American Journal on Intellectual and Developmental Disabilities

Vineland-3 Measurement Non-Invariance in Children with and without Intellectual and Developmental Disabilities

--Manuscript Draft--

Manuscript Number:	AJIDD-D-21-00075R3
Article Type:	Research Report
Keywords:	Adaptive Behavior; Vineland-3; intellectual disabilities; developmental disabilities
Corresponding Author:	Maryellen Brunson McClain Indiana University Bloomington Bloomington, IN UNITED STATES
First Author:	Maryellen Brunson McClain
Order of Authors:	Maryellen Brunson McClain
	Sarah E Schwartz
	Jac'lyn Bera
	Ryan Farmer
	Sarfaraz Serang
	Bryn Harris
	Megan E Golson
Manuscript Region of Origin:	UNITED STATES
Abstract:	Measurement of adaptive skills is important in the diagnosis, intervention planning, and progress monitoring of children with intellectual and developmental disabilities (I/DD). Thus, ensuring accurate measurement, including measurement invariance, across children with and without I/DD is critical. In this study, we evaluate the measurement invariance using multigroup confirmatory factor analysis (MG-CFA) of the Vineland-3 Comprehensive Interview (CIF) across children ages 6-21 with and without I/DD (N =1,192) using archival data. Results showed that the Vineland-3 CIF exhibits configural invariance but may show metric non-invariance in children with and without I/DD. It is possible that the Vineland-3 CIF may not be measuring adaptive behavior comparably across these populations. Suggestions for using the Vineland-3 CIF in this population are provided and future research and measure development needs are discussed.

Vineland-3 Measurement Non-Invariance in Children with and without Intellectual and Developmental Disabilities

Abstract

Measurement of adaptive skills is important in the diagnosis, intervention planning, and progress monitoring of children with intellectual and developmental disabilities (I/DD). Thus, ensuring accurate measurement, including measurement invariance, across children with and without I/DD is critical. In this study, we evaluate the measurement invariance using multigroup confirmatory factor analysis (MG-CFA) of the Vineland-3 Comprehensive Interview (CIF) across children ages 6-21 years with and without I/DD (N = 1,192) using archival data. Results showed that the Vineland-3 CIF exhibits configural invariance but may show metric non-invariance in children with and without I/DD. It is possible that the Vineland-3 CIF may not be measuring adaptive behavior comparably across these populations. Suggestions for using the Vineland-3 CIF in this population are provided and future research and measure development needs are discussed.

Keywords. Adaptive behavior, Vineland-3, intellectual disabilities, developmental disabilities, measurement invariance

Measurement Non-Invariance of the Vineland-3 in Children with and without Intellectual and Developmental Disabilities

Adaptive behavior skills are necessary for independent daily living including engaging in self-care (e.g., toileting, dressing), navigating successfully in the community, communicating effectively with others, developing social relationships, and finding and maintaining employment (Oakland & Harrison, 2008). These behaviors, conceptualized into practical, conceptual, and social domains, are used by individuals in everyday life (Schalock et al., 2021; Tassé et al., 2012). Most notably, adaptive behavior deficits are seen in children who have intellectual disability (ID) and are core diagnostic criteria for the disorder (APA, 2013; Balboni et al, 2020; Mouga et al, 2015). Specifically, for a diagnosis of ID, impairments in cognitive abilities and adaptive functioning across all three domains are required (APA, 2013). However, many children with other disabilities are at-risk for adaptive behavior deficits, including those who have autism spectrum disorder, attention-deficit/hyperactivity disorder, learning disabilities, and developmental delays (e.g., Clark et al., 2002; Ditterline et al., 2008; Frost et al., 2017; Milne et al., 2013; Vig & Jedrysek, 1995).

For any children who have disabilities and adaptive behavior deficits, the development of these skills is important as they are associated with several positive outcomes in these populations including post-secondary employment, higher education, executive functioning, and social skills (Dell'Armo & Tassé, 2019; Gligorović & Durović, 2014; Pugliese et al, 2016; Tsermentseli et al, 2018). Importantly, research has shown that children with adaptive behavior deficits may benefit from targeted educational programming and interventions (Chadwick et al., 2005; Heyman & Hauser-Cram, 2019; Woodman et al., 2018). Considering this, adaptive behavior assessment is critical for the diagnosis of ID but may also aid in the identification, development, and outcome monitoring of relevant educational and intervention efforts for children with a range of developmental disabilities and adaptive skill deficits.

