Introduction

A necessary though not sufficient part of most internationally recognised definitions of Intellectual Disability (ID) is having a measured intellectual quotient (IQ) less than a critical figure, usually 70, for example the American Association on Intellectual and Developmental Disabilities (AAIDD, 2010), the World Health Organisation (WHO, 1996), and the American Psychiatric Association (APS, 2000). Measured IQ is therefore one of the major descriptors of people with ID in the scientific literature as well as being an important independent or dependent variable. However, recent work on the accuracy to which low IQ can be measured has suggested that degree of error is much greater than had previously been thought.

Error in the measurement of low IQ

It is acknowledged that IQ tests are not accurate to one point. This error can be divided in to two broad groups: chance and systematic. Chance error occurs when there are a number of small factors that affect the score in a positive or negative way. Although the score of an individual may be affected by chance error, the mean score of a group is affected considerably less, as the errors present for different individuals tend to cancel each other out. In the measurement of low IQ there are two major sources of chance error: that due to a lack of internal consistency and temporal error. Systematic error occurs when one assessment on average scores either higher or lower than another assessment.

The degree of chance error can be indicated by giving a range of scores around the measured IQ in which the individual's true IQ has a 95% chance of falling. This is the 95% confidence interval and is reported to be about five points either side of measured IQ on the best researched tests, such as the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV Wechsler, 2003), and The Wechsler Adult Intelligence Test – Fourth Edition (WAIS-IV Wechsler, 2008). However, the estimations of the 95% confidence interval are largely based on only one source of error, the lack of internal consistency of the test. It has been suggested (Whitaker, 2010a) that IQ assessment is subject to considerably more error than a lack of internal consistency and consequently the level of accuracy, particularly at the low IQ range, will be far less. This error should be acknowledged more explicitly. The other sources of error are temporal error (Whitaker, 2008), which is a further chance error, and three systematic errors: error that is apparent from differences between different tests (Gordon, 2007; Gordon, et al, 2010), error due to tests going out of date (Flynn, 2006a, b, Whitaker, 2010b) and the floor effect (Whitaker and Wood, 2008).

Temporal error. It is generally found that if the same individual is given the same assessment twice, the two IQ scores are different. There are a number of factors that may result in changes in scores between assessments, for example, the level of distraction in the test setting, the level of alertness of the subject or how the test was administered. A meta-analysis (Whitaker, 2008) of the test re-test reliability of IQ assessments in the low IQ range (IQ<80) found that although 57% of IQ changed by less than six points between assessments, 14% changed by 10 points or more. The 95% confidence interval calculated on the basis of the test re-test reliability figure was about 13 points.

A floor effect. Both the WISC-IV and WAIS-IV measure IQ by giving the client a number of subtests measuring different aspects of intellectual ability. In order to calculate the Full Scale IQ (FS IQ) the raw scores on these subtests are converted to scaled scores with a mean of 10, a standard deviation (SD) of three, and a range between one and 19. It has been suggested (Whitaker, 2005; and Whitaker and Wood, 2008) that allocating a scaled score of one to low raw scores or a raw score of zero could result in an overestimate of intellectual ability.

The Flynn Effect. There is good evidence that the intellectual ability of the population as a whole has increased from one generation to the next at about 0.3 of an IQ point per year over the last 100 or so years. Not only has this occurred for the population in general (Flynn, 1984; 1987; 2006a), but also for those with low intellectual ability (Flynn, 1985, 2009). The implication of this for the assessment of IQ is that tests will become less accurate the longer it is since they were standardised. On average a test will overestimate an individual's IQ by about 0.3 IQ for each year since it was standardised. It therefore may be possible to compensate for this error by subtracting 0.3 of an IQ point from the measured IQ for each year between the test being standardised and given. However, although it has been argued that is still a valid correction in the US (Flynn, 2009), the rate of increase in intellectual ability is not consistent over time or between tests (Flynn, 2006b), and there is now evidence that in Scandinavia (Teasdale and Owen, 2005) and in the UK (Whitaker, 2010b) that the Flynn effect may have gone into reverse in the low IQ range. This uncertainty of the degree of the Flynn effect outside the US means that IQ scores cannot be corrected with confidence and that in effect there is additional chance error in the scores of the order of about 0.3 of a point per year (c.f. Whitaker, 2010a).

