Development of an instrument for diagnosing significant limitations in adaptive behavior in early childhood[1]

Abstract

Although adaptive behaviour became a diagnostic criterion in the 5th edition of the American Association on Intellectual and Developmental Disabilities, AAIDD (Heber, 1959, 1961), there are no measures with adequate psychometric properties for diagnosing significant limitations in adaptive behavior according to the current conception of the construct. This fact has led to an excessive reliance on intellectual functioning measures. The goal of the present paper consists of presenting the development of the AAIDD's forthcoming Diagnostic Adaptive Behavior Scale (DABS) in Spain, and, specifically, it will be focused on one of its three forms: DABS Form 4-8 years old. The sample consisted of 388 people, aged 4 to 8 years old, with and without intellectual disabilities. The functioning of an initial pool of 168 items was analyzed under the assumptions of Item Response Theory models (IRT) with the aim to select those items around the cut-off point for determining significant limitations in adaptive behavior. A set of 72 items was selected (96 items were removed due to misfit, unsatisfying response category functioning, or low precision of measurement). The final version seems to be essentially unidimensional, shows good fit to the model, and represents an accurate precision of measurement around the cutoff point for diagnosing significant limitations in conceptual, social or practical skills.

Keywords: adaptive behavior; assessment; diagnosis; intellectual disability; item response theory

[1] This project has been founded by the Ministry of Science and Innovation, MICCIN (R&D Projects, 2009) (PSI2009-10953). It has been partially supported by a grant (EDU/894/2009) given to the Excellence Group of Research (GR197) by the Junta de Castilla y León (Spain).

1. Introduction

The diagnosis and classification of intellectual disability (ID) are main topics of interest to all those who try to understand this phenomenon and want to get a little closer to the complex intervention in this field. In educational and clinical settings, the diagnostic assessment of people with ID has always been a problematic issue for teachers and psychologists (Smith, 2005). In most cases, this fact lead to an excessive reliance on standardized intelligence measures (Greenspan, 1997, 2012) in order to make eligibility decisions. In fact, the measurement of intellectual functioning has sometimes played an exclusive role in the assessment of ID, reinforcing the mistaken belief of 'incurability' (Schalock, 1999). Consequently, the familiar categories of profound, severe, moderate, and mild intellectual disabilities have become in 'watertight compartments' (Navas, Verdugo, & Gómez, 2008) or a way to understand and react to the reality of having this condition (Florian & McLaughlin, 2008).

Intellectual functioning assessment does not capture by itself the nature of disability. Significant limitations in adaptive behavior, since their inclusion in the American Association on Mental Deficiency (AMMD) definition of intellectual disability (Heber, 1959, 1961) constitutes, along with significant limitations in intellectual functioning, and an age of onset before 18, a necessary (although not sufficient) criterion for a diagnosis of ID. These criteria are also included in major international classification systems (Lecavalier, Tassé, & Lévesque, 2002).

The inclusion of adaptive behavior as one of the main diagnostic criteria emphasized that intellectual disability involves deficits in intellectual functioning but also limitations in our ability to respond to environmental requirements. However, the operational definition of adaptive behavior has been debated throughout decades and has given rise to controversy regarding its assessment.

In order to provide an answer for this lack of agreement, some authors like Greenspan (Greenspan, 1997, 2006, 2012; Greenspan & Driscoll, 1997; Greenspan, Switzky, & Granfield, 1996) or Schalock (1999) started to define a theoretical framework of personal competence; consequently, establishing a strong parallel between intelligence and adaptive behavior domains, and specifying three types of skills that could define both constructs: conceptual, social and practical skills.

Although this model still presents too many questions that have to be resolved (Tassé, Schalock, Balboni, Bersani, Borthwick-Duffy, Spreat et al., in press a), it encouraged the development of factor analysis studies on the structure of adaptive behavior (Harrison & Oackland, 2003; Mathias & Nettelbeck, 1992; McGrew & Bruininks, 1999; Widaman, & McGrew, 1996). As a result of this multidimensional approach to the conceptualization of adaptive behavior, it is defined in 2002 American Association on Mental Retardation (AAMR, formerly AAMD) Manual (Luckasson, Borthwick-Duffy, Buntinx, Coulter, Craig, Reeve, et al., 2002) as the "collection of conceptual, social and practical skills that have been learned and are performed by people in their everyday lives" (Luckasson et al., 2002, p. 17). Likewise, significant limitations in adaptive behavior are defined as "a performance that is approximately two standard deviations below the mean of either (a) one of the following three types of adaptive skills: conceptual, social or practical or (b) an overall score on a standardized measure of adaptive behavior" (Luckasson et al., 2002, p. 17). This definition has been reaffirmed in the current definition of ID proposed by the American Association on Intellectual and Developmental Disabilities (AAIDD, formerly, AAMR) (Schalock, Borthwick-Duffy, Bradley, Buntinx, Coulter, Craig, et al., 2010, p.15; Verdugo & Schalock, 2010).

