



Published in final edited form as:

J Intellect Disabil Res. 2021 March ; 65(3): 283–295. doi:10.1111/jir.12810.

Concordance of the *Adaptive Behavior Assessment System*, second and third editions

A. M. von Buttlar^{1,2}, T. A. Zabel^{1,2}, A. E. Pritchard^{1,2}, A. D. Cannon¹

¹Department of Neuropsychology, Kennedy Krieger Institute, Baltimore, MD, USA

²Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

Background—Adaptive functioning is an important area of assessment with implications for differential diagnosis, educational placement, service eligibility and criminal sentencing. While periodic normative and content updates of adaptive functioning measures are necessary to keep measures relevant, knowledge of equivalence between versions is also required if adaptive measures are to be used to track the stability of adaptive functioning skills over time.

Method—This paper presents two studies that used between-group and within-group comparison designs to examine the equivalence of the second and third editions of the Adaptive Behavior Assessment System (ABAS) in a mixed clinical sample. In study 1, ABAS-2 scores for children assessed between 2014 and 2015 ($n = 1036$; mean age = 10.24, $SD = 3.44$) were compared with ABAS-3 scores for children assessed between 2015 and 2016 ($n = 1291$; mean age = 10.51, $SD = 3.70$). Study 2 examined a separate sample of clinically referred children ($n = 572$) for whom parent ratings had been obtained on both the ABAS-2 (mean age = 9.65, $SD = 2.80$) and ABAS-3 (mean age = 13.33, $SD = 2.95$) in the course of repeated assessment.

Results—For Study 1, while no intelligence quotient score differences were observed between the ABAS-2 group (mean Verbal Comprehension Index = 93.67, $SD = 16.95$) and the ABAS-3 group (mean Verbal Comprehension Index = 93.08, $SD = 17.42$), ABAS-2 scores were lower than ABAS-3 scores on the Conceptual, Practical, and General Adaptive Composite scales. In study 2, a similar pattern was observed (ABAS-2 < ABAS-3 on the Conceptual, Practical, and General Adaptive Composite scales), and concordance correlation coefficients ranged from 0.54 [0.49, 0.58] (Practical composite) to 0.68 [0.64, 0.72] (Conceptual composite). The Practical composite had the lowest concordance correlation coefficient value and the largest mean score difference between ABAS versions.

Conclusions—The ABAS-3 scores may be higher than ABAS-2 scores in clinical populations. Knowledge of these potential discrepancies will be critical when interpreting standard score changes across ABAS versions in the course of clinical, educational and forensic assessments.

Correspondence: Ashlee M. von Buttlar, Kennedy Krieger Institute, Department of Neuropsychology, 1750 E. Fairmount Ave, 3rd Floor, Baltimore, MD 21231, USA, (mitchellas@kennedykrieger.org).

Conflict of Interest

AV, TZ, AP, and AC report no real or potential conflicts of interest.

Keywords

adaptive behavior; concordance; developmental disorder; intellectual disability; psychometrics; reliability

Assessment of adaptive functioning is a key component in the process of diagnosing intellectual disability (ID). Although the diagnosis of ID (previously mental retardation) was initially based upon significantly below average intellectual functioning, the *Diagnostic and Statistical Manual of Mental Disorders – Third Edition (DSM-III)* and subsequent editions have included the additional criteria of impairment in adaptive domains (American Psychiatric Association 1980). Adaptive functioning refers to the ability of an individual to independently manage day-to-day demands and exhibit the ‘life skills’, which are appropriate for his or her age. The elements of adaptive functioning have been defined in varying ways since the introduction of this concept in the 1950s, but repeated factor analysis has validated adaptive functioning as being composed of three major domains: conceptual, social and practical skills (Thompson *et al.* 1999; Schalock 2011). The conceptual domain refers to communication abilities and functional academic skills. Social skills are one’s ability to participate in social interactions, develop and sustain relationships and use social information. The practical domain is composed of activities of daily living and household management. Unlike intellectual functioning, these adaptive functioning constructs are thought to be directly related to the amount and types of support or intervention an individual requires to function adequately in society.

Although DSM-5 does not identify a ‘cut score’ on standardised instruments for determining the presence of adaptive dysfunction, scores on adaptive score measures can still factor heavily into disability determination processes, criminal justice proceedings, educational placements, and so on. As such, equivalence between versions of adaptive skill measures is important, and knowledge of possible disparities between versions is needed in order to prevent false conclusions concerning adaptive skill improvement or decline over time. Recently, Farmer *et al.* (2020) identified a pattern of variable concordance between versions of the Vineland Adaptive Behavior Scales, as informant ratings of individuals with neurodevelopmental disabilities produced lower standardised scores on the Vineland-3 (Sparrow *et al.* 2016) relative to the Vineland-2 (Sparrow *et al.* 2005). To date, however, there has yet to be an examination of the agreement between the current (third) version of the *Adaptive Behavior Assessment System (ABAS)*, that is, the ABAS-3 (Harrison and Oakland 2015), and the version that preceded it, that is, the ABAS-2 (Harrison & Oakland, 2003).

According to the ABAS-3 manual (Harrison and Oakland 2015), the ABAS-3 was developed based upon the content and structure of the ABAS-2, with several noteworthy updates. The item content was revised with the objectives of improving the accuracy of assessment of persons with lower or higher cognitive ability, better assessing the adaptive deficits associated with three prevalent conditions (i.e. ID, autism spectrum disorder and attention-deficit hyperactivity disorder) and updating the content to reflect newer life skill-related technologies (e.g. use of the Internet for research activities in lieu of encyclopaedias). The

ABAS-3 manual indicates that 80% of the 232 ABAS-3 Parent Form items are composed of original ($n = 138$) or revised ($n = 48$) ABAS-2 items, with the remaining 20% accounted for by introduction of 46 new items. The distribution of new and revised items across adaptive domains and individual adaptive skill areas scale is not reported in the ABAS-3 manual.

Equivalence studies comparing versions 2 and 3 of the Parent Form (ages 5 to 21 years) are described in the ABAS-3 manual (Harrison and Oakland 2015). While scores from the ABAS-2 and ABAS-3 were well correlated (corrected correlation of 0.72), these studies suggested that a degree of standardised score increase when parents rated their children on both the ABAS-2 and the ABAS-3. Of note, mean score increases of two scaled score points or higher were noted between ABAS-2 and ABAS-3 versions in the skill areas of Community Use and Home Living, and a mean score increase of 5 standard score points was noted in the Practical domain.