Adaptive Behavior Measurement

Psychologists and other professionals often use standardized rating scales and interviews to quantify a student's skills within each of the primary domains. In a recent survey of school psychologists, Benson et al (2018) found the Adaptive Behavior Assessment System, Third Edition (ABAS-3; Harrison & Oakland, 2015) and the Vineland Adaptive Behavior Scales, Third Edition (Vineland-3; Sparrow et al., 2016) were the most frequently used adaptive behavior assessment instruments. The Vineland series of adaptive behavior tests are particularly popular across professions (e.g., Peterson et al., 2014). The Vineland-3 is divided into Comprehensive- or Domain-type forms. According to promotional material for the Vineland-3, comprehensive forms are longer and may yield additional information useful for developing interventions whereas the Domain forms are briefer forms of the Vineland-3 forms. While both the Comprehensive and Domain-type forms provide information at the Adaptive Behavior Composite, Domain, and item levels, the Comprehensive form also allows review at the subdomain level. The Vineland-3 is divided into the Comprehensive Interview Form (CIF) and the Parent/Caregiver Form, which are appropriate for ages 0 to 90+ years, and the Teacher Form, which is appropriate for ages 3 to 21.

The authors of the Vineland-3 CIF (Sparrow et al., 2016) suggest a factor structure like other current and past forms of the Vineland. That is, they propose a higher-order model in which subdomains group within domains, and a positive manifold of correlations across all items on the Vineland-3 suggest an underlying general adaptive factor. This is consistent with the scoring and interpretative framework the authors recommend in which three general domains

(Communication, Daily Living Skills, and Socialization) and a general composite (Adaptive Behavior Composite) can be calculated and interpreted. In addition, the Vineland-3 produces Motor Skills and Maladaptive behavior domain scores, though these are not typically included in the calculation of the general adaptive domain. Despite consistent clinical use for many decades and past support, Sparrow and colleagues (2016) do not report factor analytic evidence for their measurement model or the interpretive framework of the Vineland-3. Despite the assertion that "...support for the Vineland-3 domain/subdomain structure rests more on the value it has demonstrated in research and practice over the years than on these intercorrelation results." (Sparrow et al., 2016, p. 54), International Test Commission (2017) recommendations are to provide evidence supporting the current norms, reliability of scale scores, and validity—including structural validity—of adapted tests (Confirmation Guideline 3).

Recently, two research teams have evaluated the measurement model underlying the Vineland-3 (Farmer et al., 2020; Pandolfi & Magyar, 2021). Farmer and colleagues (2020) conducted exploratory and confirmatory factor analysis of the Vineland-3 CIF by age-group using data from Pearson's standardization sample. In general, the higher-order model proposed by the authors fit adequately for most, but not all, age ranges evaluated. In contrast, evidence from factor extraction approaches suggested that the Vineland-3 CIF was better described as a unidimensional measure. Pandolfi and Magyar (2021) also completed an exploratory factor analysis of the Vineland-3 CIF based on correlation matrices available in the Vineland-3 manual (Sparrow et al., 2016) and found similar exploratory factor analysis results to those described by Farmer et al. (2020). That is, the data supported that the Vineland-3 CIF was largely unidimensional. Despite these findings, the prevailing theory regarding adaptive behavior is that the three primary domains are interrelated but may vary distinctly from one another. This is manifested in recent AAIDD guidelines that suggest deficits need only present with a deficit in one area of adaptive behavior (Schalock et al., 2021).

Measurement Invariance

Measurement invariance (MI) of a test (Meredith, 1993) reflects the notion that the test functions the same way across groups (e.g., race/ethnicity, dis/ability). It is required to ensure that two individuals with the same levels of latent ability the test attempts to assess also have the same observed scores on that test (Horn & McArdle, 1992). Consider two individuals with a “true” latent IQ of 105. If a test systematically produces observed IQ scores of 108 for one individual and 103 for the other, the test is biased, and the observed scores cannot be meaningfully compared side by side. To ensure that Vineland-3 CIF scores are comparable across children with and without I/DD, measurement invariance must be established.

Current Study

Clinicians and researchers commonly use the Vineland-3 CIF to measure adaptive behavior skills in children with and without I/DD (Benson et al., 2018). These usages assume that the test functions the same way across both groups, a critical assumption that has yet to be empirically demonstrated (MacLean, 2011). Lack of MI has important implications for both parties. For clinicians, a lack of MI would suggest that conclusions drawn about adaptive functioning based on the Vineland-3 CIF may be invalid and may adversely impact diagnostic decision-making. For researchers, lack of MI would imply that scores from both groups cannot be compared, either directly or in statistical models, as they may not represent the same constructs in the same way in each. For these reasons, we seek to evaluate MI the Vineland-3 CIF.

