Profile Analysis on the WISC-IV and WAIS-III in the low intellectual range: Is it valid and reliable?

By

Simon Whitaker

And

Shirley Gordon

This paper examines how far it is valid to generate a profile of an individual's cognitive abilities using the WISC-IV or WAIS-III for individuals in the low ability range. Data are presented which demonstrate that the WISC-IV and WAIS-III assessments produce different cognitive profiles, when given to the same 16-year-olds who receive special education. It is suggested that at the low IQ level, subtest and index scores may lack sufficient stability for the WISC-IV or WAIS-III to produce reliable cognitive profiles.

Introduction

Recent evidence has suggested that the accuracy with which the commonly used IQ tests such as the WISC-IV and WAIS-III can measure low IQ is considerably less than had previously been supposed. A meta analysis of the stability of low IQ by Whitaker (2008) estimated that 14% of Full Scale IQ (FS IQ) scores will change by 10 points or more when reassessed; Whitaker and Wood (2008) have shown that the WAIS-III, and more particularly the WISC-III are subject to a floor effect and Gordon et al (in press) has shown that the WISC-IV systematically gives scores on average about 12 points less than the WAIS-III when given to 16 year olds in the low ability range.

In interpreting IQ tests there is an increasing emphasis on the profile of scores an individual obtains (c.f. Prifitera et al 2008). Both the WISC-IV (Wechsler 2004) and WAIS-III (Wechsler 1998) have sections on their record forms to aid the calculation of a profile of cognitive abilities: significant disparities between index scores and strengths and weaknesses in subtests. There is, however, no evidence that the profiles obtained are either valid or reliable at the low ability range. In the light of the recent findings on the error levels of in measuring low IQ, the current paper will examine how valid it is to generate a profile of an individual's cognitive abilities using the WISC-IV or WAIS-III for individuals in the low ability range.

If it is valid to calculate a profile, then these profiles should be reliable and valid (Anastasi and Urbina 1997). Part of the test of reliability would be to demonstrate that the result of a profile analysis is stable, that is, if the same individuals were assessed twice on the same assessment their cognitive profile would not change a great deal. Part of the demonstration of validity would be to show that similar profiles are produced if an individual was assessed on two similar assessments.

Stability

Stability of IQs and Index score differences

Whitaker (2008) conducted a meta analysis of stability of IQ in the low range and reported a weighted mean stability coefficient .82 for FS IQ which gave mean Standard Error of Measurements (SEMs) of 6.4. This SEM is effectively the standard deviation (SD) of the distribution of difference in IQ scores between first and second assessments and can be used to predict the proportion of second assessments that would vary by specific amounts. It can therefore be used to predict the proportion of second test results

that would vary by specific amounts. On the basis of this, Whitaker (2008) calculated that 61% of FS IQs would change by less than six points, but that 13% would change by 10 points or more. These estimates were found to be similar to the changes that actually occurred in those studies that reported on the proportion of clients whose IQ changed by specific amounts, where 57% of FS IQs changed by less than 6 points and 14% of IQs changed by ten points or more, suggesting that this is a valid method of calculating the proportion of scores that vary by specific amounts.

Whitaker (2008) also reported mean stability coefficients of .77, and .78 for Verbal IQ (VIQ) and Performance IQ (PIQ) respectively, but did not calculate the respective SEMs or an estimate of the proportion of scores that would be expected to change by specific amounts on second testing. However, using the same calculations it would be predicted that for these stability coefficients, 55% of V IQs would change by less than 6 points but 19% would change by 10 points or more and that 57% of P IQs would change by less than 6 points, but 17% would change by 10 points or more.

This lack of stability in IQ scores will affect difference between V IQ and P IQ from one assessment to another. If V IQ and P IQ varied totally independently of one another between assessments, that is, there was no correlation between changes in V IQ and P IQ between assessments, then it would be expected that about 18% of differences between V IQ and P IQ would change by 10 points or more. As the critical values given in the WAIS-III manual for a significant difference between V IQ and P IQ are 6.43 and 8.76 for the .15 and .05 significance level respectively, it would be likely that many clients who had significant differences between their V IQ and P IQ when first assessed would not have a significant difference when they were re-assessed.