The aim of this study was to examine the equivalence of two versions of the ABAS in a mixed clinical sample. As the ABAS-2 and ABAS-3 are frequently used as outcome measures in treatment and rehabilitation settings, assessing the equivalence of these measures in a clinical sample is essential for accurately monitoring change over time. In our clinically referred sample, we tested the equivalence of the ABAS-2 and the ABAS-3 in two ways. In study 1, we examined the equivalence of the ABAS-2 and ABAS-3 using a between-groups design. As the ABAS-3 was made available for use in 2015, we compared the standardised scores of individuals for whom the ABAS-2 was administered between 2014 and 2015 with the scores of individuals for whom the ABAS-3 was administered between 2015 and 2016. In study 2, we used a within-group design and examined the equivalence of ABAS-2 and ABAS-3 standardised scores among those individuals for whom both the ABAS-2 and ABAS-3 had been administered within the context of serial/repeated clinical assessment. Based upon the comparison studies published in the ABAS-3 manual, we hypothesised that ABAS-3 scores would be higher than ABAS-2 scores in both studies.

Study 1: between-groups comparison of Adaptive Behavior Assessment System 2 and 3

Methods

Participants—Children in study 1 were referred for psychological/neuropsychological assessment and underwent assessment between 2014 and 2016 at a regional hospital in the Mid-Atlantic region of the US. Children were eligible for study 1 if their parent or caregiver had provided one set of adaptive skill ratings on the ABAS-2 or the ABAS-3 between 2014 and 2016. Children were excluded from study 1 if, in the context of repeated clinical assessment, parent ratings were provided on both the ABAS-2 and ABAS-3 (these individuals were included in study 2 instead). Participant data were included in the study 1 dataset if the individual's clinical evaluation included an estimate of verbal intelligence [i.e. Wechsler Intelligence Scales for Children – Fourth Edition (WISC-4) or Wechsler Intelligence Scales for Children – Fifth Edition (WISC-5) Verbal Comprehension Index (VCI)] and all four composite indices [i.e. general adaptive composite (GAC), Conceptual, Social, and Practical] from either the ABAS-2 or ABAS-3. Children in study 1 ($n = 2327$)

ranged in age from 6 to 16 years ($M = 10.4$ years, $SD = 2.74$ years), and 66.8% were male (Table 1). Fifty-five per cent were Caucasian, with the remaining mostly Black/African-American (27.8%), multi-racial (6.9%), Hispanic/Spanish Origin (3.2%) or Asian (3.0%). During the designated time frame, ABAS-2 and ABAS-3 scores were available for 1036 and 1291 children, respectively.

Procedure—All data were maintained in a Health Insurance Portability and Accountability Act-complaint database. Standardised test scores (described later) acquired via clinical assessment were maintained in the secure electronic health record. A waiver of consent to study these de-identified clinical data was granted by the local institutional review board.

Intelligence.—General intelligence was measured on the day of assessment using the WISC-4 (Wechsler 2003) or the Wechsler Intelligence Scales for Children-Fifth Edition (WISC-5). The WISC-4 was published in 2003 and was used at the noted regional Mid-Atlantic hospital up until the release of the WISC-5 in 2015. Both the WISC-4 and the WISC-5 generate a full-scale intelligence quotient (IQ) as well as specialised index/factor scores. For the purposes of this study, the VCI score was used as an estimate of intellectual functioning. This was possible as a VCI can be calculated within the context of both a WISC-4 and WISC-5 administration. The VCI is a standard score with a mean of 100 and a standard deviation of 15. Although the subtest composition of the VCI differs between the WISC-4 (typically similarities, vocabulary and comprehension, but with possible subtest substitution) and WISC-5 (i.e. similarities and vocabulary), the WISC-5 manual reports acceptable corrected correlation values (i.e. 0.85) in a sample of children who took both the WISC-4 and WISC-5 (Wechsler 2014).

Adaptive functioning.—Parent/caregiver ratings of adaptive skills and abilities were obtained using the ABAS-2 (Harrison & Oakland, 2003) or ABAS-3 (Harrison and Oakland 2015). The ABAS is a comprehensive assessment of a wide range of adaptive skills. Several forms are available, including a Parent Form for acquiring information from caregivers for children and adolescents between the ages of 5 and 21. On the Parent Form, adaptive behaviour is measured at several levels, including an overall GAC, three adaptive domains (Conceptual, Social and Practical) and nine individual adaptive skill areas (with the optional addition of a tenth adaptive skill area, i.e. work). The GAC and adaptive domain composite scores are standard scores ($M = 100$, $SD = 15$), and the individual adaptive skill areas are scaled scores ($M = 10$, $SD = 3$).

The ABAS-3, published in 2015, remains the most current version of the ABAS instrument available for clinical use. For the standardisation sample, the ABAS-3 manual reports good internal consistency (reliability coefficients ranging from 0.89 to 0.99 across adaptive domains and skill areas) and test–retest reliability (corrected test–retest correlations ranged from 0.73 to 0.85, with average standard score point differences ranging from 0.86 in the Conceptual domain to –0.12 in the Practical domain) for the Parent Form (ages 5–21). Construct validity has been well established, including convergent validity demonstrated via strong correlations with the Vineland Adaptive Behavior Scales, Second Edition (Vineland-2, Sparrow *et al.* 2005). The standardisation sample of the ABAS-3 included individuals with mild disabilities, ‘as long as the severity [did] not preclude mainstream

activities (such as general education)' (Harrison & Oakland, 2015, p. 60), although separate validity evidence is provided for several clinical groups in both versions (e.g. autism spectrum disorder, ID, attention-deficit hyperactivity disorder; Harrison & Oakland, 2003; Harrison & Oakland, 2015).

For the standardisation sample of the ABAS-2 Parent/Primary Caregiver and Parent Forms, the manual also reports good internal consistency (reliability coefficients ranging from 0.80 to 0.98 across adaptive domains and skill areas) and test–retest reliability (corrected test–retest correlations ranged from 0.84 to 0.93, with average standard score point differences ranging from 1.6 in the Social domain to 1.0 in the Practical domain), as well as strong convergent validity with the Vineland Adaptive Behavior Scale (Sparrow *et al.* 1984). The equivalency data presented in the ABAS-3 manual comparing it with the ABAS-2 suggest strong correlations among adaptive domain scores between versions (ranging from 0.83 to 0.88).