The agreement reached in defining the construct was also reflected in the establishment of a number of key issues that should be taken into account in its assessment

(Harrison & Boney 2002; Harrison & Raineri, 2008; Luckasson et al., 2002; Schalock, 1999; Schalock et al., 2010; Tassé, 2009; Tassé et al., in press a; Wehmeyer, Buntinx, Coulter, Lachapelle, Luckasson, Verdugo, et al., 2008): (a) adaptive behavior assessment should emphasize daily performance instead of maximum performance or thinking processes; (b) the assessment of adaptive behavior should take into account the opportunities that people with ID have to participate in community life according to their age; (c) as important as having opportunities is the cultural environment or context in which behavior takes place; therefore, the assessment should consider culturally relevant expectations, factors, and aspects related to the situational specificity in behavior assessment.

Specifically related to diagnosis function, the current manuals of ID (Luckasson et al., 2002; Schalock et al., 2010), as well as the forthcoming revision of the DSM (American Psychiatric Association, 2011), address the use of standardized measures with current norms that should be developed on a representative sample of the general population, thereby establishing the presence/absence of significant limitations in adaptive behavior (i.e., approximately two standard deviations below the mean). However, and despite the growing importance of norm-referenced adaptive behavior measures in the assessment of intellectual disability (Harrison & Boney, 2002; Harrison & Rainieri, 2008; Luckasson et al., 2002; Schalock et al., 2010), there are no standardized measures in accordance with its current definition or that are exclusively focused on diagnosis. Up to this point, approximately 200 adaptive behavior scales exist (Schalock, 1999), but there are only four adaptive behavior scales that have strong evidences of reliability and validity, and that have been standardized on the general population (Schalock et al., 2010; Tassé et al., in press a): (a) the Vineland Adaptive Behavior Scales (VABS) -2nd edition (Sparrow, Cicchetti, & Balla, 2005); (b) the Adaptive Behavior Assessment System- II (ABAS-II) (Harrison & Oakland, 2003); (c) the Scales of Independent Behavior-Revised (SIB-R) (Bruininks Woodcock,

Weatherman, & Hill, 1996); and (d) the Adaptive Behavior Scale-School Version (ABS-S: 2) (Lambert, Nihira, & Leland, 1993).

Concerned about this situation, the AAIDD began in 2006 a project focused on developing an assessment tool whose aim is to provide accurate diagnostic information around the cut-off point within an individual is deemed to have significant limitations in adaptive behavior. The forthcoming Diagnostic Adaptive Behavior Scale (DABS) (Tassé, Schalock, Balboni, Bersani, Duffy, Spreat, et al., in press b) is focused on identifying significant limitations in adaptive behavior for a diagnosis of ID, assessing those domains of adaptive behavior (i.e., practical, conceptual, and social skills) that have resulted from factor analysis work, and including relevant items related to the concepts of gullibility, vulnerability, and social cognition, which are critical to identify mild intellectual disabilities (Greenspan, 2006, 2012; Greenspan, Loughlin, & Black, 2001). Item Response Theory (IRT) has been used in its development to reliably measure individual levels of performance across the continuum of adaptive skills. Also, due to adaptive behavior increases in complexity with age and then plateaus (Luckasson et al., 2002; Schalock et al., 2010), items are being calibrated, according with IRT models, for the following age groups: DABS Form 4-8 years old; DABS Form 9-15 years old; and DABS Form 16-21 years old.

The DABS aims to become a measure exclusively focused on diagnosis that could help clinicians to assess a full range of adaptive behavior skills. An instrument with such characteristics would be very helpful in Spain, where there are no measures focused on diagnosis. In fact, the Inventory for Client and Agency Planning, ICAP (Bruininks, Hill, Watherman, & Woodcok, 1986) is the only measure that clinicians currently have in order to assess adaptive behavior, and their aim is not diagnosis, but rather services planning and program evaluation.

Due to this fact, the DABS is being developed in Spain following the development process already started by the AAIDD. The final goal is to develop a short and precise measure that comprises around 75 items to diagnosis significant limitations in adaptive behavior across three age ranges (i.e., 4-8 years old; 9-15 years old; and 16-21 years old). Because translation and adaption processes of the DABS into Spanish, in accordance with the guidelines of International Test Commission (Hambleton & Bollwark, 1991), have already been the scope of a previous article (Navas, Verdugo, Arias, & Gómez, 2010), the present paper is focused on the development of the DABS Form 4-8 years old in the Spanish context and the application of unidimensional IRT models to carry out item selection process.

2. **Method**

2.1 *Participants*

The development of the 3 Forms of the Diagnostic Adaptive Behavior Scale (DABS) in our country has involved the participation of 173 interviewers, 773 respondents and 1,047 people with and without intellectual disabilities between 4 and 21 years old. Specifically, the sample aged 4 to 8 years old (M = 5.88; DT = 1.37) was comprised of 388 people with (n = 164) and without (n = 224) intellectual disabilities. No participants were excluded based on etiology of ID, presence of sensory or motor impairments, or co-morbid conditions. All of them had already been diagnosed as having profound (n = 11), severe (n = 40), moderate (n = 46), or mild (n = 67) intellectual disability.

2.2 *Instrument*

The instrument used to measure adaptive behavior was a preliminary version of the DABS Form 4-8 years old and consisted of 168 items which were sorted into this age group, from an initial pool of 260 items, for which they had greater face validity.

According to previous factorial studies on the structure of adaptive behavior, the DABS Form 4-8 years old is divided into three subscales which are focused on the measurement of conceptual (69 items that assess language, reading and writing, money use, self-direction, time, numbers/measures, and problem-solving), practical (34 items that assess activities of daily living, maintains safe environment/safety, healthcare, schedules/routines), and social skills (65 items that assess interpersonal relationships, responsibility, self-esteem, wariness, naïveté, follows rules/obeys laws, manners, and social problem solving).

