Measurement invariance in the assessment of people with an intellectual disability

Authors:

Hannah MacLean[1] Karen McKenzie[2] Gill Kidd[1] Aja L.Murray[1] Matthias Schwannauer[2]

[1c/o]Child and Adolescent Mental Health Learning Disability Service
14-16 Hope Terrace
Edinburgh
EH9 2AR
United Kingdom

[2]Clinical Psychology
School of Health in Social Science
University of Edinburgh
Teviot Place
Edinburgh
EH8 9AG
United Kingdom

Email: kmckenzi@staffmail.ed.ac.uk
Tel: +44 (0)131 651 3953
Fax: +44 (0)131 651 3971

Abstract

Intellectual assessment is central to the process of diagnosing an intellectual disability and the assessment process needs to be valid and reliable. One fundamental aspect of validity is that of measurement invariance i.e. that the assessment measures the same thing in different populations. There are reasons to believe that measurement invariance of the Wechsler scales may not hold for people with an intellectual disability. Many of the issues which may influence factorial invariance are common to all versions of the scales. The present study, therefore, explored the factorial validity of the WAIS-III as used with people with an intellectual disability. Confirmatory factor analysis was used to assess goodness of fit of the proposed four factor model using 13 and 11 subtests. None of the indices used suggested a good fit for the model, indicating a lack of factorial validity and suggesting measurement invariance of the assessment with people with an intellectual disability. Several explanations for this and implications for other intellectual assessments were discussed.

Key words: Measurement invariance, intellectual assessment, intellectual disability

1. Introduction

Receiving a diagnosis of 'intellectual disability' can impact significantly on the individual concerned (Whitaker, 2010). On the one hand it can facilitate access to services and support but on the other can bring associated stigma (Dagnan & Waring, 2004). It is, therefore, of critical importance that diagnosis is as accurate as possible. Intellectual disability is diagnosed with respect to three criteria: scoring less than 70 on a standard IQ test, demonstrating significant impairments in adaptive functioning and childhood onset (American Psychiatric Association, 1994; British Psychological Society, 2002). Intellectual ability thus represents a third of the diagnostic criteria and it is, therefore, imperative that the intellectual assessments, from which IQ scores are derived, are demonstrated to be valid, reliable and suitable for the group being tested (Kline, 2000). This means that the assessment must meet these criteria when used with people with an intellectual disability, as well as with those who function in the normal range of intelligence.

The Wechsler Scales are commonly used in intellectual disability services (McKenzie, Murray & Wright, 2004; Whitaker, 2010), with short forms being used to provide an estimate of cognitive functioning for screening purposes (e.g. the Wechsler Adult Abbreviated Scale of Intelligence (WASI): Wechsler, 1999) and the full scales being used for diagnostic purposes with people suspected of having an intellectual disability. These scales have generally been found to have good psychometric properties in the normal range of intelligence (e.g. Kaufman & Lichtenberger, 1999; Psychological Corporation, 2008; Strauss, Sherman & Spreen, 2006; Wechsler, 2003). Psychometric properties, however, are population dependent and surprisingly little attention has been paid to the validity of such scales in people in the lower ability range, i.e. those most likely to undergo intellectual assessment for the diagnosis of an intellectual disability. Whitaker (2005, 2010) has been critical of the scales for their lack of appropriateness for this group of people and a number of limitations have been identified in respect of the standardisation samples. For example, the WASI sample only includes 119 people with an intellectual disability and the cognitive profile of those who are included may not be representative of the wider population of people with an intellectual disability (Murray, McKenzie & Lindsey, 2003). Similar difficulties have been identified in the standardisation sample of the Wechsler Adult Intelligence Scale Third Edition (WAIS-III: Wechsler, 1997) (Murray et al., 2003; Whitaker, 2005). The most recent edition of the Wechsler Adult Intelligence Scale (the WAIS-IV: Psychological Corporation, 2008) includes only 104 individuals with an intellectual disability in the standardisation sample, and the exclusion criteria used also means that it is unlikely that this sample is representative of the clinical populations with whom the test is likely to be administered in practice.

The potential lack of validity of the Wechsler scales, as used with people with an intellectual disability is problematic. Implicit in the use of such scales to diagnose an intellectual disability is the idea that the assessments have measurement invariance. This means they should work in a similar way for people in this range of intellectual ability as in the normal range of ability and that they measure the same constructs in both populations. In fact, without an explicit test of measurement invariance it is impossible to definitively conclude that this is the case (Meredith, 1993).