Error apparent from the differences between IQ scales. It is accepted that different IQ tests will give slightly different results (Floyd, et al 2008). It is also possible that these differences between tests may be greater at the low IQ range, as the tests were standardised in the main using subjects in the average intellectual ability range. An important indicator as to how much this is a problem would be given by the degree to which the two gold standard IQ tests, the WISC and the WAIS, agree with each other. Spitz (1989) showed that in the low range the WAIS-R systematically measured higher that the WISC-R and that this difference was greater for lower IQs, so that for WISC-R IQs in the 40s the WAIS-R IQ was some 15 points higher but for WISC-R IQs in the 70s the equivalent WAIS-R IQ was about 4 points higher. Gordon, et al (2010) compared the WISC-IV (UK) and the WAIS-III (UK) with 16-year-olds in special education. It was found that although there was a high correlation between the two assessments (r=.93), in each case the Full Scale IQ (FS IQ) on the WISC-IV (UK) was less than that on the WAIS-III (UK). A mean FS IQ of 53.00 was found on the WISC-IV (UK) which compared to a mean of 64.82 on the WAIS-III (UK), a difference of just less than 12 IQ points. The reason for this difference is probably multi-factorial. About 2 points may be due to the Flynn effect with the WAIS-III being standardised six years before the WISC-IV. It has also been suggested (Flynn 2009) that the WAIS-III increased scores more than would be expected by the Flynn effect by about further 2.5 points. However, the WISC-IV is subject to a greater floor effect than the WAIS-III (Whitaker and Gordon, submitted) which would decreased the difference between the tests. It is therefore unclear what factors account for about 8 points of the 12 point difference between the tests, though problems in the standardisation specifically in the low IQ range may be a factor (Whitaker, 2010a), as may the differences in test content (Flynn, 2009). What is clear is that either one or both of these assessments are failing to produce an accurate measure of an individual's true IQ. As the degree to which either assessment is in error is largely not known, it is clearly possible that either the WISC-IV (UK) is systematically underestimating true IQ by up to the order of 12 points, or the WAIS-III (UK) is systematically overestimating true IQ by the order of 12 points or both assessments are making systematic errors of less than about 12 points.

Whitaker, (2010a) combined these sources of error and calculated an estimate of the 95% confidence intervals on both the WISC-IV and WAIS-III by calculating a combined 95% confidence interval, which took into account error specifically in the low range, from both a lack of internal consistency and temporal error and added to this uncertainty due to the Flynn effect, the floor effect and the remaining systematic error apparent from the difference between the WISC-IV and WAIS-III. For the WISC-IV there is an effective confidence interval which extends 16 points below the measured IQ and 25 points above it. For the WAIS-III the effective confidence interval extends 18 points above the measured IQ and 28 points below. Whitaker, (2010a) focused on the WISC-IV and WAIS-III not because they were more likely to be subject to error than other tests but because there was far more data available to assess their accuracy than other tests. There is no reason to suppose that other intellectual assessments are not subject to at least as much error as the WISC-IV and WAIS-III.

There therefore seems to be considerably more error in the measurement of low IQ than is usually acknowledged. It may be possible to correct for some of the errors if information is available with regard to the test used, when it was given, and the raw scores obtained; notably the Flynn effect (Flynn, 2006a) and the floor effect (Whitaker and Gordon, submitted). Also the mean IQ of a group would not be subject to nearly as much chance error as would an individual IQ score. However, there is currently no way of correcting for a lot of the error that is apparent from the difference between tests. Therefore, it cannot be assumed that IQ scores obtained on different tests are equivalent.

The aim of this study is to look at the current scientific literature on ID to ascertain the extent to which the concept of intelligence and IQ is used and where it is used, how IQ was measured and reported, and specifically if IQs obtained on different tests are regarded as being equivalent. Since it would not be feasible to examine all recent papers on ID it was decided to focus on two prominent scientific journals for one year (2008): the Journal of Applied Research in Intellectual Disabilities from the UK and the American Journal on Mental Retardation, from the US.

Method

All papers in the 2008 volumes of the American Journal on Mental Retardation (AJMR) and the Journal of Applied Research in Intellectual Disability (JARID) were read. Information was extracted on how the concept of intellectual ability was used and how IQ scores were reported.

Results

There were 91 papers in the two journals, and of these 81 used the concept of intelligence. Several of the papers that used the concept of intelligence did so without reporting IQ scores. Of those studies that did report IQ scores there was a varying amount of information provided as to how the IQ was obtained or the assessments used. When IQ tests were named there was a variation in the degree of information provided about that test. In some papers the test was referenced, in others only the name of the test was given. In reporting the results here, a reference to the test used will be given if that was done in the study that is being referred to.