The DABS assesses individual's typical performance during daily routines and changing circumstances. Adaptive behavior skills are measured according to a four-point Likert scale from *rarely or never does it* (0) to *does it always or almost always independently* (3), taking into account the assistive technology that an individual could use in order to perform an ability (e.g., wheelchair, communication board).

2.3 *Procedure*

Most of the children (73.78%) were recruited from ordinary schools across 11 Spanish provinces by means of an incidental sampling.

The DABS administration procedure consisted of a face-to-face interview (carried out by qualified and trained) with one or more people who knew the person assessed very well (i.e., at least during the last 6 months), and who have had the opportunity to observe his or her behavior in a variety of settings.

2.4 *Test development and data analysis*

The Partial Credit Model (PCM) (Masters, 1982) was chosen to analyze items' and persons' functioning with the goal of selecting those items that will contribute to develop a short and precise measure of significant limitations in adaptive behavior. The PCM could be considered as a straightforward application of Rasch's model (van der Linden & Hambleton, 1997) and it has potential application in any context where the objective is to measure a latent trait through a process in which responses to items are scored with successive integers (Andrich, 1978).

2.4.1 *Prior verifications*

Prior selecting a subset of items around the cut-off point according to PCM, pertinent prior verifications of the model's data fit were carried out, which can be summarized in the following questions:

a) *Do response categories measure different levels of adaptive behavior effectively?* As Linacre (2002) pointed out, how the variable is divided into categories affects the measurement qualities of a test. Due to this fact, one of the early steps in our analysis consisted of an investigation into the functioning of rating scale categories. In this sense, PCM provides an effective framework to verify, and maybe improve, functioning of rating scale categories according to the following criteria: (a) at least 10 observations within each category and a regular observation distribution; (b) average measures advance monotonically with category; (c) *outfit* mean squares (MNSQ) less than 2.00. *Outfit* MNSQ is a fit statistic that means outlier-sensitive fit, and it is more sensitive to responses to items with difficulty far from a person, and vice-versa (Linacre, 2002).

b) *Do measures confirm the unidimensionality of each subscale? Is adaptive behavior enough to explain people's performance?* Due to DABS items are grouped into three subscales; we will

refer to each one (i.e., conceptual, social and practical skills subscales) as an essentially unidimensional latent variable. Although unidimensionality is a prerequisite for the application of unidimensional IRT models such as the PCM, in practice, the theoretical requirement of absolute unidimensionality of the scales is rarely fulfilled. A more realistic approach consists of demonstrating that there is a dominant factor that could explain the highest percentage of the variance. Some authors consider that unidimensionality exists when 40% - 60% of the variance is explained by measures (Hambleton, Swaminathan, & Rogers, 1991; Rubio et al., 2007), or when the first factor eigenvalue is five times bigger than the second factor eigenvalue (Hambleton et al., 1991). Unidimensionality can also be examined by performing Principal Component Analysis (PCA) on residuals, and it could be confirmed if the variance explained by measures is higher than 50%, and if unexplained variance is lower than 5% (Linacre, 2009).

2.4.2. *Item selection criteria*

With the aim to construct the shortest possible and most precise measure (i.e., around 75 items), after the above-mentioned verifications were examined, item selection was carried out within subscales (i.e, conceptual, social and practical skills) so that the final scales were as balanced as possible (i.e., around 25 items). Item selection criteria were based on the results of the following analysis:

a) *Assessing goodness of fit.* Residual analysis was carried out for assessing goodness of fit (i.e., the correspondence between model predictions and observed responses) according to residual statistics (MNSQ) *infit* and *outfit*. The former is sensitive to unexpected behaviors of those items located close to the skill level of the subjects, while the latter is more sensitive to responses to items with difficulty far from a person (Linacre, 2002). Item selection within subscales was carried out under the consideration that values substantially higher than 1 may indicate a lack of construct homogeneity with other items in a scale, and

smaller values may indicate item redundancy (Gómez, Arias, Verdugo, & Navas, in press; Linacre, 2002; Wilson, 2005).

b) *Selecting items that cluster around the cut-off point*. One of the main advantages of PCM is that persons and items can be mapped onto the same scale (i.e., statistically convenient log-odd units, called *logits*). *Logits* of greater magnitude represent increasing item difficulty (*b*), and higher levels of adaptive behavior (θ). Thus, the probability of a correct response could be calculated from the difference between person's adaptive behavior level (θ) and the item's difficulty (b_i). As the reader might expect, since the purpose of the DABS is diagnosis, those items involving tasks that require highest or lowest levels of adaptive behavior will not be useful for diagnostic purposes. For this reason, items around the cut-off point for determining significant limitations in adaptive behavior were selected.

c) *Analyzing Item Information Function (IIF)*: unlike Classical Test Theory (CTT), IRT assumes that measurement precision is not uniform across the variable. With the purpose of selecting those items contributing that will contribute to accurately measuring significant limitations in adaptive behavior, we analyzed IIF for each one of the 168 initial items included in DABS 4-8 years old Form, paying special attention to those items that accurately measure around the cut-off point.

These analyses were carried out for each one of the subscales (i.e., conceptual, social and practical) using WINSTEPS 3.68.1 software (Linacre, 2008).