Measurement invariance involves demonstrating that, at a minimum factor loadings are similar across the two groups in question (Meredith, 1993). In line with Whitaker (2010), most of the following examples will relate to the WAIS-III, because this is the assessment for which there is the most research in relation to people with an intellectual disability. As Whitaker notes (2010) many of the same issues apply to other intellectual assessments used with this group of people. he WAIS-III has a factor structure which is based on the theory at the time of the test's development (Kaufman, 2000). This includes four factors which correspond to the index scores of the test. There is evidence that this structure provides a good fit to the data in the normal range of intelligence in both clinical (Heijden & Donders, 2003) and non-clinical (Taub, 2001) populations. Similarly, the Wechsler Intelligence Scale for Children fourth edition (WISC-IV: Psychological Corporation, 2003) and WAIS-IV have been found to have measurement invariance across populations from different countries, (Bowden, Saklofske & Weiss, 2010; Chen, Keith, Weiss, Zhu and YuQui, 2010) although this has not been tested with participants with an intellectual disability.

It cannot, therefore, be assumed that the factor structures will be invariant across the lower range of ability, and in fact there may be several a priori reasons to believe that they will not be. Whitaker (2005) has highlighted the issue of poorer discriminability in the lowest range of ability. By converting raw scores to scaled scores, there is a loss of differentiation between people in the bottom 0.13 per cent of scores. For example, it is possible to achieve a raw score of zero by failing to score any credit on all tests, however, the corresponding scales score will be given as one. While Whitaker (2005) was commenting on the older Wechsler scales, the same issues apply to the more recent versions (Whitaker, 2010). Poorer discriminability has been linked to alterations in intercorrelations between subtests (Fogarty & Stankov, 1995) which in turn impacts on factor structure. Aside from artefacts of test construction, it has also been hypothesised that abilities become less differentiated at lower levels of ability (e.g. Deary, Egan, Gibson, Austin, Brand & Kellaghan, 1996). This too, could impact on factor structure.

Despite the concerns about measurement invariance of intellectual assessments when used with people with an intellectual disability, only one study was found that has examined the factor structure in a sample which included people with an intellectual disability. Jones, Schaik & Witts (2006) tested whether the four factor structure which underpins the WAIS III, and is reflected in the four index scores of the assessment, was applicable for individuals with low IQ (less than 74). WAIS-III data collected as part of routine clinical practice was analysed using an exploratory factor analysis. The authors found that only a two factor solution (representing the verbal and performance factors) was robust, while the four factor solution on which the WAIS III was based was not supported for the low IQ group. This lends support to the idea that the factorial validity of the four factor WAIS-III breaks down in lower ability groups. Unfortunately this study only had 105 data sets, which related to people with low IQ rather than just those with an intellectual disability, and used only exploratory techniques and so an explicit test of the WAIS-III factor structure was not conducted.

The present study seeks to build on this result by using confirmatory factor analysis to explicitly test the proposed factor structure of the WAIS-III. The factor structure underpinning different versions of the Wechsler scales will differ, due to advances in the theoretical basis underpinning the development of the assessment. However, many of the issues outlined above, which are likely to influence measurement invariance of the tests as used with people with an intellectual disability, also apply to the most recent versions of the Wechsler scales (Whitaker, 2010). The results of the study, are therefore, likely to be applicable to a number of other intellectual assessments.

2. **Method**

2.1 Procedure

2.1.1 Recruitment of Clinical Psychologists

The lead clinical psychologist for an intellectual disability service covering a large geographical area in Southern Scotland was given information regarding the study and asked if the service wished to participate. Following obtaining ethical approval from the local Caldicott guardian and permission for data to be collected from the service, the lead clinical psychologist provided contact details for four other clinical psychologists in the area who would be able to provide access to WAIS-III scores. Contact was made with these individuals to brief them about the research and to arrange access to their case files.

2.1.2 Data collection

Historical and current files stored in each intellectual disability service base were reviewed by the first author and information was gathered from those files where a WAIS-III had been scored. The assessments had all been completed by clinical psychologists or trainee psychologists under the supervision of a qualified clinical psychologist. The assessments were completed as part of routine clinical practice for a variety of reasons, including informing interventions and support requirements, as part of capacity assessments and to determine eligibility for receiving intellectual disability services. The following data were collected: gender and age; WAIS-III subtest scores; identifying code. The identifying code recorded the initials of the clinician whose service the file had come from and a data set number. The latter was matched with the client's name and was held separately by the clinical psychologists until the data were entered, after which it was destroyed. This was to ensure that data could be extracted again if lost during the inputting procedure and also to ensure entries were not duplicated. The data were collected over a six month period and covered the period from 1997 to 2010.