The same logic applies to the index scores. One of the studies (Canivez and Watkins 2001) in Whitaker's meta analysis gave stability coefficients for index scores of: .84, .87, and .81 for Verbal Comprehension Index (VCI), Perceptual Organisation Index (POI) and Freedom from Distractibility Index (FDI) respectively on the WISC-III. This corresponds to SEMs of 6.00, 5.41 and 6.54 for VCI, POI and FDI respectively. So for VCI it would be expected that 63% would change by less than 6 points and 11% by 10 points or more, for POI 69% by less than 6 points and 8% by 10 points or more and for FDI 60% by less than 6 points but 9% by 10 points or more. As the critical values for differences in index scores on the WAIS-III vary between 7.03 for VCI-WMI to 9.23 for WMI-PSI for the .15 significance level and between 9.57 for VCI-WMI to 12.56 for WMI-PSI for the .05 significance level it is clearly possible that the profile of significant differences between index scores will change between assessments.

Stability of subtest strengths and weaknesses

The same argument applies to the profile of subtest scores. According to the WISC-IV and WAIS-III manuals (Wechsler 2004, 1998) a subtest is regarded as a strength or weakness if it varies by more than a critical amount from the mean scaled score, with critical values being given for both the .15 and .05 significance levels. Whitaker (2008) gave figures for the reliability of subtests as part of his meta analysis of the stability of IQ. The weighted mean reliability figures were as follows:

Information	.62
Similarities	.48
Arithmetic	.58
Vocabulary	.61
Comprehension	.47
Digit span	.79
Picture completion	.49
Picture arrangement	.55
Block design	.62
Object assembly	.63
Coding/digit symbol	.67
Mean	**.59**

The mean stability coefficient of .59 and the SD of 3 for scaled scores gives a mean SEM of 1.92, which corresponds to 81% of re assessments varying by 2 points or less and 19% varying by 3 points or more. As the critical figures for a significant deviation from the mean subtest score are of the order of 2.5 points at the .15 significance level and 3.0 at the .05 significance level, it would be expected that there would be a different pattern of strengths and weaknesses each time the individual was assessed. The degree to which this would occur would depend on the degree to which the variation is subtest scores was independent.

A key factor in the stability of a profile based on differences between IQs or index scores and/or the strengths and weaknesses of subtests is the degree to which IQ, index or subtests scores vary independently of one another between assessments. If they varied totally independently then, as noted above, it is likely there would be major changes in profiles between assessments. However, this would only happen if each subtest was affected by its own unique set of factors, which is not likely. Such factors as distraction in the test situation, fatigue or cooperation on the part of the client are likely to affect all subtests, though not necessarily to the same extent. Therefore, the only way to find out the degree to which differences between index scores and IQ change between assessments is to do an empirical study in which this is observed but so far such a study has not been done. However, in view of this analysis it is likely that any profile produced will not be as stable as indicated in the test manuals and other recent publications (c.f. Prifitera et al 2008).

Evidence for validity

Part of a demonstration of the validity of profiles would be to show that two assessments that purport to measure the same cognitive abilities produce the same pattern of disparity between index scores/IQs and strengths and weaknesses of subtests. There has been some research in this area. In order to see if the WISC-IV and WAIS-III gave similar IQs and index scores, Gordon et al (in press), gave both assessments to seventeen 16-year-olds in special education. They found that the mean FS IQ on the WISC-IV was 53.00 but on the WAIS-III it was 64.82, which is a difference of nearly 12 points, suggesting that the two assessments are not given equivalent scores at the low IQ level.