Method

Sample

We used archival data including the norm-referenced subdomain scores for the Vineland-3 CIF. The dataset was provided by NCS Pearson in an Excel spreadsheet. Data from this spreadsheet was imported into R version 4.0.3. The entire sample included 2,749 individuals aged 0-96 years with and without disabilities. For the purposes of our study, we only included participants (N = 1,192; male = 50.1%, female = 49.9%) who were parents or caregivers of children aged 6-21 years who were Black or African American (n = 193), Latinx (n = 305), or White (n = 694). Participants from other racial/ethnic groups were not included due to small sample sizes. This form population was selected due to frequent implementation in the school setting, the alignment with school-aged children (e.g., Benson et al., 2018). Within our sample, 110 had a diagnosis of I/DD, which included developmental disability (n = 11), mild intellectual disability (n = 57), and moderate intellectual disability (n = 42).

Measures

The Vineland-3 CIF (Sparrow et al., 2016) was designed for ages 0 to 90+ years and uses 502 items to sample across five domains, including Communication, Daily Living Skills, Socialization, Motor Skills, and Maladaptive Behavior. As this study is focused on the MI of adaptive behavior, we excluded the Maladaptive Behavior and Motor Skills composites as they are not part of contemporary definitions of adaptive behavior (see Schalock et al., 2021). Each of the 9 subdomains form a single subtest, which contribute to the Communication, Daily Living Skills, or Socialization domains. Details regarding the author proposed measurement model are provided in Table 1.

Completion of a subdomains from the Vineland-3 CIF yields a raw score, which is converted and ultimately reported as a *v*-scale score (M = 15, SD = 3) using either the computer scoring software or the appendices in the Vineland-3 manual (Sparrow et al., 2016). Overall, the reliability of the subdomains for ages 3 to 97 years are adequate or better, with median internal consistency reliability, estimated via Cronbach's alpha, all above .90. Similarly, range restriction corrected test-retest reliability coefficients collected between 12 and 35 days apart ranged between .64 and .92. Within each domain, the subdomain *v*-scale scores are summed to create a sum of scores for the domain, which can then be converted into a standard score (M = 100, SD = 15) using either the computer scoring software or manually the appendices in the Vineland-3 manual. Additional psychometric data for composite scores are available in Sparrow et al. (2016) but are excluded here as subdomain scores were used for analyses.

Procedures

The archival data used in this study were collected as part of Pearson's standardization and norming process for the Vineland-3 CIF. Clinicians who conducted the CIF were those who had administered similar scales for Pearson Publishing in the past and included clinical and school psychologists, educational diagnosticians, and speech language pathologists (Sparrow et al., 2016). Prior to collecting data, all clinicians completed a web-based training provided by Pearson representatives. Parents and caregivers were recruited either by Pearson Publishing or by interviewers and in most cases were mothers. All data were collected as per standardized instructions provided in the Vineland-3 CIF manual. Much of the data were collected online via the Q Global digital platform while a smaller subset of data was collected by paper and pencil.

Because data are owned by Pearson, they are not publicly available. This study was not pre-registered.

Data Analysis Plan

We first conducted exploratory data analyses including sample demographics. Next, we ran pairwise correlations and internal consistency between and for all parcels, respectively. Parcels are those proposed by the publisher and were created using the procedures outlined by the publisher, wherein each subdomain raw score was created by summation of awarded points. These points were then translated to a v scale score with a mean of 15 and a standard deviation of 3. This procedure was completed by the publisher and supplied to the researchers. Further details are available in the Vineland-3 technical manual (see Sparrow et al., 2016). To establish the configural model, three competing confirmatory factor analyses (CFA) were fit to the full sample. The first CFA was an unequally weighted 3-factor model, which aligns with the theoretical model of the Vineland-3 CIF (Sparrow et al., 2016). The second CFA was an equally weighted 3-factor model. The third CFA was a single-factor unidimensional model. Three CFAs were evaluated to determine the best fitting model due to the proposed theoretical structure and results supporting both unidimensional and multidimensional structural support (Farmer et al., 2020). We used several indices to evaluate model fit including chi-square tests of model fit, comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). We then conducted a multigroup CFA (MG-CFA; Jöreskog, 1971) using the 3-factor unequally weighted model to assess MI across children with and without I/DD. The equally weighted model—which is reflective of the Vineland-3 use in practice—was assessed for MI and yielded almost identical results (see supplemental material[1]). All CFAs

[1] Supplemental material is housed on Open Science Framework and can be accessed at: https://osf.io/2tpz3/?view_only=1f5376cebb474ac3900d5a98ede05a4a

included the mean structure and used maximum likelihood estimation. Assessment of model comparisons included traditional Likelihood Ratio Tests (LRT) despite likelihood of over sensitivity in large samples such as ours (Limbers et al. 2009). Model fit was determined using Pendergast and colleagues' recommendations (Pendergast et al., 2017) as well as Chen's (2007) suggestion of using the change in CFI ($\Delta \leq -.005$) supplemented by RMSEA ($\Delta \geq .010$) or SRMR ($\Delta \geq .025$). Partial invariance was investigated with a multivariate score test (i.e., Lagrange Multiplier Test) and univariate score tests (Chi-square difference tests). All analyses were conducted in *R 4.0.3* (R Core Team, 2020) using *lavaan* (Roseel, 2012) and are fully detailed in the supplemental material.

