The merits of mental age as an additional measure of intellectual ability in the low ability range

By

Simon Whitaker

It is argued that mental age may have some merit as a measure of intellectual ability, particularly in the low ability range, as it can be more accurately determined than a deviation derived IQ score.

The assessment of the intellectual ability of those at the low end of the ability range is done for both clinical and diagnostic reasons; however, current IQ tests may be subject to more error than has previously been acknowledged and it is questionable whether the tests should be used for diagnostic purposes (Whitaker, 2003, 2005, 2008). What will be argued in this paper is that the clinical and research utility of the tests may be improved if, in addition to IQ, a mental age (MA) is derived from the tests.

A child's MA is the age of the average child who would obtain the same score on a test of mental ability. The concept goes back at least as far as Binet and Simon's early work at the beginning of the last century and is based on the finding that as children get older their ability to do intellectual tasks increases. The original method of calculating IQ was based on the ratio between MA and chronological age (CA), as follows:

IQ = MAx100/CA

However, MA and the MA/CA ratio derived IQ have been criticised as measures of intellectual ability. One of the main critics was David Wechsler (1949, 1955) who pointed out that MA has little meaning for adults or older children with above average mental ages. Further, it would not be possible to calculate IQs for adults using the ratio of MA and CA unless an average adult MA was used in instead of the CA. This leads to the problem of what this age should be. Wechsler (1949) considered that this is a pointless question because if one defines adult mental age as the age at which mean scores no longer increase, this will be very dependent on the tests used, with some tests stopping at 12 years and others continuing to 25. Instead of using a ratio method for calculating IQ, Wechsler used a deviation method whereby an individual's IQ was determined by the degree to which they were less or more intelligent than people of their own age.

The Wechsler assessments have become the gold standard intellectual assessments. They have evolved over 69 years since the Wechsler Bellevue was first published in 1939 (Wechsler, 1939), are apparently well standardized and are probably the most widely used tests of child and adult intelligence. IQ is assessed by giving the client a number of subtests that measure different aspects of intellectual ability. Ten subtests are

used to measure Full Scale IQ (FS IQ) on the Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV; Wechsler, 2003) and 11 on the Wechsler Adult Intelligence Scale Third Edition (WAIS-III; Wechsler, 1997). The raw scores on these subtests are converted to a normalized scaled score, which has a mean of 10, a range between one and 19, and a standard deviation (SD) of three. The conversion of raw scores to scaled scores is based on the cumulative frequency of raw scores obtained by people of that age group in the standardization sample. The scaled scores are allocated to raw scores so that the proportion of the standardization sample gaining a particular scaled score fits the normal distribution with a mean of 10 and a standard deviation (SD) of 3. The sum of the scaled scores is then calculated for all the subtests used to measure IQ. The IQ figure is derived from this sum of scaled scores in a similar manner to the way that scaled scores were derived from raw scores, with a distribution of IQs with a mean of 100 and an SD of 15.

Deriving IQ in this manner clearly has a lot of advantages:

- The SD of the IQs and scaled scores can be set so that it is a convenient whole number and does not vary across age groups as was often the case with ratio derived scores.
- The IQ can be calculated for adults without having to derive an average adult MA.
- It is meaningful for more intelligent children where their MAs would be over 16.
- It takes account of different intellectual abilities, assessed in the different subtests, developing at different rates and continuing to develop to different ages.

However, there are some disadvantages and limitations to deviation derived IQ scores that are not normally acknowledged:

1. It is not an absolute measure of intellectual ability, but rather a rank ordering of people of the same age. As noted above, IQ derived by the deviation method is a measure of an individual's intellectual ability relative to people of their age. In contrast an absolute measure of intellectual ability would be a measure of an individual's intellectual ability relative to people of all ages. It is therefore not possible to use IQ scores to compare people of different ages in terms of their absolute intellectual ability. For example, consider two children, one aged 6 years and one age 16 years, both with measured IQs of 100. The IQ of 100 shows that each of them would be at the mid point of the range of intellectual abilities of children of their own age. However, to gain the IQ of 100 the 16-year-old would have had to have get far more items correct and gain higher raw scores on the sub-tests than the 6-year-old. The 16-year-old's absolute intellectual ability would clearly be higher than the 6-year-old's.