2.2 Participants

The WAIS-III data came from individuals who had at one point in time used the intellectual disability service and who had been assessed by clinical psychologists working within the service. Participants would have been resident in the Southern Scotland at the time of the assessments. The sample can be regarded as opportunistic. Data were included for all individuals for whom there were service case files, who had been assessed using the WAIS-III and who had a Full Scale IQ of 69 or less. No information was collected in relation to adaptive functioning or childhood onset. While this raised the possibility that a few individuals may have been included in the data set who met the criterion of significant impairment in intellectual functioning, but not the other criteria for intellectual disability, this was not considered to be problematic for the present study, where the purpose was to examine the factor structure of the WAIS-III of those who were functioning in the intellectual disability range. Data with missing values were omitted from the data set and overall, data were analysed for 140 people with severe intellectual impairment (IQ<55; N= 140, male=76, female= 64; mean age= 38.5 years, sd=14.9) and 264 people with significant intellectual impairment (55<IQ<69; male=161, female=103; mean age= 31.3, sd=13.6).

2.3 Analysis

Confirmatory factor analysis is a process by which a data set can be used to check for theoretical fit against hypothesised models (Kline, 1994) and, as such, was chosen for the purposes of this study.

2.3.1 Sampling for factor analysis

There are a number of different views on the adequacy of the sample size required for factor analysis. Lewis (1995) suggests the number of cases $\geq$5x number of variables while others have suggested, N-n-1 $\geq$50 (where N = number of participants and n = number of variables; Lawley & Maxwell, 1971). The number in both of the samples used in the present study met both of these criteria.

2.3.2. Model tested

The explicit four factor model that underpins the WAIS III was tested. In this explicit model there are four common factors assumed to explain performance on the subtests. The verbal comprehension index is assumed to influence four subtests: vocabulary, similarities, comprehension and information. The perceptual organisation index is assumed to influence four subtests: picture completion, block design, picture arrangement and matrix reasoning. The processing speed index is assumed to influence two subtests: digit symbol coding and symbol search. The working memory index is assumed to influence the three subtests: letter number sequencing, arithmetic and digit span. The four indices are assumed to be explained by a second order *g* factor. The explicit model was also tested using 11 subtests (excluding the two optional subtests of comprehension and picture arrangement) which are sufficient for the calculation of index scores. The object assembly subtest does not contribute to the calculation of IQ or Index scores. Research has therefore, tended to exclude this subtest in factor analyses (e.g. Jones et al., 2006) and it was also excluded in the current study.

3. Results

3.1 Preliminary Analysis

A computer evaluation of assumptions using SPSS version 17.0 was carried out to assess the suitability of the data sets for analysis (Tabachnick & Fidell, 2001). The data were found to be acceptable in terms of factorability, multicollinearity and singularity. However, a Kolmogorov-Smirnov test suggested that all the subtests differed significantly from a normal distribution in both samples. As such, it was considered appropriate to employ EQS 6.1 (Bentler, 2007), a non parametric test, to run the confirmatory factor analysis. Yuan and Bentler (1998) suggest, in fields such as psychology when data does not always meet assumptions of multivariate normality, it is not always appropriate to use normal theory methods, which can distort the results. Using the EQS, ROBUST option a Satorra-Bentler (1994) scaled test statistic (S-B χ^2) was used to correct the mean of the sampling distribution so that it was closer to the expected mean, and to evaluate the goodness of fit of the model. Studies have shown that this statistic works well for non normal and normal data (Curran, West, & Finch, 1996) and is considered the best for dealing with non normal data (Bentler, 2007).

As the results of confirmatory factor analyses can vary markedly depending on which fit indices are used to evaluate the model, the comparative fit index (CFI) and root mean square error of approximation (RMSEA) were also obtained. Both of these were adjusted using the Satorra-Bentler scaled χ^2 statistic**.** None of the indices suggested a good fit ($p(\chi^2) < .01$ CFI<0.9, RMSEA >.05).

Table 1 illustrates the goodness of fit statistics for the proposed WAIS-III four factor model based on 13 and 11 subtests.