Results

Pairwise correlations revealed that all 9 parcels were highly correlated with one another (see supplemental material). Internal consistency for all 9 parcels was strong, $\alpha = .95$. Results yielded that the for the full sample, the first CFA (i.e., 3-factor, unequally weighted) was the best fitting model, $\chi^2(24) = 257.45$, $p < .001$; CFI = .975, RMSEA = .09, $p < .001$, when compared both the second CFA (i.e., 3 factor equally weighted), LRT χ^2 (3) = 888.67, $p < .001$, and the third CFA (i.e., single factor model), LRT χ^2 (3) = 350.94, $p < .001$. See Table 2 for other model fit statistics, as well as details regarding model fits for the MG-CFA analyses. See Figure 1 for parameter estimates of the configural and metric invariance models. In the configural invariance model, we can see from Figure 1 that the relation among factor loadings is inconsistent between children with and without I/DD. Applying Pendergast et al.'s (2017) recommendations, we found empirical support for metric (weak) non-invariance, indicating that the factor loadings differed across children with and without I/DD, $\chi^2(6) = 19.61$, $p = .003$. However, using Chen's (2007) suggested cutoffs, we did not find weak invariance, ΔCFI = -.002, ΔRMSEA = .003, ΔSRMR =

.006. Score tests (Bentler & Chou, 1993) revealed non-invariance in the loadings for two parcels: the Personal subdomain in the Daily Living Skills domain and Coping Skills subdomain in the Socialization domain. Further constraint of the intercepts of the partial metric model failed to demonstrate scalar invariance, $\chi^2(6) = 64.26$, $p < .001$, $\Delta CFI = -.010$, $\Delta RMSEA = .020$, $\Delta SRMR = .014$. See Table 2 for all model fit statistics.

We found this surprising given the vast differences in the loadings, so we further examined the context surrounding Chen's suggestions. We were especially wary of the impact of the ratio of sample sizes per group, as our 10:1 far exceeded the largest 4:1 studied in Chen's (2007) simulations. To evaluate this, we conducted a small Monte Carlo simulation study of our own. We randomly selected 110 children without I/DD to balance the 110 children with I/DD, to see if this made a difference. We then fit the same MG-CFA models to this subset of 220 children to understand how the suggested cutoffs fared (see supplemental material for code and results). We repeated this process 1,000 times to smooth out the uncertainty associated with a single random draw. We found the average ΔCFI from these samples to be -.0057, beyond the cutoff, though the average ΔRMSEA was not (-.0002), although the average ΔSRMR was close (.021). Additionally, if we were to use Chen's rules, 30% of our replications would have pointed toward metric non-invariance, despite the over 80% reduction in our sample size. We believe this is enough evidence to cast doubt on whether Chen's suggested cutoffs as appropriate for our specific dataset. Chen (2007) seemed to anticipate these sorts of concerns, concluding the paper by saying, "However, these criteria should be used with caution, because testing measurement invariance is a very complex issue" (p. 502) and that "the findings are limited to the conditions investigated in the studies reported here" (p. 502). Chen (2007) also goes on to echo the sentiments of Bollen & Long (1983), who say, "The test statistics and fit indices are very beneficial, but they are no replacement for sound judgment and substantive expertise" (p. 8). That said, we believe that the evidence we have presented suggests that it is possible the Vineland-3 CIF measures latent constructs differently for children with and without I/DD, and that their scores may not be comparable.

