schütz, A., Hallberg, S. C., Selau, T., de Mendonça Filho, E. J., & Bandeira, D. R. (2025). New psychometric evidence for the Brazilian version of the diagnostic adaptive behavior scale. *American Journal on Intellectual and Developmental Disabilities*, 1–11. <https://doi.org/10.1080/19447515.2025.2583716>
- Straccia, C., Tassé, M. J., Ghisletta, P., & Barisnikov, K. (2013). The French version of the reiss screen for maladaptive behavior: Factor structure, point prevalence and associated factors. *Research in Developmental Disabilities*, 34(11), 4052–4061. <https://doi.org/10.1016/j.ridd.2013.08.034>
- Takata, Y., Ansai, T., Soh, I., Awano, S., Nakamichi, I., Akifusa, S., Goto, K., Yoshida, A., Fujii, H., Fujisawa, 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. <https://doi.org/10.2147/cia.s43480>
- Tassé, M. J., Balboni, G., Navas, P., Luckasson, R., Nygren, M. A., Belacchi, C., Bonichini, S., Reed, G. M., & Kogan, C. S. (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., & 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., Luckasson, R., & Schalock, R. L. (2016). The relation between intellectual functioning and adaptive behavior in the diagnosis of intellectual disability. *Intellectual & Developmental Disabilities*, 54(6), 381–390. <https://doi.org/10.1352/1934-9556-54.6.381>
- Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., 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 & Developmental Disabilities*, 117(4), 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., Thissen, D., Widaman, K. F., & 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(1), 80–88. <https://doi.org/10.1111/jir.12239>
- Tassé, M. J., Schalock, R. L., Thissen, D., 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(2), 79–94. <https://doi.org/10.1352/1944-7558-121.2.79>

- Verdugo Alonso, M., Arias Martínez, B., & Navas Macho, P. (2014). La Escala de Diagnóstico de la Conducta Adaptativa (DABS): aplicaciones prácticas. *Siglo cero*, 45(249), 8–23. <https://hdl.handle.net/10366/163562>
- Von Rotz, L., Comblain, A., Courbois, Y., Des Portes, V., Lacroix, A., Touil, N., Tsao, R., & Straccia, C. (2024). The French version of the DABS: Validity and reliability preliminary analyses. *Journal of Intellectual Disability Research*, 68(7). <https://doi.org/10.1352/1944-7558-128.2.176>
- von Rotz, L., Courbois, Y., Des Portes, V., Lacroix, A., Reymond, M.-P., Tassé, M. J., Touil, N., Tsao, R., & Straccia, C. (2023). The French version of the DABS: Adaptation process and preliminary field test. *American Journal on Intellectual and Developmental Disabilities*, 128(2), 176–180. <https://doi.org/10.1352/1944-7558-128.2.176>
- Weber, D., Becker, N., Spinath, F. M., & Koch, M. (2026). The stability of IRT parameters under several test equating conditions [original Research]. *Frontiers in Psychology*, 16 - 2025. <https://doi.org/10.3389/fpsyg.2025.1652341>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
- World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11 ed.). <https://icd.who.int/>
- 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 [Research support, non-U.S. Government Research support, U.S. Government, P.H.S.]. *American Journal of Mental Retardation*, 107(3), 161–174. [https://doi.org/10.1352/0895-8017\(2002\)107%3C0161:IATPOA%3E2.0.CO;2](https://doi.org/10.1352/0895-8017(2002)107%3C0161:IATPOA%3E2.0.CO;2)

Adaptation française de l'Échelle diagnostique du comportement adaptatif : une approche fondée sur la théorie de la réponse aux items

Résumés en Français

Cette étude vise à évaluer, selon une approche fondée sur la théorie de la réponse aux items (TRI), les données recueillies à l'aide de la version française de l'Échelle diagnostique du comportement adaptatif (DABS-F) lors de la principale phase d'expérimentation du projet d'adaptation. Au total, 686 personnes au développement neurotypique et 199 personnes présentant une déficience intellectuelle ont été recrutées dans trois pays européens francophones, soit la Belgique, la France et la Suisse. Les données ont été analysées en reproduisant les analyses effectuées par les concepteurs de la version originale de la DABS. Les résultats montrent que l'unidimensionnalité, les items d'ancrage, les paramètres de discrimination et de difficulté issus des analyses fondées sur la TRI, ainsi que la validité discriminante de la DABS-F, étaient comparables aux résultats obtenus avec la version originale de l'instrument. Ces résultats sont discutés au regard de l'ensemble du projet d'adaptation de la DABS-F.

Adaptación francesa de la Escala de Diagnóstico de Conducta Adaptativa (DABS): un enfoque basado en la Teoría de Respuesta al Ítem (TRI)

Resúmenes en Español

El objetivo de este estudio fue evaluar los datos recopilados mediante la versión francesa de la Escala de Diagnóstico de Conducta Adaptativa (DABS-F) durante la prueba de campo principal del proyecto de adaptación, utilizando un enfoque basado en la Teoría de Respuesta al Ítem (TRI). Se reclutó a 686 participantes con desarrollo típico y a 199 participantes con discapacidad intelectual en tres países europeos de habla francesa (Bélgica, Francia y Suiza). Los datos se analizaron replicando los análisis realizados por los autores originales de la DABS. Los resultados mostraron que la unidimensionalidad, los ítems de enlace, los parámetros de discriminación y dificultad derivados de los análisis de la TRI, así como la validez discriminante de la DABS-F, fueron comparables a los resultados obtenidos con la versión original del instrumento. Estos resultados se discutieron en el contexto del proyecto global de adaptación de la DABS-F.