3. Results

3.1. *Prior verifications*

a) Do item-category measures appear in the proper order according to the level of attribute?

The analysis of response categories effectiveness stated that 4 items had less than 10 observations in category response '0' (*rarely or never does it*): "*Follow verbal instructions"* and "*Follow simple one-step directions*" (within Conceptual Skills subscale); "*Acknowledges familiar people (e.g., smiles, nods, etc)"* (within Social Skills subscale); and "*Eats without making a mess"* (within Practical Skills subscale). It could signify that people have higher abilities than those being assessed with this response category.

Overall, average measures advanced monotonically with category, but the analysis of empirical item-category measures indicated that three items did not meet this assumption: "*Writes at least 20 words from memory"* and "*Read at least 25 words"* in Conceptual Skills subscale, and "*Controls bladder at night time*" in Practical Skills subscale. This means that, regarding these items, as adaptive behavior level increases, higher categories are not likely to be chosen, and as adaptive behavior level decreases, lower categories are not likely to be chosen.

On the other hand, 14 items from the initial pool of 168, had *outfit* MNSQ values greater than 2.00, suggesting that there is more unexplained noise than explained noise, and indicating that there is more misinformation than information in the observations (Linacre, 2005). Their inclusion in the final scale should be considered, because the use of their response categories is unexpected (e.g., people with lowest level of ability scored with a '3' in a difficult item).

b) Do measures confirm the unidimensionality of each subscale?

We carried out a parallel analysis with 1,000 replications by performing a Monte Carlo simulation, and eigenvalues for the three variables (Conceptual, Social and Practical Skills) were calculated. The results showed that more than 60% of the variance was explained by measures and the eigenvalue of the first factor (39.6, 38.72, and 19.73,

respectively) was five times bigger than the eigenvalue of the second factor (7.17, 4.14, and 2.01, respectively) in all the subscales. Principal Component Analysis (PCA) on residual scores indicated that more than 60% of the variance was explained by measures, and unexplained variance was lower than 5% (3.1%, 3.00%, and 3.1% for conceptual, social and practical subscale, respectively). Because the prerequisite of unidimensionality is rarely fulfilled, these results offered some support for the unidimensionality of the DABS.

3.2. *Item selection and test development*

a) *Assessing Goodness of Fit.* Goodness of fit analysis showed that 19 of the 69 items within Conceptual Skills subscale had *infit* or *outfit* MNSQ values above (i.e., there is unmodeled noise), and below (i.e, observations are too predictable) the acceptable levels (i.e., values substantially greater or lower than 1.00). Regarding Social Skills subscale, this percentage fell to 4.62% (3 of 65 items), whereas in the case of Practical Skills subscale this percentage was 11.76% (4 of 34 items showed misfit). These results are shown in Table 1.

<Table 1>

b) *Selecting items that cluster around the cut-off point for determining significant limitations in adaptive behavior.* As it is shown in item-person maps (Figure 1), items that are closer to the top are more difficult to endorse, and moving downward the items are easier to endorse. M, S and T parameters indicate the mean, one standard deviation and two standard deviations for items (right) and persons (left) in each subscale. As it can be seen, items are reasonably targeting the persons. However, some items are difficult for individuals to endorse (those that are closer to the top and closer to T parameter) and they will be performed only by those subjects with the highest level of ability, and, therefore, their inclusion in the diagnostic adaptive behavior measure is not recommended.

<Figure 1>

c) *Analyzing Item Information Function (IIF).* For diagnostic purposes, there is no interest in selecting those items that are accurately measuring higher levels of adaptive behavior. Due to that, mean and standard deviations for the Non-ID sample were calculated within subscale and domains, with the aim to choose those items that measure with highest precision around the sample mean and two standard deviations below it. This signifies, for example, considering the domain "Number and Measures" (which comprised 5 items), and individuals' scores in it (M = 1.02; SD = 1.24), that item 01("*Counts at least 10 objects, one by one*") and 02 ("*Makes accurate comparisons between two objects of different sizes*") measures with more precision those subjects around the cut-off point (Figure 2), while item 05 ("*Uses ruler or tape measure to measure length*") is more appropriate to assess those individuals with highest levels of adaptive behavior.

<Figure 2>

3.3 *Selected Items*

According to the results obtained from the detailed analysis in section 3.1 and 3.2, 72 items were chosen (a sample of them is shown in Table 2) whose *infit* MNSQ values were within 0.76 – 1.46. A total of 96 items were removed due to misfit, unsatisfying response category functioning, or low precision of measurement.

<Table 2>

Regarding the precision of measurement of the subset of selected items, Test Information Function (i.e., the sum of IIF at each level of adaptive behavior skills), indicated that we are measuring with precision (standard error of measurement lowest than .20) around the cut-off point (Figure 3), while greater error is associated with the highest

and lowest scores. In this figure, relative frequencies of different levels of abilities are shown, expressed by a bar chart that allows us to know the percentage of people receiving a specific score. Furthermore, these figures provide information about *logit* scores, standard scores and the standard error of measurement associated with each score. As can be seen in Figure, DABS appears especially appropriate for θ levels around the cut-off point whereas it measures less precisely those subjects with extreme values.