2. It is an ordinal measure. Deviation derived IQ is an ordinal measure of ability. This means that it can rank order people of the same age along a continuum of intellectual ability. It does not tell us the degree to which people of different IQs differ in intellectual ability. While it is perfectly correct to say that an individual with an IQ of 120 is more intelligent than an individual of the same age with an IQ of 60, it is not possible to say that they are twice as intelligent.

3. It may not be possible to measure deviation IQ reliably in the low ability range. Whitaker (2003, 2005, 2008) has pointed out that, when used with individuals who have low intellectual ability (IQs below about 80) the current commonly used IQ tests are likely to be subject to more error than has previously been acknowledged. This is best illustrated by considering the disparities between assessments when used with clients of low intellectual ability. Either the WISC or the WAIS can be used with adolescents of 16 years old. Whitaker (2008) and Whitaker and Wood (2008) examined the criteria for a 16-year-old gaining a scaled score of two on subtests common to the WISC-III, WISC-IV and the WAIS-III, and found evidence that the criteria were significantly more difficult on the WISC-III and WISC-IV than on the WAIS-III. If this is the case it would result in the WAIS-III giving higher IQ scores than either the WISC-III or WISC-IV when applied to clients with low intellectual ability. This suggestion is supported by an empirical study (Gordon, 2007; Gordon, Duff, Davidson & Whitaker, submitted) in which the WISC-IV and WAIS-III were both given to 17 sixteen-year-olds in special education in counterbalanced order. The mean FS IQ on the WISC-IV was 53.00 but 64.82 on the WAIS-III; a difference of 11.82 IQ points.

The reason for this disparity may stem from the way in which the assessments were standardised. Although standardised using relatively large stratified samples of the US population (2,450 for the WAIS-III, 2200 for the WISC-IV), the samples were then divided into sub samples of 200 subjects in specified age ranges. Effectively the standardisation was done separately for each age range. This meant that there were only five people below two SDs (IQ 70) and none below 2.8 SDs (IQ 58). The accuracy of the assessment of low IQ is dependent on how representative the subjects in the sample with low intellectual ability are of the people with low IQs generally. A failure to include sufficient people with low abilities would result in the assessment underscoring low IQs, while having too many people of low ability would result in the assessment over-scoring IQ at the low level. Therefore, in the IQ 55 to 70 ranges, there is potential for error due to inaccurate sampling. Below IQ 55, where there would not be expected to be any subjects in the sample, the relationships between raw scores and scaled scores and sums of scaled scores and IQs must have been based on extrapolation, which may not reflect the true relationship between test performance and IQ.

Given these limitations and problems with deviation derived IQ as a measure of low intellectual ability it may be worth examining if MA could offer any advantages as an additional measure of intellectual ability.

Derivation of mental age
If an individual has been given an intellectual assessment standardised on children, then their MA on individual subtests can be obtained by finding the age band for which the person's raw score is given a scaled score of 10 on the raw score to scaled score conversion tables. The overall MA could then be obtained by calculating the mean MA for all the subtests used to assess FS IQ. An alternative to finding the MA appropriate to a raw score in the "raw score to scaled score" tables would be to use Table A.9 (page 257) of the WISC-IV manual (Wechsler, 2004), which gives a broad conversion of raw score to mental age based on the US standardisation sample. If an MA is required for an adult

then it may be possible to give them the WISC-IV or alternatively give them the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). The WASI derives an IQ score from four subtests: Vocabulary, Similarities, Block Design and Matrix Reasoning and was standardized on a sample age between 6 years and 89 years so is suitable for both adults and children. Table A.7 (page 194) in the manual converts raw scores to age equivalents.