Insert table 1 about here

4. Discussion

The present study used the WAIS III intellectual assessment to assess whether the explicit four factor structure on which it is based also held for people with an intellectual disability. The study found evidence that the factor structure of the WAIS-III does break down in people with an intellectual disability and thus suggests that the instrument lacks measurement invariance in this group. This was true irrespective of the extent of intellectual impairment, as the structure was not supported in either those with a significant intellectual impairment or in those with a severe intellectual impairment.

There are a number of possible explanations for the failure to find a good fit of the WAIS-III factor structure in people with an intellectual disability. Firstly, the result may have statistical cause. The data in both groups were positively skewed- with scores concentrated in lower levels of ability and likely to be indicative of floor effects. Floor effects can act to increase the inter-correlations between subtest scores because scoring is similarly restricted across the whole range of subtests. The implication of this is that, until tests with adequate floors are developed, factorial validity will be unattainable in low ability groups. It is fair to expect the same limitation to apply to the WAIS-IV, which despite having somewhat improved floors, will be unlikely to be sufficiently discriminating in the lower range of ability to do much better than the WAIS-III in this respect.

Secondly, because the groups were defined on the basis of full scale IQ and subtest scores all contribute to this score, there may have been an artificial attenuation of test inter-correlations. This is because when groups are defined by full scale IQ, in order to fall within a given IQ range, extreme scores in one test must be offset by extreme scores in the opposite direction in another and this can produce spurious negative correlations and an overall reduction in intercorrelations (Jensen, 2003). This is a general problem associated with taking a subgroup of a normally distributed population and can only be satisfactorily overcome by using an independent measure of IQ to define ability groups.

Finally, it is possible that the results reflect real differences in the factor structure of intelligence in lower ability groups. This idea follows from the differentiation hypothesis, first advanced by Charles Spearman in 1927 and appearing in various forms since (e.g. Deary et al., 1996; Reynolds, Keith & Beretvas, 2010; Saklofske, Yang, Zhu, & Austin 2008). The hypothesis states that at higher levels of ability, narrow abilities, such as those measured by the Wechsler scales, become more differentiated and thus less correlated with one another. Conversely, lower ability groups will exhibit higher intercorrelations between abilities. This effect, if of a sufficient magnitude, could result in differing factor structures in people with intellectual disabilities and would apply irrespective of the particular measuring instrument. Therefore, although the WAIS-III has now been superseded by the WAIS-IV, if the results of the present study are attributable to dedifferentiation, the same limitations identified here will apply to this instrument too. Another facet of this problem, of course, is that in the absence of instruments with adequate floors, it is difficult to establish if factor structure differences reflect true differentiation or are simply an artefact of test construction whereby discriminability is poorer in the lower range of ability.

The idea that there may be a different factor structure of intelligence may also be implicit in the idea that there are qualitative differences in ability between people with an intellectual disability and the rest of the population. The IQ<70 cut off point is then argued to represent not just an arbitrary point, but a meaningful reflection of these qualitative differences. Until, however, tests with adequate discriminability in people with an intellectual disability can be developed, the questions of differentiation and qualitative difference in ability structure will remain unanswered. The results also highlight the importance of using additional information, such as that obtained from measures of adaptive functioning, in decisions about diagnosis, support and treatment planning. Tests such as the Vineland Adaptive Behaviour Scale – Second Edition (Vineland II; Sparrow, Cicchetti, & Balla, 2005) and Adaptive Behaviour Assessment System - Second Edition (ABAS-II; Harrison & Oakland, 2003) are standardised tests of functioning, which have been normed for people with an intellectual disability.

5. Conclusions

The present study failed to establish factorial validity of the WAIS-III in people with an intellectual disability. This suggests that the instrument lacks measurement invariance in this group. A number of possible explanations for this have been discussed, however, definitive interpretation is made difficult in the presence of floor effects. Regardless, the results highlight the importance of developing tests which have adequate floors, if such tests are to be used either to test hypotheses about the factor structure of intelligence in lower ability groups or to demonstrate measurement invariance of major intelligence scales in people with an intellectual disability. Until then, clinicians should continue to interpret the results of intellectual assessments used with people with an intellectual disability with caution, bearing in mind their potential limitations with this group of people.

6. References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: APA.

Bentler, P. M. (2007). EQS 6.1. *Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bowden, S.C., Saklofske, D.H. & Weiss, L.G. (2010). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale–IV in the United States and Canada. *Educational and Psychological Measurement, 22*, Early view: doi: 10.1177/0013164410387382

British Psychological Society (2002). *Learning Disability: Definitions and Contexts.* Leicester: The British Psychological Society.