This study used the paper version of the Parent Form of the ABAS according to the standardised procedures identified in the ABAS-3 manual. It was utilised in the clinical activities of licensed psychologists and/or their supervised trainees. If participants had more than one available ABAS-2 or ABAS-3, their most recent ABAS-2 and earliest ABAS-3 were selected for analyses.

Data analytic strategy—Independent-samples *t* tests were used to examine differences in ABAS-2 and ABAS-3 ratings between composite scores, including 95% confidence intervals of the difference between means. Effect sizes were calculated using Cohen's *d* using pooled standard deviations.

Results—Results of study 1 are depicted in Tables 2 and 3. In this mixed clinical sample, the mean VCI was in the low 90s for children rated on the ABAS-2 ($n = 1036$, mean VCI = 93.67, $SD = 16.95$) and for children rated on the ABAS-3 ($n = 1291$, mean VCI = 93.08, $SD = 17.42$), and there was no indication of a difference between groups with regard to verbal intelligence (Table 1). GAC scores were higher in the ABAS-3 cohort as compared to the ABAS-2 cohort (scaled score difference of 5.37 points; see Table 2). Similar differences were noted on the Conceptual and Practical domain scales, but not the Social scale. In an exploratory analysis of the individual skill areas, the largest differences between ABAS versions were noted on the Home Living, Community Use, Self-Direction and Self-Care scales (Tables 3 and 7). Of note, three of these four skill area scales also comprise the Practical domain (i.e. Community Use, Home Living, Health and Safety and Self-Care).

Study 2: within-group comparison of Adaptive Behavior Assessment

System 2 and 3

Participants

Participants in study 2 ($n = 572$; Table 4) were referred for psychological/neuropsychological evaluation and underwent assessment between 2006 and 2019 at a regional hospital in the Mid-Atlantic region of the US. Participants were eligible for study 2

if: (1) they had been assessed two times and their parent or caregiver had provided two sets of adaptive skill ratings, one on the ABAS-2 (time 1) and one on the ABAS-3 (time 2); and (2) the individual's clinical evaluation included all four composite indices from both the ABAS-2 and the ABAS-3. There was no overlap between study 1 and study 2 participants.

Participants in study 2 ranged in age from 5 to 21 years, and 67.9% were male (refer to Table 4 for demographic information). The majority of the participants were Caucasian (56.9%), followed by Black/African-American (26.2%), Hispanic (4.4%) and Asian (3.5%). Approximately 6% of the sample identified as multi-racial. The average age of participants at time 1 was 9.65 years ($SD = 2.80$) and was 13.33 years ($SD = 2.95$) at time 2. The average number of years between evaluations was 3.68 years ($SD = 1.61$, range = 0–11.34 years). The median number of years between evaluations was 3.43 years [interquartile range (IQR) = 2.63–4.61 = 1.98]. The correlation between time to follow up and absolute change in GAC scores between ABAS versions was quite small ($r = 0.043$, $P = 0.30$, 95% confidence interval [-0.04, 0.12]), and this suggests no meaningful relationship between changes in ABAS-2 and ABAS-3 ratings and the amount of elapsed time between evaluations. Therefore, time to follow up was not included in subsequent analyses. Similar to study 1, the mean VCI for the sample was in the low 90s (mean VCI = 91.29, $SD = 18.02$).

Procedure

The same instruments and procedures were used as those described in study 1 above.

Statistical analysis

Paired-samples *t* tests were used to examine differences in ABAS-2 and ABAS-3 ratings between composite scores, including 95% confidence intervals of the difference between means. Effect sizes were calculated using Cohen's *d* using pooled standard deviations. Concordance correlation coefficients (CCCs) were calculated for all ABAS composites to estimate the degree of agreement between ratings on ABAS versions. A value of 1.0 reflects exact agreement between tests.

Results

Results of study 2 are depicted in Tables 5 and 6. As hypothesised, we found differences between ABAS-2 scores and ABAS-3 scores in overall adaptive functioning as measured by the GAC, with higher scores (mean standard score increase of 6.7 points) observed on the ABAS-3. Scores on the ABAS-3 were also higher than the ABAS-2 on the Conceptual composite (mean scaled score increase of 3.3 points) and Practical composite (mean scaled score increase of 8.6 points), with small-to-medium effect sizes, respectively (average standard scores and effect sizes are presented in Table 5). Contrary to our hypotheses, there were no differences in ABAS-2 and ABAS-3 ratings on the Social composite.

Concordance correlation coefficients of ABAS composites ranged from 0.54 [0.49, 0.58] (Practical composite) to 0.68 [0.64, 0.72] (Conceptual composite; Table 5). The CCC 95% confidence bands of the Conceptual and Social composite scales overlapped, but the CCC 95% confidence band of the Practical composite fell below the confidence bands of both of the Conceptual and Social composite scales, making it the composite with the most

discordance between ABAS versions. Not surprisingly, the composite scale with the lowest CCC (Practical) also had the largest mean score difference (8.5 standard score points) between ABAS versions.

In order to better understand the individual factors driving these composite score differences, paired-samples *t* tests were used to assess any differences between ABAS-2 and ABAS-3 subscales. These analyses identified mean differences between all subscales at time 1 and time 2, with higher scaled scores on the ABAS-3, with the exception of the Leisure subscale (Table 6). Scaled score point changes ranged from -0.21 (Leisure) to 2.24 (Home Living), with an average scaled score point increase of 1.06. The largest effect sizes were found on subscales included in the calculation of the Practical and Conceptual composites, as would be expected given the differences between ABAS versions on these composites. Specifically, approximately 33% of the sample ($n = 191$) had a Practical composite score that moved from the 'below average' range or below ($SS = 79$ or below) to the 'low average' range or higher ($SS = 80$ or above) between their two evaluations. Effect sizes were moderate-to-large for Community Use and Self-Direction (both included in calculation of the Conceptual composite), as well as for Home Living, Community Use and Self-Care (used in calculation of the Practical composite; refer to Table 7 for side-by-side comparisons of studies 1 and 2 mean differences and effect sizes).