A number of the studies used a single IQ test and stated which test it was. For example, MacMahon and Jahoda, (2008) used the WASI (Wechsler, 1999), Mayes et al, (2008), Mildon et al, (2008) and Weiss, (2008) used the Kaufman Brief Intelligence Test (K-BIT) (Kaufman and Kaufman, 1990) though only Mayes et al, (2008) reference it. Other studies were less precise as to what test they used or seemed to treat different tests as giving equivalent results.

Studies that used various IQ assessments (tests treated as equivalent). Nine of the studies used different IQ assessments to determine the participants' IQs. Of these, four acknowledged that the results of the different tests may not be equivalent in some respect. Krinsky-McHale et al, (2008) acknowledged that scores on the WAIS-R may not be equivalent to those on the Stanford-Binet or the Leiter International Performance Scale. Whitaker and Wood, (2008) compared the floor effect on the WISC-III and the WAIS-III and found it was significantly greater on the WISC-III than the WAIS-III. In interpreting this result they suggest that the WISC-III may systematically give lower FS IQ at the low range than the WAIS-III. Harris et al, (2008) assessed the intellectual ability of people with Fragile-X syndrome using the WPPSI-III, WISC-III and the WAIS-III or the Kaufman Assessment Battery for

Children (K-ABC) depending on their age. There therefore seemed to be an assumption that these tests were equivalent, however, it is stated: “Although these cognitive measures are used routinely for individuals with Fragile X syndrome, to our knowledge there have been no validity or reliability studies done regarding their utilization in this population.” (p. 430). Wheeler et al, (2008) assessed boys with fragile X on either the Mullen Scale of Early Learning (Mullen, 1995) or the Leiter International Performance Scale-Revised (Leiter-R) (Roid and Miller, 1997) depending on age. It was noted that the Mullen Scale of Early Learning was designed to measure development and the Leiter-R was developed to assess nonverbal IQ, hence scores on these two measures were not comparable. With the other studies there was no suggestion that the different tests used were not producing equivalent results.

Studies that specify IQ scores without stating the assessment. A number of studies stated IQ scores without saying how they were obtained. For example, Melville et al, (2008) simply state that “the primary source of level of intellectual disability was taken from documented intelligence quotient (IQ) test and Vineland Scale (survey form) assessments.” (p. 427). Carr, (2008) reports on a follow up of a cohort of people with Down’s syndrome who were born between December 1963 and November 1964. She cites the mean and range of IQs but does not make it clear how or when these IQs were obtained. In addition, three literature reviews (Jahoda et al, 2008, Wade et al, 2008), and the International Association for the Scientific Study of Intellectual: Disability Special Interest Research Group, IASSID SIRG, 2008) cite IQs without indicating the assessments that were used.

Measuring very low IQ. A further issue that emerges from some of these studies is that of the measurement of very low IQ. A number of studies report IQs well below the floor level of IQ 40 on the WISC-IV and WAIS-IV but do not say how these IQs were measured. Hickson et al, (2008) used selection criteria including an IQ in the range 35 to 75 but without giving any indication as to how the IQ scores were obtained. Urv et al, (2008) cite mean IQs of**:** 32.47, 30.94, 32.36 and 31.42 respectively for people with Down's syndrome in four different stages of dementia but do not say how the IQs were obtained. Thompson et al, (2008) investigated the inter-rater reliability of the Supports Intensity Scale (Thompson et al, 2004), which assesses the support needs of clients with ID. The participants were grouped according to the following IQ ranges: 51-69, 36-50, 20-35 and <20. However there was no indication as to how their IQs were assessed.

Discussion

It is apparent from this review that the concept of intelligence is widely used within the scientific literature on ID, with 81 of the 91 papers in the 2008 volumes of JARID and AJMR making reference to it. Measured IQ is also frequently reported. A few studies indicated that there may be some errors in the measurement of low IQ and that different intellectual assessments may not be equivalent to one another. However, to a large extent, when IQ is cited in papers there is little or no indication as to the accuracy of the assessments. There seems to be an implicit assumption that an IQ assessment done with a particular test under a particular set of conditions is equivalent to another assessment done with a different test under a different set of conditions.