<Figure 3>

4. Discussion

The adoption of new approaches to the study of intellectual disability has highlighted the need for multidimensional classification systems based on the particular needs and circumstances of each individual. Although achieving a consensus on diagnostic instruments to be used in daily practice is not easy, it is essential to ensure the existence of at least instrument validity and reliability for the assessment process (Schalock et al., 2010).

The development of the Diagnostic Adaptive Behavior Scale in our context constitutes the first real effort to develop an instrument exclusively focused on the diagnosis of significant limitations in adaptive behavior. Item Response Theory allowed us to develop a precise measure around the cut-off point, including those items that maximize discrimination between the presence and absence of intellectual disability. Also, the analysis of item parameter estimates and item and test information curves provided information about the correlation between skills required in different age ranges and successful performance in everyday activities.

The implications of using a diagnostic measure like the DABS are profound given the consideration that it would be used for determining specific programs or service

eligibility criteria. For this reason, we would like to point out some of its strengths and limitations, being aware that there are aspects that need to be improved.

The major strength of this study lies in the methodology used. Because the main goal of DABS is to facilitate the diagnosis of intellectual disability through the identification of significant limitations in adaptive behavior, it was critical to develop an accurate and short measure in order to get maximum precision around the cut-off point. In this sense, Item Response Theory allowed us to analyze measurement precision across the scale. Although the selection of the sample was undertaken by means of an incidental sampling (and this could be considered as a weakness of the study), to contact those people who demonstrated wide interest in the study has also encouraged its spread, because, as a result, we held meetings with both federations and organizations which provide services to people with ID that allowed us to bring them much closer to adaptive behavior assessment, reminding them of the importance of this diagnostic criterion.

However, the study also has some limitations. We have already noted the first, about sample selection process. Secondly, another limitation identified during the data collection procedure was the difficulties in accessing the data from other measures of adaptive behavior in order to analyze concurrent validity of the scale. On the other hand, this leads us to emphasize the importance of using a multidimensional perspective when diagnosing, leaving behind systems that only include intelligence measures.

Thirdly, this study does not provide data for purposes of estimating inter-rater reliability, an issue that becomes critical for assessing reliability of any information provided by third party respondents. Also, it would be desirable to analyze test-retest reliability for testing the stability and reliability of the scale over time. We hope to solve these limitations through future research.

References

American Psychiatric Association. (2011). *American Psychiatric Association DSM-5 development.* Retrieved from: http://www.dsm5.org/ProposedRevision/Pages/proposedrevision.aspx?rid=384

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2,* 581-594. doi:10.1177/014662167800200413.

Bruininks, R. H., Hill, B. K., Weatherman, R. F., & Woodcock, R. W. (1986). *ICAP: Inventory for Client and Agency Planning. Examiner's Manual.* Allen, TX: DLM Teaching Resources.

Bruininks, R. H., Woodcock, R., Weatherman, R., & Hill, B. K. (1996). *Scales of Independent Behaviour-Revised.* Park Allen, TX: DLM Teaching Resources.

Florian, L. & McLaughlin, M. J. (2008). *Disability classification in education: Issues and perspectives.* Thousand Oaks, CA: Corwin Press.

Gómez, L.E., Arias, B., Verdugo, M.A., & Navas, P. (in press). Application of the Rasch Rating Scale Model to the Assessment of Quality of Life of Persons with Intellectual Disability. *Journal of Intellectual and Developmental Disabilities.*

Greenspan, S. (1997). Dead manual walking? Why the 1992 AAMR definition needs redoing. *Education and Training in Mental Retardation and Developmental Disabilities, 32*(2), 179-190.

Greenspan, S. (2006). Functional concepts in mental retardation: Finding the natural essence of an artificial category. *Exceptionality, 14*(4), 205-224.

Greenspan, S. (2012). How do we know when it's raining out? Why existing conceptions of intellectual disability are all (or mostly) wet. *Psychology in Intellectual and Developmental Disabilities 37*(2), 4-8.

Greenspan, S. & Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D. P. Flanagan & P. T. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 131-150). New York, NY: Guilford.

Greenspan, S. & Grandfield, J. M. (1992). Reconsidering the construct of mental retardation: Implications of a model social competence. *American Journal of Mental Retardation, 96*(4), 442-453.

Greenspan, S., Loughlin, G., & Black, R. S. (2001). Credulity and gullibility in persons with developmental disorders: A framework for future research. *International Review of Research in Mental Retardation, 24,* 101–135. doi:10.1016/S0074-7750(01)80007-0.

Greenspan, S., Switzky, H. N., & Granfield, J. M. (1996). Everyday intelligence and adaptive behavior: A theoretical framework. In J. W. Jacobson & J. A. Mulick (Eds.), *Manual of diagnosis and professional practice in mental retardation* (pp. 127-135). Washington, DC: American Psychological Association.

Hambleton, R. K. & Bollwark, J. (1991). Adapting tests for use in different cultures: technical issues and methods. *Bulletin of the International Test Commission, 18*, 3-32

Hambleton, R. K., Swaminathan, J., & Rogers, H. J. (1991). *Fundamentals of IRT*. Newbury Park, CA: Sage.

Harrison, P. L. & Boney, T. L. (2002). Best practices in the assessment of adaptive behavior. In A. T. J. Grimes (Ed.), *Best practices in school psychology IV* (pp. 1167-1179). Bethesda, MD: National Association of School Psychologists.