Studies 1 and 2 sensitivity analyses

Procedure

Additional sensitivity analyses were also conducted for both studies 1 and 2. Given the retrospective nature of these data, the only available diagnostic information was the billing diagnosis that was used for the clinical visit. Because ID is typically not a billable diagnosis in the US, we have not reported billing diagnostic information for fear that it would underestimate the number of individuals with an ID diagnosis in our sample. Therefore, although we recognise that a low IQ score is only one of the necessary components of a diagnosis of ID, we chose to split both samples at an IQ standard score estimate of 75 as a way of better characterising participants who may be at risk for a diagnosis of ID. Participants from each study were separated into two groups based on VCI (at or below 75 and at or above 76), and analyses described earlier were repeated.

Results

Findings from the sensitivity analyses were similar to the main analyses of studies 1 and 2. Specifically, when compared with ABAS-2 scores, ABAS-3 scores were higher on the GAC, Conceptual and Practical composites in both study 1 (Table 8) and study 2 (Table 10). Of note, these mean ABAS composite score differences were qualitatively larger in the sample of participants with a VCI of 75 or lower; a similar pattern was found for ABAS subscales (Tables 9 and 11).

Discussion

This research is, to our knowledge, the first independent study examining the equivalence of the ABAS-2 and ABAS-3 in mixed clinical samples using both between-subjects (study 1)

and within-subjects (study 2) designs. When compared with ABAS-2 scores, these two studies found higher ABAS-3 scores for overall adaptive functioning (the GAC), as well as for the Conceptual and Practical composites. Of note, the Social composite did not differ between ABAS versions in either study. As such, these studies revealed evidence of a systematic positive age-corrected score increase when ABAS-3 scores are compared to ABAS-2 scores in a clinical setting. Consistent with the trend of higher ABAS-3 scores noted in the ABAS-3 manual's ABAS-2/ABAS-3 comparability study, the results of the current study provide additional evidence of a pattern of discrepancy between these two versions of the ABAS. These findings have clinical relevance, as changes in adaptive skill ratings are often used to document clinical improvement or decline, and changes of this type can factor heavily into disability determination processes, criminal justice proceedings and educational placement decision-making.

The pattern of standardised score increase between ABAS-2 and ABAS-3 versions was greater in some adaptive domains than others and was most evident on the Practical composite. The ABAS Practical composite is an important indicator of an individual's self-care, home living and community use skills, and scores on this scale are used clinically to gauge the person's ability to manage their activities of daily life. Across both studies, the largest ABAS-2/ABAS-3 mean score differences were noted on the Practical composite, with mean standard score ABAS-2/ABAS-3 differences of +7.22 for study 1 and +8.6 for study 2. Moreover, while the overall concordance (CCC) coefficients varied between ABAS composite scales in study 2, the lowest concordance indicator came from the comparison of the ABAS-2/ABAS-3 Practical scales (CCC = 0.54, [0.49, 0.58]). Given this pattern of lower version concordance and increased standard score discrepancy, we believe that the findings of these studies suggest that comparisons of Practical composite scores across ABAS versions 2 and 3 are the most problematic clinically.

Although relevant for our mixed clinical samples as a whole, the differences noted between ABAS-2 and ABAS-3 scores were particularly noteworthy for subsamples that had IQ (WISC-4 or WISC-5 VCI) standard score estimates of 75 or lower. In both studies, caregiver ratings of these individuals at risk for ID produced higher scores on the ABAS-3 than the ABAS-2. For instance, in the study 2 (within groups) subsample of individuals with a VCI standard score of 75 or below, the average ABAS-2 ($M = 65.12$, $SD = 13.95$) and ABAS-3 ($M = 73.95$, $SD = 12.71$) GAC scores differed by over +8 standard score points, that is, over half a standard deviation. This was also true for the Conceptual composite (+5 points) and the Practical composite (+11 points).

Discrepancies between ABAS-2 and ABAS-3 composite scales have implications for how the adaptive functioning of an individual is qualitatively described, particularly for persons at risk for an ID diagnosis. For example, approximately 70% of study 2 participants with a VCI of 75 or lower had an ABAS-2 GAC standard score below 70, whereas only 40% of these same participants had a GAC standard score below 70 on the ABAS-3. Although the DSM-5 does not specify a specific adaptive skill scale cut-score for the diagnosis of ID, it is possible that higher scores on the ABAS-3 could still influence clinical judgement and diagnostic decision making and lead to an overestimation of adaptive skill improvement over time.

It remains unclear what accounts for the increased ABAS-3 scores in our clinical samples. The authors of the ABAS-3 do not elaborate upon potential differences between scores on the ABAS-2 and ABAS-3 Parent Forms identified in their equivalence study. They do, however, elaborate upon a large difference in scaled score in the Self-Care skill area ($d = -1.22$) on the Parent/Primary Caregiver Form (ages 0–5). The authors of the ABAS-3 suggest that this lack of concordance is unlikely to be due to revisions in item content, but instead may be due to reduced caregiver expectations (Harrison & Oakland, 2015). To account for the increase in standard score between versions of the Parent/Primary Caregiver Form (Ages 0–5), the ABAS-3 authors suggest that at the time of the ABAS-2 standardisation, ‘parents generally saw their youngest children as more independent and more able to care for themselves than did parents during the much more recent period of ABAS-3 standardization.’ Although this explanation may be possible, it is our view that this type of sociological explanation is hard to reconcile with the findings of Farmer *et al.* (2020), as they identified standardised score *decreases* when comparing the Vineland-2 and Vineland-3 interview forms.

As an alternative explanation, we propose that the scores on the ABAS-2 in our sample may have been artificially depressed in the final years in which the ABAS-2 was used. We propose this as many of the ABAS-2 items anecdotally appeared outdated by 2014, including items referencing tasks or materials that were no longer commonly used or available. For instance, ABAS-2 items such as ‘Finds and uses a pay phone’ or ‘Reads classified ads for purchases and services’ had both become increasingly less relevant for children in 2014. We theorise that assessing more contemporary children using these types of outdated items may have resulted in lower ratings of adaptive functioning, particularly when these contemporary children were compared with children in the normative sample for whom these items were more functionally and culturally relevant. In contrast, we theorise that these same contemporary children may present as more functional when assessed using more relevant item content from an updated measure like the ABAS-3. One of the potential risks of extended use of adaptive skill inventories is the possibility that items may no longer be relevant to the functioning of the individual being assessed. Of note, there were no content changes between the original version of the ABAS (Harrison & Oakland, 2000) and the ABAS-2 (Harrison & Oakland, 2003), meaning that the item content of the ABAS-2 had not been updated for nearly 15 years prior to the publication of the ABAS-3 (Harrison and Oakland 2015). These findings highlight the importance of regularly updating content on adaptive skill inventories to ensure they remain relevant for the population to which they are being administered.