Possibly one of the reasons for this apparent assumption as to the accuracy of IQ assessments is that researchers believe that different IQ and adaptive behaviour tests are equivalent and can measure true IQ to within 5 points. However, the evidence now suggests that this is not the case; well standardised tests such as the WISC-IV and WAIS-IV may only measure to an accuracy of 15 points either side of the measured IQ and vary systematically from each other by about 10 points (Whitaker 2010a). We do not have the data as to the accuracy of other assessments in the low range but there is no reason to suppose that they would be any better.

The failure to take this lack of accuracy in the measurement of low IQ into account could have varying impacts on studies, and may on some occasions lead to false conclusions being drawn. The chance and systematic errors apparent in the measurement of low IQ will have different effects. If a study presents the mean IQ of a group of participants then the chance error will be greatly reduced as the errors will be cancelled out by averaging the scores. However, a mean score will be still subject to systematic error. If on the other hand the study makes used of IQ in correlations then the systematic error will not affect the results, however, correlations and tests of statistical significance would be affected by the chance error. It may be useful to consider some examples in more detail.

The greatest impact of the error will be in the assessment of the IQ of a single individual, for example when reporting on a case study, as this will be subject to both chance and systematic errors and so true IQ may vary by up to 25 points either side of the measured IQ. If there is a large disparity between measured and true IQ the study could give the false impression that the factor under investigation had a particular effect with individuals at that IQ level. The mean IQ of groups may well be more accurate as the chance errors will tend to be cancelled out, but they will still be subject to systematic error such as that evident from the difference between tests, the floor effect, and the Flynn effect.

Chance error will cause a particular problem if focus of the study is a correlation between IQ and another variable. Correlation between measured IQ and variables that are dependent on true IQ will be less than they would be if we were able to measure IQ without error. This could lead to small but nonetheless theoretically important effects of intelligence being missed. For example, a position paper on parenting by people with ID (IASSID SIRG, 2008) stated that there is no relationship between IQ (in the 60s) and parenting ability. It is possible that researchers have failed to find a relationship between IQs in the 60s and parenting ability, not because one does not exist**,** but because the error in measuring IQ at this level has reduced the correlation. Whitaker, (2010a) reports that the overall reliability figure, taking into account both temporal error and error due to a lack of internal consistency, is .74. Therefore, one would expect that a correlation between a variable and measured IQ would be .74 of the correlation between that variable and true IQ. It is therefore possible that one reason why a significant correlation between IQ and parenting has not been found is due to the correlations being made between true measured IQ and not true IQ.

A failure to recognise that there may be difference between tests due to systematic error may lead to wrong conclusions being drawn. A study by Russell et al, (1997) is an example of a study where this may have happened. Russell et al, (1997) investigated whether schizophrenia reduced IQ in the UK. They compared the IQs of adults who had developed schizophrenia with their IQs as children, before developing schizophrenia. However, as children they were mainly assessed on the WISC-R, though some were assessed on the WISC and as adults they were assessed on the WAIS-R. They reported the mean WISC-R/WISC IQ to be 84.1 and the mean WAIS-R IQ to be 82.2 and conclude that schizophrenia did not result in a significant reduction in IQ. They make the point that previous studies claiming schizophrenia reduced IQ had been flawed and that it was just that schizophrenics had a lower pre morbid IQ. However, they fail to consider that the WISC-R may systematically measure lower than the WAIS-R at these IQ levels which is probable for two reasons. First, it is likely that the scores on the WAIS-R would have been elevated far more by the Flynn effect than the WISC-R they took as children. This is because the majority of the assessments on the WISC-R occurred in the mid 1970s, soon after the UK version of the test came out, so would have been subject to about a one point Flynn effect. As the paper was published in 1997 and the WAIS-R assessment was done on adults, one assumes these assessments took place in the mid 1990s, about 18 years after the test was normed, meaning it would be subject to about a 6 point Flynn effect. The paper does not report how many participants were assessed on the WISC as children, which would have been subject to a Flynn effect of about 10 points. However, assuming that only a small proportion of the children were given the WISC, it is likely that the WAIS-R scores were systematically elevated by the Flynn effect on average by about 4 points above the WISC and WISC-R assessments. Secondly, there is the additional systematic difference of about 4 points between the WISC-R and WAIS-R found by Spitz, (1989) at this IQ level. Therefore, in order to get an estimate of true change in IQs between the participants when they were children and later as adults, one need to subtract from the WAIS-R score 4 points for the Flynn effect and 4 additional points for the measure being done in the low range. Therefore, with the 1.9 points difference that was found in the measured scores in the study, it would appear that as adults the participants were effectively scoring about 10 IQ points lower than they did as children. This may well be seen as good evidence that schizophrenia does reduce IQ.