Discussion

The results of this study indicate that the configural structure of the Vineland-3 CIF is potentially invariant between children ages 6-21 years with and without I/DD. However, to our knowledge we present the first evidence of partial metric non-invariance of the Vineland-3 CIF, indicating that the test may not measure the domains in the same way for children with and without I/DD. It is critical that we point out that the Vineland-3 CIF had *partial* metric non-invariance, meaning that some loadings were the same in both groups, a significant finding unto itself. While these results stem from a small heterogenous sample (*n* = 110), they should give us pause. We know of no other analysis of the Vineland-3 CIF's metric non-invariance; given that the Vineland-3 CIF introduced variations to items and a new normative sample (Sparrow et al., 2016), it is necessary to evaluate its psychometric properties independently of its previous versions (American Education Research Association, American Psychological Association, and National Council on Measurement in Education, 2014; International Testing Commission, 2017). It is possible research comparisons of children with and without I/DD using the Vineland-3 CIF may not be valid as the results of this study demonstrate that the test may not measure adaptive functioning in the same way in each of these populations, leading to inaccurate comparison and understanding of adaptive skills. Mean comparisons of the Vineland-3 CIF across children with and without I/DD may not be valid either, as these require scalar (strong) invariance, whereas we did not even achieve metric (weak) invariance. The results of this study also broadly demonstrate potential challenges using the Vineland-3 CIF as a measurement tool as the factor loadings within the configuration may not be equivalent and perhaps should not be scored in the same way for both clinical and non-clinical populations. Ultimately, researchers may not be measuring adaptive behavior in the same way on the Vineland-3 CIF across children with and without I/DD leading to potentially inaccurate interpretations of the scores and invalid comparison between groups. The same concerns may arise if using the Vineland-3 in clinical contexts when a comparison of standardized scores across individuals with and without I/DD are needed. However, because of the small and heterogenous sample coupled with the importance of replicating studies prior to making such significant claims, the most important conclusion resulting from our finding of potential measurement non-invariance of the Vineland-3 CIF is the imperative need for more research on comprehensively evaluating its psychometric properties and subsequent applications in research and clinical contexts.Moving beyond correlational and factorial data may be critical in understanding how best to use the Vineland-3 CIF in clinical practice. For instance, examining classification accuracy (e.g., sensitivity, specificity, positive likelihood ratio) may reveal that the test scores differentiate well between those with and without I/DD.

As metric invariance is truly fully met, there are several potential paths forward. An adaptation of the Vineland-3 CIF could be more intentional about establishing MI by selecting invariant items. A recalibration of the scoring procedure that considers MI may also be beneficial. Until the potential measurement non-invariance of the Vineland-3 CIF across children with and without I/DD is addressed, professionals may consider relying on the qualitative information provided by the Vineland-3 CIF's items in addition to the quantitative information provided by its scores to assess and compare adaptive behavior in children with and without

I/DD. The benefit of criterion-based measurement is that a person's abilities are compared against a predetermined standard, and not against other children. Although our study illuminates considerations for the use of the Vineland-3 CIF in children with and without I/DD we believe it still plays a valuable role in providing information regarding adaptive functioning skills in this population. Our aim is not to discourage the use of the Vineland-3 CIF in research and clinical practice, but rather encourage clinicians and researchers to place the appropriate weight on conclusions drawn from the test.

Limitations

One limitation of this study is that we lacked sufficient information regarding the original clinicians conducting the Vineland-3 CIF in the dataset. Such information would reflect the backgrounds of the raters (e.g., discipline, clinical experience, race/ethnicity) which would allow for us to hypothesize potential explanations more holistically for our findings and may be important in understanding the results of our study which suggested measurement non-invariance. Another limitation of our study is that the diagnostic categories used in the data set were unclear and not inclusive. Although the clinical sample included parents of children with I/DD, there was not a clear definition of what qualified as DD specific to their dataset. Unfortunately, the publishers did not provide a description or definition of inclusion criteria for DD specifically either in sharing the data or in their published material. Moreover, ID was only inclusive of mild and moderate, which limits the application and generalizability for severe or profound ID populations. Finally, there is limited racial and ethnic representation included in the sample, especially for the clinical group. Considering the complex intersectionality of race/ethnicity and dis/ability and the historic I/DD identification disparities many racially minoritized children and their families experience (e.g., Maenner et al., 2020, Mañgana et al., 2016) generalizability of the Vineland-3 CIF within these more specific populations is crucial.

Future Research

Although our study suggests potential metric non-invariance, future research is certainly needed to further evaluate this claim. We see several directions moving forward. First, we only examined the Vineland-3 CIF, which involves a trained clinician conducting a structured interview with a parent/caregiver and scoring items based on gleaned information. Future research should examine the MI of the Vineland-3 parent and teacher rating forms that are commonly used in I/DD evaluations and research. We also only examined MI of the Vineland-3 CIF for children aged 6-21 years. Future research should determine the MI of the Vineland-3 CIF across the lifespan, as the measure is normed for individuals ages 0-90+ years. Although the Vineland-3 is a commonly used adaptive instrument there are other measures—such as the ABAS-3 (Harrison & Oakland, 2015)—that are also regularly used in clinical and research contexts. Thus, research investigating the MI other adaptive measure may also be beneficial for clinicians and researchers alike.