Advantages of mental age
There is a clear stipulation in the WISC-IV manual that the age-equivalents given in Table A.9 are provided for establishing comparisons with age norms of other scales and they should not be used for reporting purposes. However, if MA results are used with appropriate caution they may well give additional helpful information about an individual's intellectual ability. In a number of areas MA would seem to be a better measure than the deviation derived IQ scores.

- It is a measure that will compare across age groups provided the same measure has been used. An individual aged 30 years who has a MA of 10 years will have the same overall mental ability as a child of eight who gains an MA of 10 years. Although further research would have to be done on this it may well mean that MA would be a much better predictor of what an individual is able to do intellectually than IQ, for example predicting academic ability.

- It could be regarded as a ratio rather than ordinal measure of intellectual ability. This is because it is given in units of age, which is clearly a ratio scale. Therefore, it is reasonable to say that a person with an MA of 10 years has double the MA of somebody with an MA of 5 years. This may be useful in some circumstances, for example, if one wanted to look at the amount of change between groups. However, using MA as an interval measure should be done with caution as, although MA may increase evenly over time, actual mental ability does not. It is quite clear from looking at the raw score to MA conversion Table A.9 in the WISC-IV manual and Table A.7 in the WASI manual that, although the relationship between MA and raw score is apparently linear between MAs of 6 years and 10 years on most subtests, for MAs greater than 10 years there is a much slower increase in raw score, with MA suggesting that actual mental ability is not increasing as fast with age. This would mean that increases in MA over the age of ten are not indicative of equivalent increases in mental ability.

- MA is probably more accurately determined in the low ability range than is a deviation derived IQ. If, as suggested above, the inaccuracy in the deviation derived IQs in the low ability range is due to poor sampling of subjects with low ability, and inaccurate extrapolation of the relationship between raw scores and scaled scores, then a MA would be more accurate as these errors do not apply to its derivation. An individual's MA on a subtest is determined by finding the age group for which the raw score would be given a scaled score of 10. This is the same as the age group for which half the sample would have a raw score less

than that client's raw score and half the sample would score greater. There is therefore no extrapolation involved and it would not make any difference if there was poor sampling of people with low intellectual ability.

Issues to be resolved
From the above it seems that MA can be more accurately derived than a deviation IQ in the low ability range. Also between MAs of 6 years and 10 years it may be a reasonable measure of mental ability. It therefore could be used as a measure of intellectual ability in the low ability range for MAs between 6 years and 10 years. However, there are a number of issues where further work needs to be done in order to better understand the strengths and limitations of MA as a measure of mental ability. The issues are:
  • When does increase in ability begin to tail off with increase in age?
  • Which sub-tests are the best predictors of MA?
  • What is the level of accuracy of a measure of MA?
  • To what extent can MA predict ability in the real world as opposed to IQ?

Until these issues are resolved, any reporting of MA should be done with a cautionary statement saying that it is not clear how accurately MA can be obtained or the degree to which MA reflects mental ability.

Affiliation
South West Yorkshire Mental Health Trust and The University of Huddersfield

Address
The Learning Disability Research Unit, Room HW G/16, University of Huddersfield, HD1 3DH; s.whitaker@hud.ac.uk

References

Gordon, S. (2007). *Comparison of the WAIS-III and WISC-IV in 16 year olds who receive special education.* Unpublished doctorate thesis, Division of Clinical Psychology, University of Liverpool

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

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale.* New York: The Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children Manual.* New York: The Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Test Manual.* New York: The Psychological Corporation.

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

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence Manual.* London: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children – Fourth Edition: Technical and Interpretive Manual.* San Antonio, TX: The Psychological Corporation.

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

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

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

Whitaker, S. (2008). WISC-IV and Low IQ: review and comparison with the WAIS-III. *Educational Psychology in Practice*, 24, 129-137

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

Word count: 2698