Harrison, P. & Raineri, G. (2008). Best practices in the assessment of adaptive behaviour. In A. Thomas & J. Grimes (Eds.) *Best practices in school psychology (5th Ed.)* (pp. 605-616). Bethesda, MD: NASP Press.

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

Heber, R. (1959). A manual on terminology and classification in mental retardation. *American Journal on Mental Deficiency Monographs, 64*(2).

Heber, R. (1961). *A manual on terminology and classification on mental.* Washington, DC: American Association on Mental Deficiency.

Lambert, N., Nihira, K., & Leland, H. (1993). *AAMR Adaptive Behavior Scales—School (ABS-S:2). Examiner's Manual* (2nd ed.). Austin, TX: Pro-Ed.

Lecavalier, L., Tassé, M. J., & Lévesque, S. (2002). Assessment of mental retardation by school psychologists. *Canadian Journal of School Psychology, 17*(1), 97-107. doi: 10.1177/082957350201700108.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106. Retrieved from: http://www.winsteps.com/a/linacre-optimizing-category.pdf.

Linacre, J. M. (2005). Dichotomous and politomous category information. *Rasch Measurement Transactions, 19,* 1005-1006.

Linacre, J. M. (2008). *Winsteps. Rasch measurement computer program* [computer software]. Chicago, IL: Winsteps.com.

Linacre, J. M. (2009). *A user's guide to Winsteps*. Retrieved from: http://www.winsteps.com/winman/index.htm?copyright.htm

Luckasson, R., Borthwick-Duffy, S., Buntix, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., et al. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington DC: American Association on Mental Retardation.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. doi: 10.1007/BF02296272.

Mathias, J. L. & Nettelbeck, T. (1992). Validity of Greenspan's Models of Adaptive and Social Intelligence. *Research in Developmental Disabilities, 13*(2), 113-129. doi:10.1016/0891-4222(92)90019-3

McGrew, K. S. & Bruininks, R. H. (1990). Defining adaptive and maladaptive behavior within a model of personal competence. *School Psychology Review, 19,* 53-73.

Navas, P., Verdugo, M.A., Arias, B., & Gómez, L. E. (2010). La conducta adaptativa en personas con discapacidad intelectual (Adaptive behavior in people with intellectual disabilities). *Siglo Cero, 41*(3), 28-48.

Navas, P., Verdugo, M.A., & Gómez, L. E. (2008). Diagnóstico y clasificación en discapacidad intelectual (Diagnosis and classification of intellectual disabilites). *Intervención Psicosocial, 17*(2), 143-152.

Rubio, V. J., Aguado, D., Hontagas, P. M., & Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure. An application of the Graded Response Model. *European Journal of Psychological Assessment, 23*(1), 39-46. doi: 10.1027/1015-5759.23.1.39.

Schalock, R. L. (1999). *Adaptive behavior and its measurement: Implications for the field of mental retardation*. Washington, DC: American Association on Mental Retardation.

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V., Buntix, W. H. E., Coulter, M. D., Craig, E. M., et al. (2010). *Intellectual Disability. Definition, classification and systems of supports.* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Smith, T. (2005). Assessment of individuals with mental retardation: Introduction to special issue. *Assessment for Effective Intervention, 30*(1), 1-4. doi: 10.1177/073724770503000401.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Adaptive Behavior Scales, Second Edition (Vineland-II)*. Circle Pines, MN: American Guidance Service.

Tassé, M. J. (2009). Adaptive behavior assessment and the diagnosis of mental retardation in capital cases. *Applied Neuropsychology, 16*(2), 114-123.

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Spreat, S., et al. (in press a). The Construct of Adaptive Behavior: Its Conceptualization, Measurement, and Use in the Field of Intellectual Disability. *American Journal on Intellectual and Developmental Disabilities.*

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Spreat, S., et al. (in press b). *Diagnostic Adaptive Behavior Scale: User's Manual.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer.

Verdugo, M.A. & Schalock, R.L. (2010). Últimos avances en el enfoque y concepción de las personas con discapacidad intelectual (Changes in the understanding and approach to persons with intellectual disability). *Siglo Cero, 41*(4), 7-21.

Wehmeyer, M. L., Buntinx, W. H. E., Coulter, D. L., Lachapelle, Y., Luckasson, R. A., Verdugo, M. A., et al. (2008). The intellectual disability construct and its relation to human functioning. *Intellectual and Developmental Disabilities, 46*(4), 311-318. doi: 10.1352/2008.46:311–318.

Widaman, K. F. & McGrew, K. S. (1996). The structure of adaptive behavior. In J. W. Jacobson & J. S. Mulick (Eds.), *Manual of diagnosis and professional practice in mental retardation* (pp. 97-110). Washington, DC: American Psychological Association.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Table 1.