The most significant limitation of this paper is the small and heterogenous sample on which we base these analyses. It is not clear whether the sample size permits us to adequately rule out that our findings are false positives; post-hoc power is largely redundant with finding a statistically significant effect and simulation data would have been poorly informed by the absence of an agreed-upon smallest effect size of interest. However, we contend that any assumption made by the publisher supporting non-invariance between groups on the Vineland-3 CIF would need to rely on this same sample. Thus, our analysis relies on a sample *at least as good as* those used in alternative analyses performed by the publisher to provide initial support of their assumed non-invariance. We use the terms ‘assumption’ and ‘assumed’ due to the absence of published analyses regarding the non-invariance of the Vineland-3 CIF. Regardless, more research evaluating MI of the Vineland-3 in children with and without I/DD is necessary considering the importance of replication studies and the impact of such findings on how researchers and clinicians use the Vineland-3 with this population. Finally, the purpose of our study was to evaluate MI in the Vineland-3 in children with and without I/DD. Considering our findings revealed not only measurement non-invariance but also mediocre model fit for both children with and without I/DD, we recommend future research explore how to improve model fit for the Vineland-3 in a range of populations as this would have important implications and support to the use of the Vineland-3 in research and practice.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Balboni, G., Rebecchini, G., Elisei, S., & Tassé, M. J. (2020). Factors affecting the relationship between adaptive behavior and challenging behaviors in individuals with intellectual disability and co-occurring disorders. *Research in Developmental Disabilities*, 104. https://doiorg.dist.lib.usu.edu/10.1016/j.ridd.2020.103718

Benson, N.F., Floyd, R.G., Kranzler, J.H., Eckert, T.L., Fefer, S.A., Morgan, J.B. (2019) Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology*, 72, 29-48. https://doi.org/10.1016/j.jsp.2018.12.004.

Bentler, P.M., & Chou, C.P. (1993). Some new covariance structure model improvement statistics. Sage Focus Editions, 154, 235-255. https://doi.org/10.1177/0049124192021002006

Bollen, K. A., & Long, J. S. (Eds.). (1983). *Testing structural equation models*. Sage.

Chadwick, O., Cuddy, M., Kusel, Y. and Taylor, E. (2005), Handicaps and the development of skills between childhood and early adolescence in young people with severe intellectual disabilities. *Journal of Intellectual Disability Research*, 49, 877-888. https://doi-org.dist.lib.usu.edu/10.1111/j.1365-2788.2005.00716.x

Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. *Journal of Child Psychology and Psychiatry*, 43(6), 785–796.

Dell'Armo, K.A, Tassé M.J., (2019) The Role of Adaptive Behavior and Parent Expectations in Predicting Post-School Outcomes for Young Adults with Intellectual Disability. *Journal of Autism and Developmental Disorders.* 49, 1638-1651. https://doi.org/10.1007/s10803-018-3857-6

Ditterline, J., Banner, D., Oakland, T., Becton, D. (2008) Adaptive Behavior Profiles of Students with Disabilities, *Journal of Applied School Psychology*, 24:2, 191-208, DOI: 10.1080/15377900802089973

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662–680. https://doi.org/10.1037/0021-9010.70.4.662

Farmer, R. L., Floyd, R. G., & McNicholas, P. J. (2020). Is the Vineland-3 Comprehensive Interview Form a Multidimensional or Unidimensional Scale?: Structural Analysis of Subdomain Scores Across Early Childhood to Adulthood. *Assessment*, https://doi.org/10.1177/1073191120947804

Frost, K.M., Hong, N., Lord, C., (2017), Correlates of adaptive functioning in minimally verbal children with autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, 112(1), 1-10. DOI: 10.1352/1944-7558-122.1.1

Gligorović, M., & Đurović, N. B. (2014). Inhibitory control and adaptive behavior in children with mild intellectual disability. *Journal of Intellectual Disability Research*, 58(3), 233–242. https://doiorg.dist.lib.usu.edu/10.1111/jir.12000

Harrison, P. L., & Oakland, T. (2015). ABAS-3. Torrance: Western Psychological Services.

Heyman, M., & Hauser-Cram, P. (2019). The influence of the family environment on adaptive functioning in the classroom: A longitudinal study of children with developmental disabilities. *Research in Developmental Disabilities*, 86, 20–30. https://doi-org.dist.lib.usu.edu/10.1016/j.ridd.2019.01.001

Horn, J.L., McArdle, J.J. A practical and theoretical guide to measurement invariance in aging research. *Exp Aging Res.* 1992 Autumn-Winter;18(3-4):117-44. doi: 10.1080/03610739208253916. PMID: 1459160.

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests*. www.InTestCom.org.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.

Limbers, C. A., Newman, D. A., & Varni, J. W. (2009). Factorial invariance of child self-report across race/ethnicity groups: A multigroup confirmatory factor analysis approach utilizing the PedsQL™ 4.0 Generic Core Scales. *Annals of Epidemiology, 19*(8), 575-581.

MacLean, H., McKenzie, K., Kidd, G., Murray, A. L., & Schwannauer, M. (2011). Measurement invariance in the assessment of people with an intellectual disability. *Research in developmental disabilities*, 32(3), 1081-1085.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Pyschometrika*, 58, 525-543.