In addition to changes in format or version, there are other potential explanations for the differences between ABAS versions. It is possible that the larger differences found between ABAS versions in our sample compared to the differences noted in the ABAS-3 manual’s equivalence studies are due to sample characteristics. Specifically, the overall levels of adaptive functioning in our samples were below average as compared with the broader range of adaptive ability represented in the ABAS equivalence studies. Indeed, similar to the present study, recent research on concordance of the Vineland-2 and Vineland-3 found greater score differences between versions in individuals with lower levels of adaptive functioning (Farmer *et al.* 2020).

It is also important to note that it is possible that the differences in ABAS ratings between versions are due to true improvements in adaptive functioning in our samples. However, given that ABAS-3 ratings were higher than ABAS-2 ratings when examining both cross-sectionally in a between-group sample and longitudinally in a within-group sample, it is also possible that these differences are a product of the measure itself. Furthermore, several studies suggest that age-corrected adaptive behaviour ratings tend to *decline* over time in individuals with developmental disabilities such as Fragile X (Klaiman *et al.* 2014), Williams syndrome (Fisher *et al.* 2016) and other genetic conditions associated with ID (Fisch *et al.* 2012). This is not necessarily because these individuals are experiencing a functional decline, but their adaptive skill development is not increasing at the same rate as typically developing peers. Therefore, it is all the more surprising that we found an increase in ABAS scores over time, which was especially pronounced in individuals with a VCI at or below 75, rather than a decrease as might be expected in a clinically referred sample. Also important to consider is the possibility that measurement error is a driver of these observed score changes.

As noted earlier, in contrast to the findings presented in the current paper, Farmer *et al.* (2020) found that caregiver ratings were typically higher on the Vineland-2 than on the Vineland-3. The authors note that it is possible that substantial changes made to the scoring procedures between Vineland versions may have contributed to these differences.

Specifically, new items were added at earlier/easier skill levels, ratings of motor functioning are no longer factored into overall adaptive composites for children 7 years or younger, and there were changes in ratings for behaviours that are inconsistently or incompletely demonstrated without prompting (i.e. behaviours that may have earned a score of '1' on the Vineland-2 may earn a score of '0' on the Vineland-3; Farmer *et al.* 2020; Sparrow *et al.* 2016). Importantly, however, there were no changes made to the scoring procedures between ABAS versions, only changes to item content, so this is not a viable explanation for our ABAS findings.

It is also possible that the subjects in study 2 had received intervention services between administration of the ABAS at time 1 and time 2, resulting in actual improvements in adaptive functioning. A strength of this research is the use of both cross-sectional, between-subjects and longitudinal, within-subjects designs, which allowed us to better control for the potentially confounding effects of participant/cohort characteristics. While the within-groups design used for study 2 was vulnerable to potential intervention related confounds, it is unlikely that intervention-related effects would account for the ABAS-2/ABAS-3 score discrepancies seen in the between-subjects design used in study 1. As such, it remains our view that the mean score differences observed between ABAS versions in studies 1 and 2 are a result of a lack of concordance between versions.

Our results indicate that a caregiver's ratings on the ABAS-3 will likely be higher than their ratings on the ABAS-2, which has important clinical implications. Because adaptive functioning is one of the key criteria in the diagnosis of ID, discrepancies between assessments can potentially raise doubt as to whether the individual still meets diagnostic criteria for ID. This in turn can have substantial impact on treatment decisions (e.g.

decisions made as a result of progress monitoring), as well as service eligibility and provision across the lifespan. We recommend that clinicians be cautious in their interpretation of score increases when changing from the ABAS-2 to the ABAS-3 in an individual's serial assessment, as these increases may be a testing artefact and may not be reflective of actual clinical or functional improvement. These findings also have important implications for research utilising different versions of the ABAS, including treatment studies, and suggests that form version may be important to account for in research design and analyses.

Limitations

This research is limited by the fact that it utilised convenience samples of clinically referred children who demonstrated overall adaptive functioning skills that were lower than the sample of children in the original ABAS validation studies. However, our samples reflect the typical population of children who are likely to need and undergo adaptive behaviour assessment. That being said, we cannot reasonably expect these findings to generalise to a sample of non-clinically referred children. Furthermore, the use of a convenience sample precludes more in-depth assessment of any meaningful differences between participants in study 1 (with only one ABAS) versus study 2 (with both ABAS versions), as administration of these forms was at the discretion of individual clinicians.

Another limitation of this study was that we did not counterbalance the administration of forms, as was done in the original ABAS-2 and ABAS-3 equivalence studies conducted by the authors. However, these forms were administered as part of clinical care and mirror real-world assessment practices. Additionally, participants were only included in this study if they had complete data for the variables of interest; as such, it was not possible to assess for any meaningful differences between individuals who had complete versus incomplete questionnaire data.

Conclusion

The ABAS-3 is considered a state-of-the-art tool for assessing adaptive functioning across domains, and this 2015 revision of the ABAS instrument represents an important improvement over the ABAS-2 content. However, we would advise clinicians who complete serial assessments of individuals for whom both ABAS-2 and ABAS-3 scores are available to consider the impact of incongruence between measures along with individual factors when interpreting such findings and providing recommendations. In addition to adding to the adaptive functioning research literature, we are hopeful that the information provided by these studies can be used to inform clinical practice and treatment planning.

Acknowledgments

We would like to thank the children and families at Kennedy Krieger Institute who made this research possible.

Source of Funding

There was no funding source for this research.