Recommendations

Now that it is apparent that there are major problems in the measurement of low intellectual ability, the scientific literature needs to take account of them. At a minimum, if assessments are referred to in studies, the assessment used should be named and referenced. It would also be helpful for authors to make it clear to the readers that measured IQ may differ considerably from true intellectual ability and that results of studies may be affected by this.

If information is available as to reliability and validity of the assessment used, specifically in the low range, this could be referred to and referenced.

Giving more details as to how the tests were carried out could help a reader make allowances for some of the errors and evaluate how accurate the assessment may be. Authors therefore may wish to state how long it was between the test being standardised and being given, whether there was a floor effect and whether the conditions under which the assessments took place were known to be optimal. If scores can be corrected it may be helpful for the author to do so and possibly state both corrected and uncorrected scores.

Summary

Recent work has shown that there is more error in the measurement of low IQ than has previously been acknowledged, notably: poor stability, a floor effect, the Flynn effect and large disparities between different tests. In order to assess the degree to which these errors impact on scientific studies all the papers in two leading journals on intellectual disability, the *Journal of Applied Research on Intellectual Disability* and the *American Journal on Mental Retardation,* for the year 2008, were read and use of intellectual assessments noted. It was found that the majority of the papers referred to the concept of intelligence. However, only a few papers acknowledged that there may be some additional error in its measurement. Several papers simply cited IQ scores without any further information as to how they were obtained. It is argued that there is a need for published studies to give more information as to how intellectual ability is assessed, to acknowledge the errors that may occur, and to consider how such errors may have affected their results.

Reference

American Association on Intellectual and Developmental Disabilities. (2010) *Intellectual Disability: Definition, classification and systems of supports* (11th ed.). Washington DC: AAIDD.

American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision.* Washington DC: American Psychiatric Association.

Carr, J. (2008) The everyday life of adults with Down syndrome. *Journal of Applied Research in Intellectual Disabilities*: 21, 389-397.

Floyd, R.G., Clark, M.H. & Shadish, W.R. (2008) The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice*, 39, 414-423.

Flynn, J.R. (1984) The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.

Flynn, J.R. (1985) Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency*, 90, 236-244.

Flynn, J.R. (1987) Massive gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.

Flynn, J.R. (2006a) Efeito Flynn: Repensoado a inteligencia e seus efecitos [The Flynn Effect: Rethinking intelligence and what affects it]. In C Florres-Mendoza & R Colon (eds.) *Introducao a Psicologia das Diferencas Individuais [Introduction to the psychology of individual differences]* (pp. 387-411). Porto Alegre: ArtMed. (English trans: from jim.flynn@stonebow.otago.ac.nz).

Flynn, J.R. (2006b) Tethering the elephant capital cases, IQ and the Flynn Effect. *Psychology, Public Policy and Law*, 12, 170-189.

Flynn, J.R. (2009) The WAIS-III and WAIS-IV: Daubert motions favor the certainly false over the approximately true. *Applied Neurology*, 16, 98-104.

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

Gordon, S., Duff, S. Davison, T & Whitaker, S. (2010) Comparison of the WAIS-III and WISC-IV in 16 year old special education students. *Journal of Applied Research in Intellectual Disability*.23, 197-200.

Harris, S.W., Hessl, D., Goodlin-Jones, B., Ferranti, J., Bacalman, S., Barbato, I., Tassone, F., Hagerman, P.J., Herman, K. & Hagerman, R.J. (2008) Autism profile of males with Fragile X syndrome. *American Journal on Mental Retardation,* 113, 427-438.

Hickson, L., Khemka, I., Golden H., & Chatzistyli A. (2008) Profiles of women who have mental retardation with and without a documented history of abuse. *American Journal on Mental Retardation*: 113, 133-142.

IASSID SIRG. (2008) Parents labelled with intellectual disability: Position of the IASSID SIRG on parents and parenting with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*: 21, 296-307.

Jahoda A., Kemp J., Riddell S. & Banks P. (2008) Feelings About Work: A review of the Socio-emotional Impact of Supported Employment on People with Intellectual Disabilities' *Journal of Applied Research in Intellectual Disabilities*: 21, 1-18.