Item misfit

Subscale	Statement	Infit MNSQ	Outfit MNSQ
CS	Answer yes or no when asked a simple question	1.06	2.53
	Uses three- our four-word sentences	.56	.39
	Uses prepositions in a sentence	.75	.42
	Writes first name, copying from an example	1.24	2.12
	Writes short notes o messages	.69	.44
	Follows written classroom or work schedule	1.00	2.58
	Arranges coins from highest to lowest values	.68	.42
	Arranges to get to an event or activity he or she likes	1.20	1.72
	Makes decisions about important life events	1.55	1.68
	Goes out during the day	1.78	2.81
	Takes appropriate action when an item is broken or malfunctioning (e.g., repair, discard, seek assistance)	1.59	2.01
	Shows an understanding of time equivalents (e.g., "9:15" is the same as "quarter past nine")	.72	.38
	Demonstrates understanding of temperature in relation to weather (e.g., that 100 degrees F. is very warm, 20 degrees F. is cold).	1.43	2.04
	Demonstrates understanding of weight (e.g., own body weight, weighing of produce)	1.63	1.52
SS	Goes out with friends	1.59	1.41
	Accurately evaluates own abilities (i.e., doesn't Over or under estimate)	.84	2.60
	Stands a comfortable distance from others during Conversations, as appropriate for culture	1.87	1.96
PS	Controls bladder at night time	1.58	3.55
	Uses the restroom in familiar settings	.90	2.06
	Uses the restroom in unfamiliar settings	1.28	2.43
	Rides only with people known to him/her.	1.60	2.29

Notes: CS = Conceptual Skills; SS = Social Skills; PS = Practical Skills

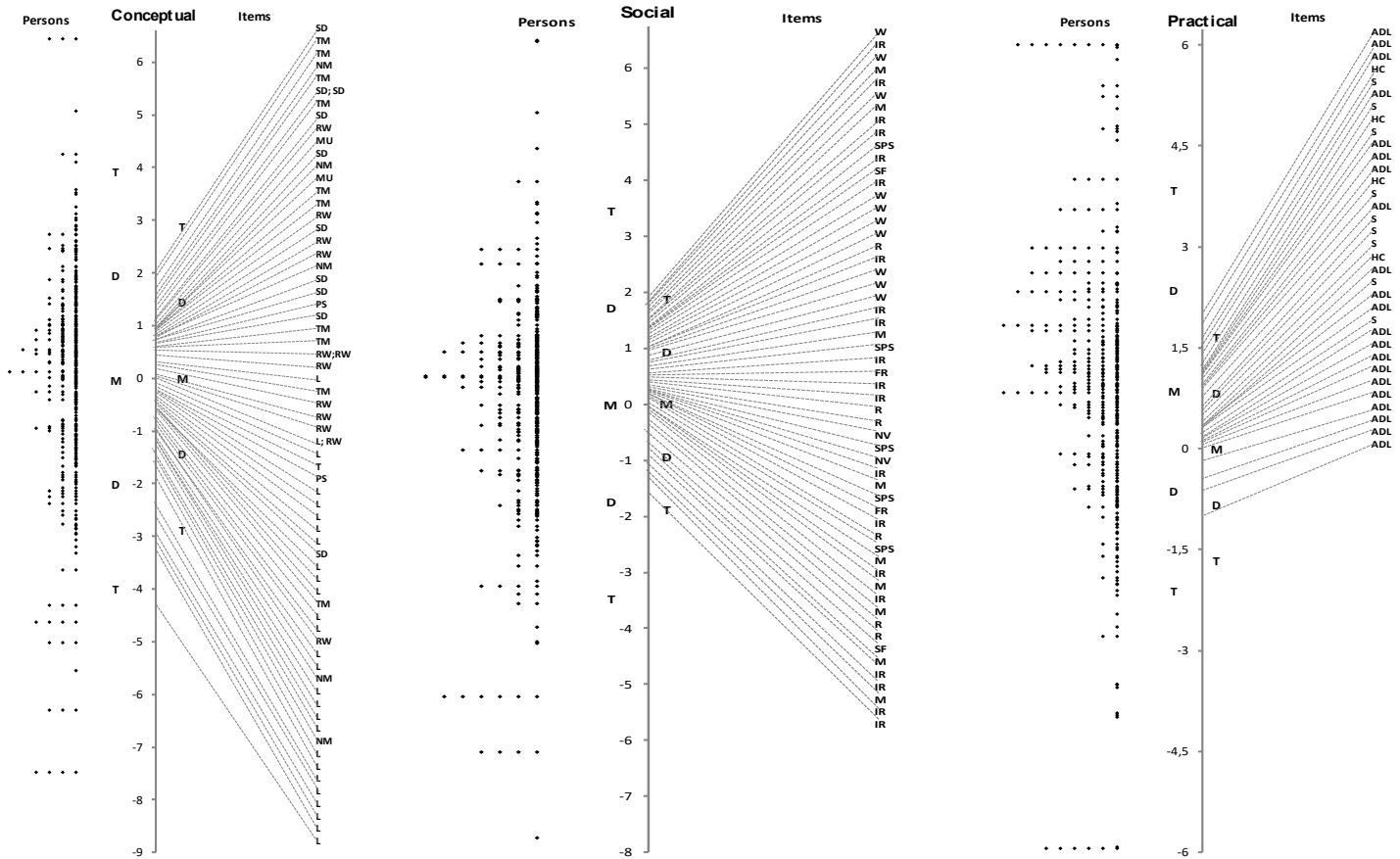

Figure 1. Item-person maps

Notes: M = Mean; S = Standard Deviation; T = Two Standard Deviations.
SD = Self-Direction; TM = Time; NM = Numbers/Measures; RW = Reading and Writing; MU = Money Use; L = Language; PS = Problem Solving;
W = Wariness; IR = Interpersonal; M = Manners; SPS = Social-Problem Solving; SF = Self-esteem; R = Responsibility; FR = Follow Rules; NV= Naïveté; ADL = Activities of Daily Living; HC = Healthcare; S = Safety