References

- American Psychiatric Association (1980) Diagnostic & Statistical Manual of Mental Disorders, 3rd edn. American Psychiatric Association, Washington, D.C.
- Farmer C, Adedipe D, Bal VH, Chlebowski C & Thurm A (2020) Concordance of the Vineland Adaptive Behavior Scales, second and third editions. *Journal of Intellectual Disability Research* 64, 18–26. [PubMed: 31657503]
- Fisch GS, Carpenter N, Howard-Peebles PN, Holden JA, Tarleton J, Simensen R et al. (2012) Developmental trajectories in syndromes with intellectual disability, with a focus on Wolf–Hirschhorn and its cognitive–behavioral profile. *American Journal on Intellectual and Developmental Disabilities* 117, 167–79. [PubMed: 22515830]
- Fisher MH, Lense MD & Dykens EM (2016) Longitudinal trajectories of intellectual and adaptive functioning in adolescents and adults with Williams syndrome. *Journal of Intellectual Disability Research* 60, 920–32. [PubMed: 27273269]
- Harrison P, & Oakland T (2000) Adaptive Behavior Assessment System. Harcourt Assessment, San Antonio, TX.
- Harrison P, & Oakland T (2003) Adaptive Behavior Assessment System. 2nd edn. Harcourt Assessment, San Antonio, TX.
- Harrison P & Oakland T (2015) Adaptive Behavior Assessment System (ABAS-3). The Psychological Corporation, San Antonio, TX.
- Klaiman C, Quintin E, Jo B, Lightbody AA, Hazlett HC, Piven J et al. (2014) Longitudinal profiles of adaptive behavior in Fragile X syndrome. *Pediatrics* 134, 315–24. [PubMed: 25070318]
- Schalock RL (2011) The evolving understanding of the construct of intellectual disability. *Journal of Intellectual and Developmental Disability* 36, 227–37.
- Sparrow SS, Balla DA, & Cicchetti DV (1984) Vineland Adaptive Behavior Scales. American Guidance Service, Circle Pines, MN.
- Sparrow SS, Cicchetti DV & Balla DA (2005) Vineland Adaptive Behavior Scales, Second Edition (Vineland-II). Pearson, San Antonio, TX.
- Sparrow SS, Cicchetti DV & Saulnier CA (2016) Vineland Adaptive Behavior Scales, Third Edition (Vineland-3). Pearson.
- Thompson JR, McGrew KS & Bruininks RH (1999) Pieces of the puzzle: measuring the personal competence and support needs of persons with intellectual disabilities. *Peabody Journal of Education* 77, 23–39.
- Wechsler D (2003) Wechsler Intelligence Scale for Children, Fourth edn. The Psychological Corporation, San Antonio, TX.
- Wechsler D (2014) Wechsler Intelligence Scale for Children, Fifth edn. The Psychological Corporation, San Antonio, TX.

Table 1

Study 1: demographic characteristics ($n = 2327$)

	ABAS-2 cohort ($n = 1036$)	ABAS-3 cohort ($n = 1291$)
Child age (M, SD)	10.24 (3.44)	10.51 (3.70)
Child ethnicity (%)		
Caucasian	55	56
Black/AA	28	27
Other/unknown	17	17
Gender (%)		
Female	34	33
Male	66	67
WISC-4 (n)	608	5
WISC-5 (n)	428	1286

Note. AA, African-American; ABAS, *Adaptive Behavior Assessment System*; M , mean; SD , standard deviation; WISC, Wechsler Intelligence Scale for Children.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study 1: comparison of ABAS-2 and ABAS-3 standardised scores; mean (*SD*) (*n* = 2327)

Table 2

	ABAS-2 cohort (<i>n</i> = 1036)	ABAS-3 cohort (<i>n</i> = 1291)	<i>t</i> value [†]	<i>P</i> value	Cohen's <i>d</i>	95% CI [lower; upper]
WISC-4/5 VCI ABAS composites	93.67 (16.95)	93.08 (17.42)	0.82	0.41	0.03	[-0.82, 2.00]
GAC (SS)	77.65 (16.46)	83.02 (12.21)	9.04	<0.001	0.37	[4.21, 6.54]
Conceptual (SS)	79.92 (14.71)	82.61 (12.27)	4.81	<0.001	0.20	[1.60, 3.79]
Social (SS)	85.21 (16.89)	85.88 (13.68)	1.05	0.29	0.04	[0.58, 1.91]
Practical (SS)	77.90 (18.16)	85.12 (12.71)	11.25	<0.001	0.47	[5.96, 8.47]

Note. ABAS, *Adaptive Behavior Assessment System*; CI, confidence interval; *SD*, standard deviation; SS, standard score; VCI, Verbal Comprehension Index; WISC, Wechsler Intelligence Scales for Children.

[†]Degrees of freedom = 2325 for all *t* tests.

Table 3

Study 1: Comparison of ABAS-2 and ABAS-3 Scaled Scores; mean (*SD*)

ABAS skill areas	ABAS-2 cohort (<i>n</i> = 1036)	ABAS-3 cohort (<i>n</i> = 1291)	<i>t</i> value [†]	<i>P</i> value	Cohen's <i>d</i>	95% CI [lower, upper]
Communication	6.99 (3.29)	7.46 (2.65)	3.75	<0.001	0.16	[0.22, 0.70]
Community use	5.99 (3.39)	7.85 (2.73)	14.68	<0.001	0.60	[1.61, 2.11]
Functional academics	6.68 (2.94)	7.30 (2.82)	5.24	<0.001	0.22	[0.39, 0.86]
Home living	4.91 (3.67)	7.07(2.45)	16.92	<0.001	0.69	[1.90, 2.40]
Health and safety	7.65 (3.29)	7.72 (2.77)	0.58	0.57	0.02	[0.17, 0.32]
Leisure	7.68 (3.24)	7.40 (2.84)	2.20	0.03	0.09	[-0.53, 0.03]
Self-care	6.39 (3.34)	7.95 (2.77)	12.28	<0.001	0.51	[1.31, 1.80]
Self-direction	4.93 (3.44)	6.41 (2.27)	12.41	<0.001	0.51	[1.24, 1.71]
Social	6.38 (3.93)	7.27 (2.81)	6.38	<0.001	0.26	[0.62, 1.17]

Note. ABAS, *Adaptive Behavior Assessment System*; CI, confidence interval; *SD*, standard deviation.

[†]Degrees of freedom = 2325 for all *t* tests.