The data from this study are used here to calculate and compare the profiles of the disparities between index scores and the strengths and weaknesses of subtests between the two assessments. This analysis is presented in Table 1, which shows all the significant disparities between index scores and all the strengths and weaknesses for the common subtests, for both the WISC-III and WAIS-IV at the .15 level of significance (those that are also significant at the .05 level are marked with an *). In doing this analysis it is assumed that the Perceptual Organisation index (POI) on the WAIS-III is equivalent to the Perceptual Reasoning Index (PRI) on the WISC-IV.

The percentage agreement between the two assessments was calculated by finding the percentage of times a significant disparity or strength/weakness on one assessment was also shown on the other assessment. For index score disparities this was found to be 27.12% at the .15 level and 25.00% at the .05 level. For the strengths and weaknesses in the subtests there was a 10.26% agreement at the .15 level and a 13.64% at the .05 level. It is also notable that on index disparities Subjects 5 and 13 show index score disparities in the opposite direction on the WISC-IV and WAIS-III. On the WISC-IV subject 5 had the Working Memory Index (WMI) significantly higher than Verbal Comprehension Index (VCI) at the .15 level, but on the WAIS-III VCI was significantly higher than WMI at the same level. Subject 13 showed two disparities in the opposite directions with the WISC-IV and WAIS-III, with the WISC-IV WMI significantly higher at the .15 level than both VCI and PRI, however, on the WAIS-III the WMI was significantly lower than either VCI or PRI at the .15 level. The same was seen at the .05 level of significance for VCI WMI disparity.

Discussion

The above analysis has established the following:

- Subtests lack stability which, depending on the degree to which they vary independently of one another, would mean that a profile of strengths and weaknesses would change from assessment to assessment.
- Although Index scores have about the same degree of stability as IQs and so would not be expected to vary as much as subtests, they may still vary sufficiently to produce different profile disparities in index scores between assessments, depending on the degree to which they vary independently of one another between assessments.
- When the WISC-IV and WAIS-III were given to the same people different profiles of cognitive abilities were produced for each assessment. Whether this is due to a lack of stability in the index scores and subtests or whether it is due to problems with validity, in that apparently similar subtest and index ability assessments are in fact measuring different things, it is not posible to tell from this data.

It therefore seems that there are problems with using the WISC-IV and the WAIS-III to produce a profile analysis in individuals in the low IQ range. Clearly more work needs to be done on this. What is not known is the degree to which both subtests and index

scores vary independently of one another between assessments. If they do not very independently of one another, even though subtest and index scores may lack stability, the profiles would not change between assessments. Stability studies need to be done in which the same assessment is given to people in the low ability range twice and the degree to which their profile of ability change assessed. Until such studies are conducted with the WISC-IV and WAIS-III any profile produced should be treated with caution.

References

Anastasi, A. and Urbina,S. (1997). *Psychological Testing (seventh edition)*. Upper Saddler River: Prentice-Hall Inc

Gordon, S., Duff, S. Davison, T and Whitaker, S. (in press). Comparison of the WAIS-III and WISC-IV in 16 year old special education students. *Journal of Applied Research in Intellectual Disability*.

Prifitera, A., Saklofske, D.H. and Weiss, L.G. (2008). WISC-IV Clinical Assessment and Intervention. San Diego: Second Edition.

Wechsler, D. (1998). *Wechsler Adult Intelligence Scale – Third UK Edition: Administrative and Scoring Manual*. London: The Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children – Fourth UK Edition: Administrative and Scoring Manual*. London: The Psychological Corporation.

Whitaker, S. (2008). The stability of IQ in people with low intellectual ability: An analysis of the literature. *Intellectual and Developmental Disabilities*, 46, 120-128.

Whitaker, S. & Wood, C. (2008). The distribution of scale score and possible floor effects on the WISC-III and WAIS-III. *Journal of Applied Research in Intellectual Disabilities*. 21, 136-141.

