The uses of the WISC-III and the WAIS-III with people with a learning disability: Three concerns

By

Simon Whitaker

Published in Clinical Psychology, 50 July 2005, 37-40

Summary
From information in the WISC-III and WAIS-III manuals it is apparent that these assessments may have anomalous item difficulty, a higher floor, and inaccurate percentiles, at lower IQ levels.

Introduction

This paper considers three potential concerns with regard to the use of the WAIS-III and the WISC-III with people who have a learning disability, which are:

- There may be anomalies in the degree of difficulty in some test items associated with low scale scores.
- The floor effect on both tests may be greater than is acknowledged in the test manuals.
- The percentile quoted in the manual make the false assumption that IQ is normally distributed at the low end and consequently are inaccurate.

The standardization of the WISC-III and WAIS-III

The WISC-III (UK), published in 1992, is an Anglicised version of the WISC-III (US); however, care was taken when developing the US version not to use language that would need altering when used in the UK. It makes use of UK norms, from a stratified sample of 824 children, from eleven age groups between 6 and 16 years, with between 67 and 87 children per age group. However, no children from special schools were included in the samples, so there may not have been any children with IQs 2 standard deviations (SDs) below the norm.

The WAIS-III (UK), published in 1998, is an Anglicised version of the WAIS-III (US). It is specified in the manual that the degree of change from the US version was conservative in order to minimise any changes in the item difficulty. Unlike the WISC-III (UK), it makes use of the US norms which have has demonstrated to be roughly equivalent to the norms in the UK (Wechsler 1998).

Both assessments measure intelligence by giving the client a series of subtests. The WISC-III has 13 subtests, 10 of which are used to calculate IQ. The WAIS-III has 14 subtests, 11 of which are used to calculate IQ. The two assessments have 12 subtests in common, 9 of which are used to calculate IQ. The raw score a client obtains on a subtest is converted to a normalized scale score with a mean of 10 and an SD of 3. The conversion of raw score to scale score is based on cumulative frequency distributions of raw scores for each age group, normalizing these distributions, and then calculating the appropriate scale score for these raw scores. The difficulty level for each scale score, that is, the percentage of people in a particular age group who obtain a particular scale score,

should therefore be the same across the subtests.

Degree of difficulty of test items

As noted above, although both the WISC-III and the WAIS-III were standardised on large stratified samples, it is unclear to what extent this sample included people with learning disabilities. Even if it did include a representative sample of people with learning disabilities there still would have been relatively few people with learning disability at each age level. The WISC-III (UK) sample used between 67 and 87 children per age group, therefore you would only expect to have between 1.5 and 2 children with IQs or scale scores two or more standard deviations below the average, in any age group. The WAIS-III (US) standardisation, (with the exception for the older age groups) used 200 adults per age group, which would mean one would expect to have only 5 people scoring 2 SDs below the norm. Therefore very few people would be failing the subtest items at the low scale score levels. Scale scores of 2 being 2.67 SDs below the mean, one would expect only 0.38% of the population as a whole to fail, or less than one person per age group in the stratified sample on the WAIS-III. Therefore it would be expected that items given early in subtests would not be failed by any of the sample and so it would not be possible to say empirically how difficult they were compared with other early items. Therefore allocation of raw scores to scale scores at these low scale score levels must have been based on an extrapolation of the performance of the more intellectually able, rather than an empirical testing of the degree of difficulty of the test items. It therefore seems reasonable to consider if there are any apparent anomalies in the degree of difficulty of test items within the two assessments at these low scale score levels. There seems to be at least two possible sources of error in item difficulty: the use of American English in the UK version of the tests, and the use of language in the subtest instructions that is more complex than the level of the item.

The use of American English. When the WAIS-III (UK) was produced attempts were made to minimise the changes made to the US version, which therefore raises the possibility that some items use a version of English that is either not used in the UK or is used with far less frequency than it is in the US. Possible example of this are found in some early items in the Comprehension subtest on the WAIS-III, for example item 6 refers to the uses of a “parole system” and item 8 to people in some professions needing “licence” before offering services to the

public. In the UK it is rare to refer to somebody to be “released on parole” (except in US films) and the term used for being legally able to practise is usually to be registered. A similar problem may occur in the instructions for Digit Span, which are as follows: “I am going to say some numbers. Listen carefully, and when I am through (American), I want you to say them right back to me (American). Just say what I say.” Although with all these examples it is apparent what the questioning is getting at if one thinks it through, it is not at all clear that the degree of intellectual ability needed to work out that “licensed” means “registered” is not greater than that required to understand why it is necessary to be licensed.