Table 4

Study 2: demographic characteristics ($n = 572$)

Child age, time 1 (<i>M, SD</i>)	9.65 (2.80)
Child age, time 2 (<i>M, SD</i>)	13.33 (2.95)
Child ethnicity (%)	
Caucasian	57
Black/AA	26
Other/unknown	17
Gender (%)	
Female	32
Male	68

Note. AA, African-American; *M*, mean; *SD*, standard deviation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Study 2: comparison of ABAS-2 and ABAS-3 standardised scores; mean (*SD*) (*n* = 572)

	ABAS-2	ABAS-3	<i>t</i> value [‡]	<i>P</i> value	Cohen's <i>d</i>	95% CI [lower, upper]	CCC, 95% CI [lower, upper]
GAC (SS)	72.87 (17.20)	79.56 (13.37)	12.75	<0.001	0.37	[5.66, 7.72]	0.61 [0.57, 0.66]
Conceptual (SS)	76.57 (15.66)	79.84 (12.77)	7.03	<0.001	0.20	[2.36, 4.18]	0.68 [0.64, 0.72]
Social (SS)	81.68 (17.22)	83.37 (14.59)	2.98	0.003	0.11	[0.57, 2.80]	0.64 [0.59, 0.68]
Practical (SS)	72.29 (19.10)	80.89 (13.92)	9.82	<0.001	0.47	[7.38, 9.82]	0.54 [0.49, 0.58]

Note. ABAS, *Adaptive Behavior Assessment System*; GAC, general adaptive composite; *SD*, standard deviation; SS, standard score.

[‡]Degrees of freedom = 571 for all *t* tests.

Table 6
 Study 2: comparison of ABAS-2 and ABAS-3 scaled scores; mean (*SD*) (*n* = 566)

	ABAS-2	ABAS-3	<i>t</i> value [†]	<i>P</i> value	Cohen's <i>d</i>	95% CI [lower, upper]
ABAS skill areas						
Communication	6.28 (3.43)	6.94 (2.85)	6.26	<0.001	0.21	[0.45, 0.87]
Community use	5.09 (3.50)	6.73 (3.04)	12.79	<0.001	0.50	[1.39, 1.89]
Functional academics	6.10 (3.28)	6.69 (2.82)	5.65	<0.001	0.19	[0.39, 0.80]
Home living	4.32 (3.39)	6.56 (2.42)	17.02	<0.001	0.76	[1.98, 2.50]
Health and safety	6.67 (3.55)	7.09 (3.06)	0.69	0.001	0.13	[0.16, 0.69]
Leisure	6.91 (3.35)	6.64 (3.21)	-2.22	0.027	0.08	[-0.51, -0.03]
Self-care	5.50 (3.15)	7.39 (2.96)	16.66	<0.001	0.62	[1.67, 2.12]
Self-direction	4.46 (3.33)	5.98 (2.49)	3.84	<0.001	0.52	[0.23, 0.72]
Social	5.84 (3.93)	6.65 (3.05)	6.17	<0.001	0.23	[0.56, 1.08]

Note. ABAS, *Adaptive Behavior Assessment System*; *SD*, standard deviation.

[†]Degrees of freedom = 565 for all *t* tests.

Table 7

ABAS-2 versus ABAS-3 mean scaled/standard score differences and effect sizes for studies 1 and 2

	Study 1 mean difference (<i>n</i> = 2327)	Study 1 effect sizes (<i>d</i>)	Study 2 mean difference (<i>n</i> = 572)	Study 2 effect sizes (<i>d</i>)
GAC	5.37	0.37	6.69	0.43
Conceptual	2.69	0.20	3.27	0.23
Communication	0.47	0.16	0.66	0.21
Functional academics	0.62	0.22	0.59	0.19
Self-direction	1.48	0.51	1.52	0.52
Social	0.67	0.04	1.69	0.11
Leisure	-0.28	0.09	-0.27	0.08
Social	0.89	0.26	0.81	0.23
Practical	7.22	0.47	8.60	0.51
Community use	1.86	0.60	1.64	0.50
Home living	2.16	0.69	2.24	0.76
Health and safety	0.06	0.02	0.42	0.13
Self-care	1.56	0.51	1.89	0.62

Note. Composites are presented as standard scores ($M = 100, SD = 15$), subscales are presented as scaled scores ($M = 10, SD = 3$); mean difference = ABAS-3 — ABAS-2.

ABAS, *Adaptive Behavior Assessment System*.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 8

Study 1: comparison of ABAS-2 and ABAS-3 standard scores by VCI; mean (*SD*); *n* = 2327

	ABAS-2 cohort VCI (<i>n</i> = 153)	75	ABAS-3 cohort VCI (<i>n</i> = 169)	75	<i>t</i> value [†]	<i>P</i> value	Diff.	ABAS-2 cohort VCI (<i>n</i> = 883)	76	ABAS-3 cohort VCI (<i>n</i> = 1122)	76	<i>t</i> value [‡]	<i>P</i> value	Diff.
WISC-4 (%)	62.75		1.18		–	–	–	57.98		0.27		–	–	–
WISC-5 (%)	37.25		98.82		–	–	–	42.02		99.73		–	–	–
Composites														
GAC	68.88 (16.21)		75.64 (10.50)		4.48	<0.001	6.76	79.16 (16.03)		84.14 (12.07)		7.92	<0.001	4.98
Conceptual	69.60 (13.54)		73.13 (9.96)		2.68	0.008	3.53	81.70 (14.17)		84.04 (11.94)		3.99	<0.001	2.34
Social	78.07 (15.27)		80.02 (12.54)		1.26	0.208	1.95	86.45 (16.86)		86.76 (13.63)		0.46	0.65	0.31
Practical	70.01 (20.52)		79.25 (11.50)		5.04	<0.001	9.24	79.27 (17.37)		86.00 (12.65)		10.04	<0.001	6.73

ABAS, *Adaptive Behavior Assessment System*; Diff., difference score (ABAS-3 — ABAS-2); GAC, general adaptive composite; *SD*, standard deviation; VCI, Verbal Comprehension Index; WISC, Wechsler Intelligence Scales for Children.

[†]Degrees of freedom = 320.

[‡]Degrees of freedom = 2003.