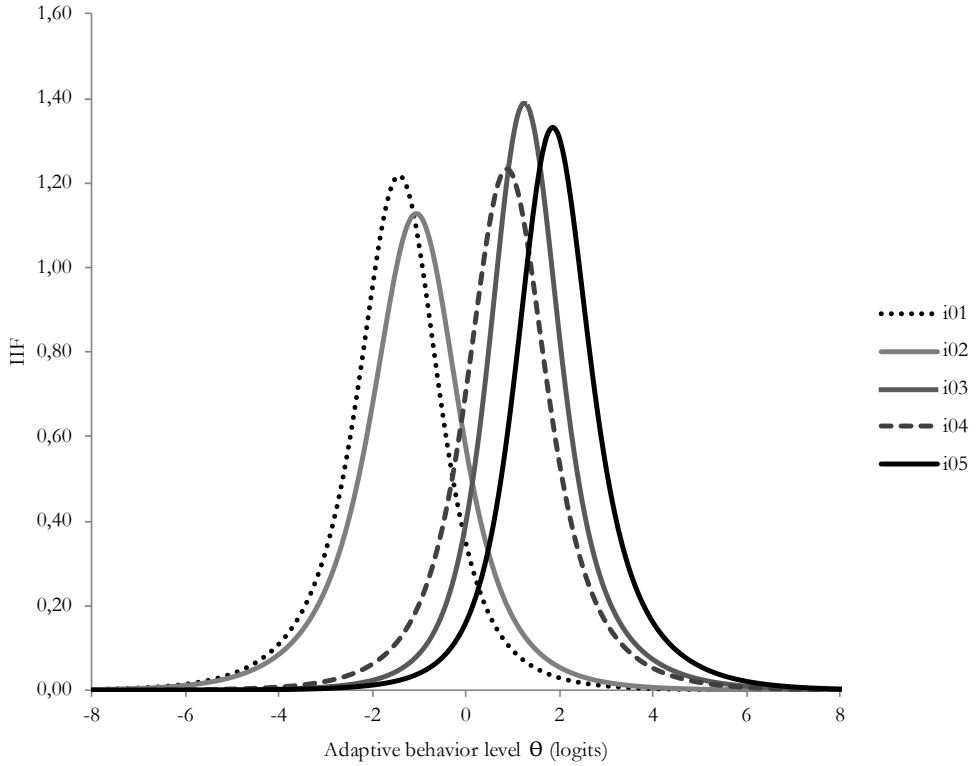

Figure 2. Item Information Function

Table 2.

Sample of selected items

Subscale	Domain	Item statement
Conceptual Skills	Language	Communicates complex ideas through oral, sign or written language
	Reading/Writing	Follows written classroom or work schedule
	Self-direction	Make plans to complete projects in logical steps
	Time	Refers to things he or she has done or will do within a specified time, such as last week or next month
	Numbers/measures	Counts at least 10 objects, one by one
	Problem solving	Can identify when there is a problem
Social Skills	Interpersonal	Stays on the topic in group conversations
	Responsibility	Modifies his/her behavior in accordance to social situation (e.g.. quiets down when entering a library. auditorium. etc)
	Self-esteem	Accepts compliments
	Wariness	Recognizes whom to trust when making a decision
	Follow rules	Follows rules and regulations when playing games
	Manners	Waits turn when waiting to use an object in use by someone else
	Social Problem-solving	Responds appropriately to humor
Practical Skills	Activities of daily living	Adjusts shirt/sweater if in-side is out
	Safety	Uses caution when around hot surfaces
	Healthcare	Communicates to others when not feeling well
	Routines	Follows a day schedule

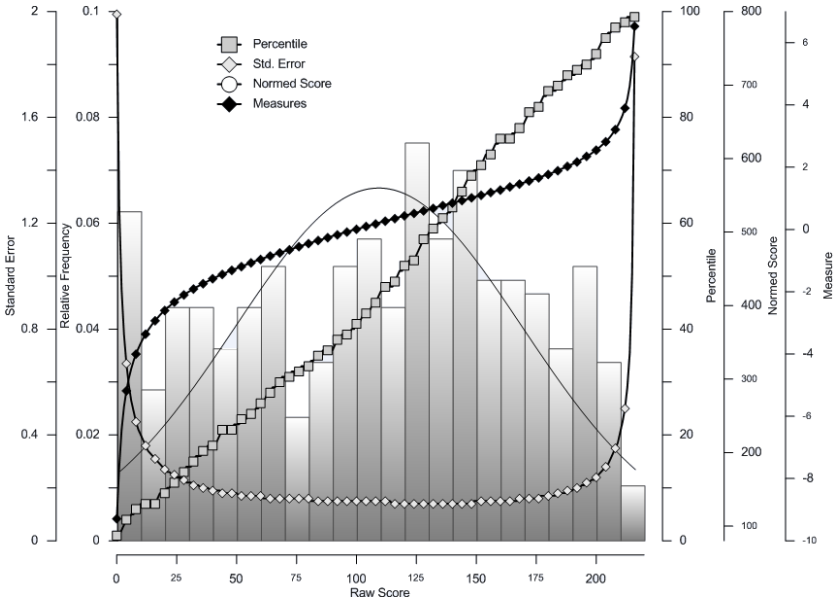

Figure 3. Test Information Function