Kaufman, A.S. & Kaufman, N.L. (1990) *Kaufman Brief Intelligence Test*. American Guidance Services, Circle Press.

Krinsky-McHale, S. J., Devenny D.A., Kittler P. & Silverman W. (2008). Selective attention deficits associated with mild cognitive impairment and early stage Alzheimer's disease in adults with Down syndrome. *American Journal on Mental Retardation*: 113, 369-386.

MacMahon, P. & Jahoda, A. (2008) Social comparison and depression: People with mild and moderate intellectual disabilities. *American Journal on Mental Retardation*: 113, 307-318.

Mayes, R., Llewellyn, G., & McConnell, D. (2008) Active negotiation: Mothers with intellectual disabilities creating their social support networks. *Journal of Applied Research in Intellectual Disabilities*: 21, 341-350.

Melville, C.A., Cooper S-A., Morrison J., Allan L., Smiley E. & Williamson A. (2008) The prevalence and determinants of obesity in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*: 21, 425-437.

Mildon, R., Wade, C., & Matthews, J. (2008) Considering the contextual fit of an intervention for families headed by parents with an intellectual disability: An exploratory study. *Journal of Applied Research in Intellectual Disabilities*: 21, 377-387.

Mullen E.M. (1995) *Mullen Scale for Early Learning*. Circle Pines, MN: American Guidance Service.

Roid, D. and Miller, L. (1997). *Leiter International Perforamance Scale-Revised*. Wood Dale, IL: Stoelting.

Russell, A.J., Munro, J.C., Jones, P.B., Hemsley, D.R., & Murray, R.M. (1997) Schizophrenia and the myth of intellectual decline. *American Journal of Psychiatry*, 154, 635-639.

Spitz, H.H. (1989).Variations in the Wechsler interscale IQ disparities at different levels of IQ. Intelligence, 13, 157-167.

Teasdale, T.W. & Owen, D.R. (2005) A long-term rise and recent decline in intellectual test performance: the Flynn Effect in reverse. *Personality and Individual Differences*, 39, 837-843.

Thompson, J. R., Bryant, B. R., Campbell, E. M., Craig, E. M., Hughes, C. M., Rotholz, D. A., Shalock, R., Silverman, W., Tasse, M., & Wehmeyer, M. (2004) *Supports Intensity Scale* (SIS). Washington, DC: American Association on Mental Retardation.

Thompson, J. R., Tasse, M. J., & McLaughlin, C. A. (2008).Interrater reliability of the supports intensity scale (SIS). *American Journal on Mental Retardation*: 113, 231-237.

Urv, T. K., Zigman, W. B., & Silverman, W. (2008). Maladaptive behaviours related to dementia status in adults with Down syndrome. *American Journal on Mental Retardation*: 113, 73-86.

Wade, C., Llewellyn, G., & Matthews, J. (2008) Review of parent training interventions for parents with intellectual disability. *Journal of Applied Research in Intellectual Disabilities*: 21, 351-366.

Wechsler, D. (1999) *Wechsler Abbreviated Scale of Intelligence (WASI).* San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (2008) *Wechsler Intelligence Scale for Children – Fourth Edition: Technical and Interactive Manual*. San Antonio, TX: The Psychological Corporation.

Weiss, J.A. (2008) Role of Special Olympics for mothers of adults athletes with intellectual disabilities. *American Journal on Mental Retardation*, 113, 241-253.

Wheeler, A. C., Skinner, D. G., & Bailey, D. B. (2008) Perceived quality of life in mothers of children with fragile X syndrome. *American Journal on Mental Retardation*: 113. 159-177.

Whitaker, S. (2005) The use of the WISC-III and the WAIS-III with people with a learning disability: Three concerns. *Clinical Psychology*, 50, 39-40.

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. (2010a) 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.

Whitaker, S. (2010b) Are people with intellectual disabilities getting more or less intelligent? *British Journal of Developmental Disabilities*, 56, 49-55.

Whitaker, S. & Gordon, S. (submitted) Floor effects on the WISC-IV and WAIS-III. Submitted to Journal of Applied Research in Intellectual Disabilities.

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.

World Health Organisation (1996) *Diagnostic and Management Guidelines for Mental Disorders in Primary Care: ICD-10 Primary Care Version*.
Seattle: Hogrefe & Huber Publishers.