Table 1

Profile Analysis

| | | |Pr .15 |and |.05* | | | | | | | | | | | |Subject | |VCI-PRI |VCI-WMI |VCI-PSI |PRI-WMI |PRI-PSI |WMI-PSI |VC |SI |DS |CO |LN |BD |MR |SS |Cod | |1 |wisc-iv | | | | | | |S* | | | | | | | | | |wais-iii | | | | | | | | |S | | | | | | | |2 |wisc-iv |neg | |neg* |pos* |neg* |neg* | | | |W |W* |S* | |W* |S | | |wais-iii |neg* | | |pos* | |neg* | | |W | | |S* | | | | |3 |wisc-iv |neg* |neg* |neg* |pos |pos* | | |W |S* | |W* |S |S* | |W | | |wais-iii | |neg* | |pos* | | | | |S | | | | | | | |4 |wisc-iv | | | | | | | | | | | | | | | | | |wais-iii | | | | | | | | | | | | | | | | |5 |wisc-iv |neg |neg |neg* | | | |W | | | |S |W | | | | | |wais-iii | |pos* | |pos | |neg | | | | | | | | | | |6 |wisc-iv |neg* | | |pos* | |neg* |W | | | |W | |S | | | | |wais-iii |neg |pos* | |pos* |pos | | | | | | | | | | | |7 |wisc-iv | | | |pos* | |neg* | | | | |W* | | | | | | |wais-iii | | |neg | |neg |neg* | | | | | | | | |S* | |8 |wisc-iv | | |neg* | |neg | | | | | | | | | | | | |wais-iii | | |neg | | |neg | | | | | | | | | | | |9 |wisc-iv | |neg | | | | | | | | | | | | | | | | |wais-iii |neg* | |neg* |pos* | |neg* |W* | | | | | |S* | | | | |10 |wisc-iv | |pos* | |pos* |pos |neg | | | | |W* | | | | | | |wais-iii |neg | | |pos* | | | | | | |W* | | | | | |11 |wisc-iv |neg* |neg* |neg* | | | |W* | | | |S* | | |S* | | | | |wais-iii |neg |neg* |neg* |neg* | | | | | | | | | | | | |12 |wisc-iv | | | | | | | |S | | |S* | | | |W | | |wais-iii |pos |pos* | | | | | | | | | | | | | | | |13 |wisc-iv | |neg* | |neg | | | | |S | | | | | | | | | |wais-iii | |pos* | |pos | | | | | | | | | | | | |14 |wisc-iv | | | | | | | | | | | | | | | | | |wais-iii | |pos | | | | | | | | | | | | | | | |15 |wisc-iv |neg* |neg* | | |pos* |pos | | |S* | | | |S* | | | | |wais-iii | | | | | | | | |S | | | | | | | |16 |wisc-iv |neg* | |neg* | |neg |neg* |W* |W* | | | | | |S* | | | |wais-iii | | |neg* | |neg* |neg* | | | | | | | | | | |17 |wisc-iv | | | |neg* | | |W | | | | | | | | | | | |wais-iii |pos* |pos* | | |neg |neg | |S | | | | | | | | |

Table caption

Table 1
The analysis of the cognitive profiles for each of the seventeen 16-year-olds on both the WISC-IV and WAIS-III. VCI=Verbal Comprehension Index; PRI=Perceptual Reasoning Index (on the WISC-IV) and Perceptual Organisation Index (on the WAIS-III), WMI=Working Memory Index, PSI=Processing Speed Index, VC=Vocabulary, SI=Similarities, DS=Digit Span, C0=Comprehension, LN=Letter-Number Sequencing, BD=Block Design, MR=Matrix Reasoning, SS=Symbol Search, Cod=Coding (on the WISC-IV) and Digit Symbol-Coding (on the WAIS-III), neg=significant negative difference between the two index scores at the .15 level, neg*= significant negative difference between the two index scores at the .05 level, pos=significant positive difference between the two index scores at the .15 level, pos*=significant positive difference between the two index scores at the .05 level, S=a subtest that shows a strength at the .15 level S*=a subtest that shows a strength at the .05 level W=a subtest that shows a weakness at the .15 level W*=a subtest that shows a weakness at the .05 level.