The degree of difficulty in item directions exceeds the apparent degree of difficulty of the item. In some items the degree of intellectual ability required to understand the instruction or the requirements of an item may exceed the level of ability that that item is testing. Although some of the Performance subtests give demonstrations or practice items to the client, many rely on a set of verbal instructions only. Some idea of the level of understanding that would be expected from a client with low intellectual ability can be obtained from the Vocabulary subtest. A person in the reference group age range of 20 to 34 years on the WAIS-III would only be required to get a raw score 5 in order to get scale score of 2, which could be done by giving basic one point responses to the following words: “Bed”, “Ship”, “Penny”, “Winter”, and “Repair”, all common words. However, to gain the same scale score on other scales, they have to understand quite long and relatively complex instructions. The most obvious example is Letter-Number Sequencing, which has an 80-word set of instructions, which requires the client to understand what “in alphabetical order” means. Other subtests on both the WISC-III and the WAIS-III also seem to be relatively complex. For example on the WAIS-III Picture Completion has a twenty six word instruction which includes the word “important”, Vocabulary has a 38 word instruction, Similarities requires the client to understand what the word “alike” means. Picture Arrangement has 38 word instructions that require the client to understand the word “order”, and although a demonstration is allowed if the client fails on the first trials, this would result in them loosing a point, which could result in a reduction in a scale score.

The degree to which there are anomalies in item difficulty could be tested empirically if the assessments were given to a large

number of people with learning disabilities. However, until that work is done there must be some doubt as to the validity of both assessments at the lower IQ levels, due to inappropriate test items.

The Floor Effect

The WISC-III measures down to IQ 40 and the WAIS-III to IQ 45. However, both these floor IQs correspond to a scale score of 1 in all the subtests contributing to IQ. As a scale score of 1 is given even if the client obtains a raw score of zero, there is a potential for a floor effect that would result in people with very low “true IQs” (in the 30s) being given IQs in the 40s. To some extent the test designers recognise this as a problem, as both assessments state that a full scale IQ should not be given unless the client has scale scores above zero on at least 3 Verbal and 3 Performance subtests. However, there still is a logical problem with the current floors of the assessments.

It is not logical to give a scale score of 1 to a raw score of 0. A raw score of zero could imply an ability level just below that measured on the subtest; however, it could also imply ability well below this or no ability at all. There must be a point below which there is a raw score of 0. As it is not clear what ability level this corresponds to, except that it would less than 3.3 standard deviations below the norm, it not possible to allocate an IQ score to it. It is therefore not clear what raw score should correspond to a scale score of 1. Although a raw score of 1 may correspond to a scale score of 1 in some age ranges, there is no evidence to support this, it also would not be all age ranges, otherwise we would be saying that a raw score of 1 in a 6 year old was equivalent to a raw score of 1 in a 16 or 30 year old. Therefore, the lowest scale score we can confidently calculate from a raw score is scale score 2. This would mean that an IQ couldn’t be given unless a client has obtained at least a scale score of 2 in all the subtests contributing to IQ, which on both assessments gives an IQ of 49. However, if the client gets a scale score below 2 in any of these subtests then an IQ score cannot be properly calculated.

The Percentile Ranking

Both the WAIS-III and the WISC-III manuals give percentile ratings for IQ scores. These are based on a theoretical normal distribution of IQ with a mean of 100 and an SD of 15. An IQ of 70 is given a percentile ranking of 2 and an IQ of 60 a ranking

of 0.4, and IQs below 50 a ranking of <0.1. The problem with this is that at low IQs the normal distribution does not apply, there being far more people with severe and profound learning disability than would be predicted from the normal distribution. The prevalence of severe learning disabilities seems to be about 0.4% rather than <0.1. Abramowicz and Richardson (1975) reviewed epidemiological studies of severe learning disability in children and found the average and consistent rate of about 0.4% of the population, an average prevalence of 0.4% for people with severe learning disabilities (IQ<50) was also reported by Roeleveld et al (1997) in a more recent review. This means that true prevalence may well be that predicted by the normal curve plus .03%. While it may be reasonably accurate to quote a prevalence rate of 2% for IQ 70, the rate probably being about 2.6%, it is clearly inaccurate to quote a rate of 0.4% for IQ 60 when the true rate is 0.7% or <0.1% for IQ 50 when the true rate is probably about 0.4%.

Discussion

The issues raised in this paper have implications for the degree of confidence we can have with the results obtained from the WISC-III and WAIS-III at low IQs. These concerns, together with others raised elsewhere (Whitaker 2003; submitted), mean that psychologists should be cautious about quoting the IQ figures and other statistics from the manuals when they are assessing a client with a low IQ particularly if this is for diagnostic purposes. Clearly what is currently needed is further research on intelligence in people with learning disabilities so that we can be confident with item difficulty, the floor of the tests and the percentiles, and that all tests roughly agree.

References

Abramowicz, H.K. and Richardson, S.A. (1975)
Epidemiology of severe mental retardation in children: Community studies.
American Journal of Mental Deficiency, 80, 18-39

Roeleveld, N., Zielhuis, G.A. and Gabreels, F. (1997)
The prevalence of mental retardation: A critical review of recent literature.
Developmental Medicine and Child Neurology, 39, 125-132

Wechsler, D. (1992)
Wechsler Intelligence Scale for Children – Third Edition.
London: The Psychology Corporation Ltd

Wechsler, D. (1997)
WAIS-III, WMS-III, Technical Manual
San Antonio: The Psychological Corporation

Wechsler, D. (1998)
Wechsler Adult Intelligence Scale – Third Edition.
London: The Psychology Corporation Ltd

Whitaker (2003)
Should we abandon the concept of mild learning disability?
Clinical Psychology, 29, 16-19

Whitaker (submitted)
Is the WISC harder than the WAIS for people with learning disabilities?
Submitted to Clinical Psychology