Chen, H., Keith, T.Z., Weiss, L., Zhu, J. & YuQui, L. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences, 49 (7)*, 677-682

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Dagnan, D. & Waring, M. (2004). Linking stigma to psychological distress: Testing a social-cognitive model of the experience of people with intellectual disabilities. *Clinical Psychology and Psychotherapy, 11*, 247-254

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. doi: DOI: 10.1016/S0160-2896(96)90008-2. *Intelligence, 23*(2), 105-132.

Fogarty, G. J. & Stankov, L. (1995). Challenging the 'Law of Diminishing Returns'. *Intelligence, 21*, 157-174.

Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System* (2nd ed.). Minneapolis, MN: Pearson Assessment.

Heijden, P. & Donders, J. (2003). WAIS-III factor index score patterns after traumatic brain injury. *Assessment 10(2),* 115-122.

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. doi: DOI: 10.1016/S0160-2896(01)00094-0. *Intelligence, 31*(2), 95-105.

Jones, J, Schaik, P. van. & Witts, P. (2006). The factor structure of the WAIS-III in persons with low IQ. *British Journal of Clinical Psychology, 45* (2), 145-152.

Kaufman, A.S. & Lichtenberger, E.O. (1999). *Essentials of WAIS III Assessment*. New York: John Wiley & Sons.

Kaufman, A.S. (2000). Two and three factor solutions of the WAIS-III. *Assessment, 8*(3), 267-280.

Kline, P. (2000). *A Psychometrics Primer*. London: Free Association Books.

Lawley, D.N. & Maxwell, A.E. (1971). *Factor analysis as a statistical method*. London: Butterworth and Co.

Lewis, J.R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7* (1), 57-78.

McKenzie, K., Murray, G.C. & Wright, J. (2004). Adaptations and accommodations: The use of the WAIS III with people with a Learning Disability. *Clinical Psychology, 43*, 23-26.

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

Murray, G., McKenzie, K. & Lindsay, W.R. (2003). A pilot study of the use of the WAIS III with people with a learning disability. *Clinical Psychology*, *28*, 17-21.

Psychological Corporation (2003). *WISC-IV technical and interpretive manual*. San Antonio, TX: Psychological Corporation.

Psychological Corporation, (2008). *WAIS-IV Technical and interpretive manual*. San Antonio, TX. Psychological Corporation.

Reynolds, M. R., Keith, T.Z. & Beretvas, S.N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence, 38*(2), 231-241.

Saklofske, D. H., Yang, Z. M., Zhu, J. J., & Austin, E. J. (2008). Spearman's law of diminishing returns in normative samples for the WISC-IV and WAIS-III. *Journal of Individual Differences, 29*(2), 57-69.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland Adaptive Behavior Scales* (2nd Ed.). Minneapolis, MN: Pearson Assessment

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Strauss, E., Sherman, E.M.S. & Spreen, O. (2006). *A compendium of neuropsychological tests. Administration, norms and commentary.* New York: Oxford University Press.

Tabachnick, B.G. & Fidell, L.S. (2001). *Using Multivariate Statistics. Fourth Edition*. MA: Allyn & Bacon.

Taub, G.E. (2001). A confirmatory factor analysis of the Wechsler Adult Intelligence Scale – third edition: is the verbal/performance discrepancy justified? *Practical Assessment, Research & Evaluation, 7(22), pp n/a*

Wechsler, *D. (1997). Wechsler Adult Intelligence Scale - Third Edition.* London: Psychological Corporation.

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *WISC-IV technical and interpretative manual.* San Antonio: Psychological Corporation.

Whitaker, S. (2005). The uses of the WISC-III and the WAIS-III with people with a learning disability: Three concerns. *Clinical Psychology Forum, 50*, 37-40.

Whitaker, S. (2010). Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III. *Personality and Individual Differences*, 48, 517–521

Yuan, K–H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 51*, 289-309.

Table 1: Goodness of fit statistics for WAIS-III four factor models based on all 13 subtests and excluding subtests Comprehension and Picture Arrangement

Group	S-B χ^2	d.f.	p	CFI	RMSEA (90% confidence interval)
		All 13 subtests			
Severe	134.5	61	.00000	.687	.093 (.071-.114)
Significant	131.5	61	.00001	.853	.066 (.051-.082)
		Excluding subtests Comprehension and Picture Arrangement			
Severe	87.9	40	0.0005	.746	.090 (.063-.116)
Significant	86.6	40	0.0003	.869	.067 (.047-.085)

Conflict of interest: None
Sources of funding: None