Table 9

Study 1: comparison of ABAS-2 and ABAS-3 scaled scores by VCI; mean (*SD*)

ABAS subscales	ABAS-2 VCI 153	75 (<i>n</i> = 153)	ABAS-3 VCI 169	75 (<i>n</i> = 169)	<i>t</i> value [†]	<i>P</i> value	Diff.	ABAS-2 VCI 883	76 (<i>n</i> = 883)	ABAS-3 VCI 1122	76 (<i>n</i> = 1122)	<i>t</i> value [‡]	<i>P</i> value	Diff.
Communication	5.04 (2.93)		5.59 (2.36)		1.87	0.062	0.55	7.33 (3.23)		7.74 (2.58)		3.11	0.002	0.32
Community use	4.85 (3.30)		6.53 (2.39)		5.27	<0.001	1.68	6.19 (3.37)		8.05 (2.73)		13.70	<0.001	0.62
Functional academics	4.44 (2.53)		4.97 (2.56)		1.93	0.055	0.53	7.06 (2.84)		7.66 (2.71)		4.76	<0.001	0.22
Home living	4.59 (3.89)		6.58 (2.41)		5.56	<0.001	1.99	4.97 (3.63)		7.14 (2.44)		15.96	<0.001	0.89
Health and safety	6.17 (3.46)		6.44 (2.49)		0.80	0.42	0.27	7.90 (3.19)		7.91 (2.76)		0.07	0.944	0.13
Leisure	6.29 (2.99)		6.15 (2.59)		-0.47	0.64	-0.15	7.92 (3.22)		7.59 (2.83)		-2.43	0.015	0.02
Self-care	5.27 (3.55)		7.01 (2.95)		4.78	<0.001	1.73	6.58 (3.26)		8.09 (2.72)		11.25	<0.001	0.71
Self-direction	3.65 (3.20)		5.55 (2.01)		6.43	<0.001	1.90	5.15 (3.43)		6.54 (2.28)		10.80	<0.001	0.54
Social	4.98 (3.60)		6.21 (2.70)		3.48	0.001	1.23	6.62 (3.93)		7.43 (2.79)		5.39	<0.001	0.32

ABAS, *Adaptive Behavior Assessment System*; VCI, Verbal Comprehension Index; *SD*, standard deviation; Diff., difference score (ABAS-3 — ABAS-2).

[†]Degrees of freedom = 320.

[‡]Degrees of freedom = 2003.

Table 10
 Study 2: comparison of ABAS-2 and ABAS-3 standard scores by VCI; mean (*SD*)

	ABAS-2 VCI 75 (n = 58)	ABAS-3 VCI 75 (n = 58)	t value [†]	P value	Diff.	CCC, 95% CI [lower, upper]	ABAS-2 VCI 76 (n = 300)	ABAS-3 VCI 76 (n = 300)	t value [‡]	P value	Diff.	CCC, 95% CI [lower, upper]
Composites												
GAC	65.12 (15.46)	73.95 (12.71)	5.02	<0.001	8.83	0.46 [0.29, 0.63]	76.12 (14.73)	82.69 (11.42)	9.24	<0.001	6.57	0.50 [0.43, 0.56]
Conceptual	67.00 (13.52)	72.21 (11.49)	3.17	0.002	5.21	0.46 [0.28, 0.65]	79.88 (13.64)	82.99 (10.86)	4.74	<0.001	3.11	0.56 [0.49, 0.63]
Social	77.26 (17.84)	78.55 (13.51)	0.72	0.474	1.29	0.63 [0.48, 0.78]	84.53 (15.74)	86.64 (13.77)	2.72	0.007	2.11	0.58 [0.51, 0.66]
Practical	66.02 (19.03)	77.33 (13.95)	5.65	<0.001	11.31	0.47 [0.31, 0.63]	75.31 (16.26)	83.66 (11.99)	10.02	<0.001	8.35	0.42 [0.34, 0.50]

ABAS, *Adaptive Behavior Assessment System*; CCC, concordance correlation coefficient; *SD*, standard deviation; VCI, Verbal Comprehension Index.

[†]Degrees of freedom = 57.

[‡]Degrees of freedom = 299.

Table 11

Study 2: comparison of ABAS-2 and ABAS-3 scaled scores by VCI; mean (*SD*)

	ABAS-2 VCI 75 (<i>n</i> = 57)	ABAS-3 VCI 75 (<i>n</i> = 57)	<i>t</i> value [†]	<i>P</i> value	Diff.	ABAS-2 VCI 76 (<i>n</i> = 297)	ABAS-3 VCI 76 (<i>n</i> = 297)	<i>t</i> value [‡]	<i>P</i> value	Diff.
ABAS subscales										
Communication	4.04 (3.01)	5.26 (2.74)	3.30	0.002	1.22	6.86 (3.19)	7.77 (2.43)	5.91	<0.001	0.91
Community use	3.68 (3.20)	5.79 (2.46)	4.80	<0.001	2.11	5.48 (3.21)	7.30 (2.68)	10.59	<0.001	1.82
Functional academics	3.95 (2.68)	4.79 (2.19)	2.52	0.01	0.84	6.87 (3.06)	7.48 (2.40)	4.01	<0.001	0.61
Home living	4.19 (3.32)	6.28 (2.70)	5.32	<0.001	2.09	4.37 (3.09)	6.77 (2.41)	13.06	<0.001	2.40
Health and safety	5.18 (3.51)	5.95 (3.28)	1.88	0.07	0.77	7.34 (3.13)	7.73 (2.71)	2.08	0.04	0.39
Leisure	5.98 (3.46)	5.86 (2.96)	-0.31	0.76	-0.12	7.53 (3.06)	7.48 (2.86)	-0.31	0.76	-0.05
Self-care	4.89 (3.16)	6.74 (3.14)	5.40	<0.001	1.85	5.93 (2.79)	7.87 (2.66)	12.55	<0.001	1.94
Self-direction	3.58 (3.11)	5.46 (2.42)	4.49	<0.001	1.88	4.78 (3.15)	6.26 (2.27)	8.55	<0.001	1.48
Social	4.79 (4.20)	5.77 (3.01)	2.33	0.02	0.98	6.31 (3.65)	7.35 (2.76)	5.75	<0.001	1.04

ABAS, *Adaptive Behavior Assessment System*; Diff., difference score (ABAS-3 — ABAS-2); *SD*, standard deviation; VCI, Verbal Comprehension Index.

[†]Degrees of freedom = 56.

[‡]Degrees of freedom = 296.s