Milne, S. L., McDonald, J. L., & Comino, E. J. (2013). Adaptive function in preschoolers in relation to developmental delay and diagnosis of autism spectrum disorders: Insights from a clinical sample. *Autism*, 17(6), 743-753.

Mouga, S., Almeida, J., Café, C., Duque, F., & Oliveira, G. (2015). Adaptive profiles in autism and other neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, 45(4), 1001–1012. https://doi-org.dist.lib.usu.edu/10.1007/s10803-014-2256-x

Oakland, T., & Harrison, P.L. (2008) Adaptive Behavior Assessment System-II., Clinical use and interpretation., *Academic Press.* 399-410. https://doi.org/10.1016/B978-0-12-373586-7.X0001-X

Pandolfi, V., & Magyar, C. I. (2021). Vineland-3 Structural Validity and Interpretability of Domain Scores: Implications for Practitioners Assessing Adolescents with Developmental Conditions. *American Journal on Intellectual and Developmental Disabilities, 126*(3), 216–229. https://doi.org/10.1352/1944-7558-126.3.216

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, *60*, 65-82.

Peterson, C. H., Lomas, G. I., Neukrug, E. S., & Bonner, M. W. (2014). Assessment use by counselors in the United States: Implications for policy and practice. *Journal of counseling & development*, 92(1), 90-98.

Pugliese, C. E., Anthony, L. G., Strang, J. F., Dudley, K., Wallace, G. L., Naiman, D. Q., & Kenworthy, L. (2016). Longitudinal examination of adaptive behavior in autism spectrum disorders: Influence of executive function. *Journal of Autism and Developmental Disorders,* 46(2), 467-477.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. Retrieved from https://www.jstatsoft.org/v48/i02/.

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). *Intellectual disability: Definition, diagnosis, classification, and systems of support* (12th ed.). American Association on Intellectual and Developmental Disabilities.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales, Third Edition*. Pearson.

Tassé, M. J., Schalock, R. L., Balboni, G., Bersani Jr, H., Borthwick-Duffy, S. A., Spreat, S., ... & Zhang, D. (2012). The construct of adaptive behavior: Its conceptualization, measurement, and use in the field of intellectual disability. *American journal on intellectual and developmental disabilities,* 117(4), 291-303.

Tsermentseli, S., Tabares, J. F., & Kouklari, E. C. (2018). The role of every-day executive function in social impairment and adaptive skills in autism spectrum disorder with intellectual disability. *Research in Autism Spectrum Disorders*, 53, 1–6. https://doi-org.dist.lib.usu.edu/10.1016/j.rasd.2018.05.006

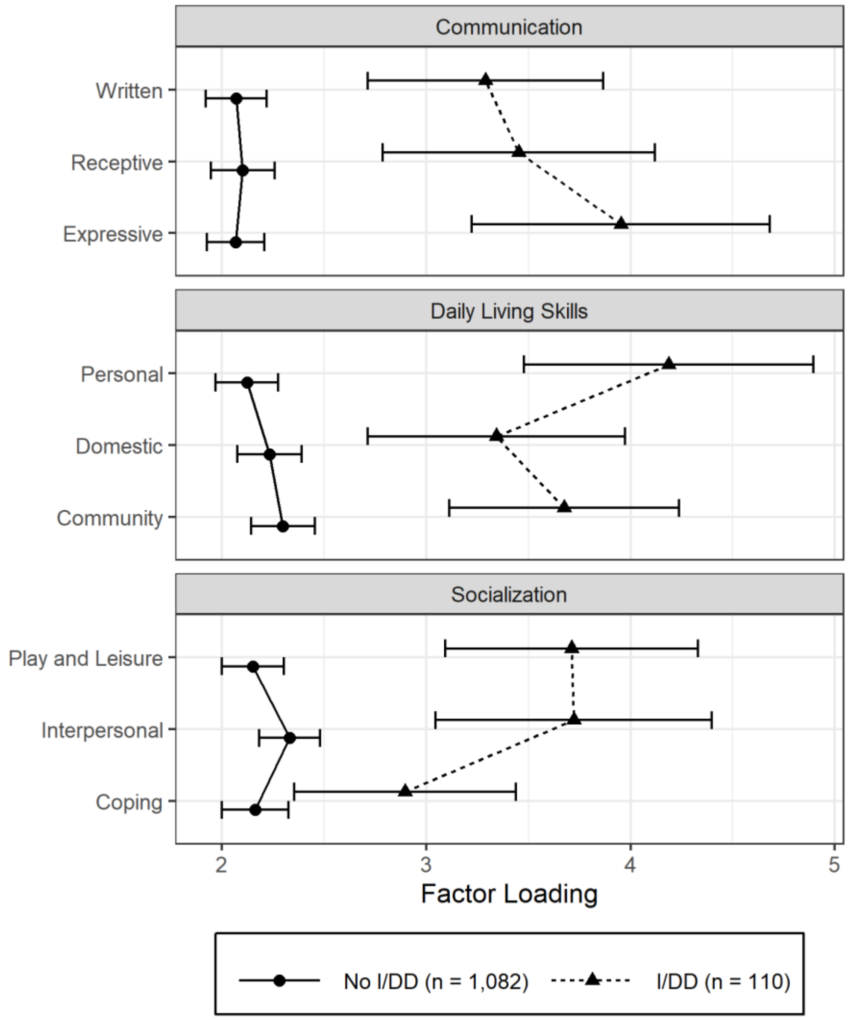
Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European journal of developmental psychology*, *9*(4), 486-492.

Vig, S., & Jedrysek, E. (1995). Adaptive behavior of young urban children with developmental disabilities. *Mental Retardation*, 33(2), 90–98.

Woodman, A. C., Demers, L., Crossman, M. K., Warfield, M. E., & Hauser-Cram, P. (2018). Part C Early Intervention dosage and growth in adaptive skills from early childhood through adolescence. *Early Childhood Research Quarterly*, 43, 73–82. https://doi-org.dist.lib.usu.edu/10.1016/j.ecresq.2018.01.007

Figure 1

Parameter Estimates of the Multigroup Confirmatory Factor Analysis (MG-CFA) Configural Invariance for the Vineland 3 Subdomains (9 parcels)



Note. I/DD = intellectual and developmental disabilities. Each panel represents each of the three domains of the Vineland-3 CIF. The X axis is the factor loading and the Y axis includes the subdomains within each of the three domains. Error bars represent a 95% Wald confidence interval.

Table 1

Domains and Subdomains in the Vineland-3 CIF

Domain	Subdomain	Number of Items	Content Measured	Cronbach Alpha
Communication				.94
	Expressive	49	Expressive language	.97
	Receptive	39	Receptive language	.93
	Written	38	Written communication	.94
Daily Living Skills				.97
	Community	58	Community engagement	.96
	Domestic	30	Cleaning and home management	.95
	Personal	55	Self-care skills	.94
Socialization				.95
	Coping Skills	33	Emotion regulation, rule following, and decision making	.96
	Interpersonal Relationships	43	Social interaction and relationships	.95
	Play and Leisure	36	Entertainment and leisure skills	.95

Note: Information in this table were adapted from the Vineland Adaptive Behavior Scales, Third Edition manual (Sparrow et al., 2016). Cronbach alpha data were obtained from Table 6.7 on page 187.

Table 2

Comparison of Vineland-3 CIF model fits for the full sample and samples of children with and without I/DD using a CFA approach

	Model Fit						LRT*		
Sample	χ^2	*df*	*p*	CFI	TLI	RMSEA	χ^2	*df*	*p*
SINGLE GROUP									
Full, *N* = 1192									
3 factor, unequally weighted	257.45	24	<.001	0.97	0.96	0.09	--	--	--
3 factor, equally weighted	263.38	26	<.001	0.97	0.97	0.09	5.93	2	.052
Unidimensional	608.39	27	<.001	0.94	0.92	0.13	350.94	3	<.001
Children without I/DD, *n* = 1082									
3 factor, unequally weighted	208.82	24	<.001	0.96	0.94	0.08	--	--	--
3 factor, equally weighted	213.41	26	<.001	0.96	0.95	0.08	4.59	2	.101
Unidimensional	438.30	27	<.001	0.92	0.89	0.12	229.49	3	<.001
Children with I/DD, *n* = 110									
3 factor, unequally weighted	60.24	24	<.001	0.96	0.94	0.12	--	--	--
3 factor, equally weighted	79.69	26	<.001	0.94	0.91	0.14	19.45	2	<.001
Unidimensional	159.80	27	<.001	0.84	0.79	0.21	99.57	3	<.001
MULTI-GROUP, 3 factor, unequally weighted									
Configural	269.05	48	<.001	0.96	0.94	0.09	--	--	--
Metric[a]	288.66	54	<.001	0.96	0.95	0.08	19.61	6	.003
Partial Metric[a]	273.82	52	<.001	0.96	0.95	0.08	4.77	4	.312
Scalar[b]	338.07	58	<.001	0.95	0.94	0.09	64.26	6	<.001

Note: Partial metric invariance model freed the loadings for two parcels (Personal subdomain in the Daily Living Skills domain, Coping subdomain in the Socialization domain). Scalar invariant model also allowed these loadings to be free but all intercepts to the same. LRT = Likelihood Ratio Test. CFI = comparative fit index. TLI = Tucker-Lewis index. RMSEA = root mean square error of approximation. *LRT tests compare each model to the first one listed for each section. [a]LRT tests compared to configural. [b]LRT compared to partial metric.