Journal of Applied Research in Intellectual Disabilities 2007, **20**, 323–330

A Comparison of WAIS-R and WAIS-III in the Lower IQ Range: Implications for Learning Disability Diagnosis

Suzanne Fitzgerald, Nicola S. Gray and Robert J. Snowden
School of Psychology, Cardiff University, Cardiff, UK

Accepted for publication *20 September 2006*

Background Whether the Flynn effect (the increase in the populations' IQ over time) affects the IQ scores of people with learning disability or borderline learning disability remains unclear. The issue is important as the Flynn effect should alter the number of people eligible for health service resources. A comparison of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) with the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) in individuals with learning disability or borderline learning disability was conducted.
Method All participants completed the WAIS-R and the WAIS-III. Discrepancy scores were calculated for the Full Scale IQ score and the Verbal and Performance subscale scores.
Results WAIS-III Full Scale scores were significantly lower by over 4 IQ points. Verbal and Performance Scale IQ scores were also significantly lower than the corresponding WAIS-R scores.
Conclusion The shift from WAIS-R to WAIS-III means that 66% more people meet criterion A for the diagnosis of learning disability and hence this has major resource implications for health service providers.

Keywords: Flynn effect, learning disability diagnosis

Introduction

The populations' intelligence increases by roughly 0.3 IQ points each year, a phenomenon known as the Flynn effect (Flynn 1984). To counteract this effect, IQ tests are occasionally re-normed to the current population. Consequently, a person's IQ score will be lower when tested on the updated version of the test, altering the percentage of people with a Full Scale IQ score below 70 (Flynn 2000). The introduction of the newest version in the Wechsler Adult Intelligence Scale (WAIS) series, the WAIS-Third Edition (WAIS-III; Wechsler 1997) 16 years after the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler 1981) would be expected to decrease participants Full Scale IQ score by 4.8 points.

Wechsler (1997) investigated the Flynn effect and compared a sample of 192 adults aged 16–74 years (mean age 43.5 years; SD 20.2 years) on the WAIS-III and the previous test version, the WAIS-R, in a counterbalanced order. The interval between testing was 2–12 weeks, the median being 4.7 weeks. The WAIS-R was found to be 2.9 Full Scale IQ points higher than the WAIS-III, the WAIS-R Verbal Scale 1.2 IQ points higher and the WAIS-R Performance Scale 4.8 IQ points higher. The difference that Wechsler reported is therefore less than that would be expected from the Flynn effect. This is the only study to have investigated the Flynn effect on IQ scores as measured by the WAIS-R and the newer WAIS-III, therefore it is not possible to directly compare the Wechsler data to any other data in order to consider reasons for this seemingly low affect of the Flynn effect on Full Scale IQ scores.

The Flynn effect was originally described in people of around average intelligence. Some researchers have found that the Flynn effect is uniform across the borderline (Full Scale IQ 70–80) and mental retardation (Full Scale IQ <70) intellectual ability levels (Kanaya *et al.* 2001; Truscott & Frank 2001), but others have found it to be smaller in those with lower IQ, and that it can even reverse for those with scores in the learning disability range (Spruill & Beck 1988; Spitz 1989).

Truscott & Frank (2001) analysed the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler 1974) and the more recent Wechsler Intelligence Scale

10.1111/j.1468-3148.2006.00349.x

for Children-Third Edition (WISC-III; Wechsler 1991) scores for a sample of children receiving special education because of learning difficulties ($n = 171$). However, only children with an IQ of 80 or more were included in the sample, as Spitz (1989) reported an atypical effect of the Flynn effect in children with mental retardation (children with a diagnosis of mental retardation typically have a Full Scale IQ of <70). Therefore, the sample is only representative of children with intelligence in the borderline intelligence range and above. A significant difference between test scores of −4.77 for Full Scale IQ, −2.93 for Verbal Scale IQ and −6.68 for Performance Scale IQ was found. The Flynn effect was present across all three subscales of the WISC but was not uniform, with an inflated effect on the Performance Subscale.

In the Truscott & Frank (2001) study a large gap between test administrations means that the participants' history was not controlled for [the average time between test administration was 3.32 years (SD, 0.99)]. As the authors suggest, it may have been beneficial to have administered the WISC-R and the WISC-III at approximately the same time to ensure that participants' histories did not differentially affect performance on the IQ tests. Events in the 3-year gap between administrations could have accounted for some of the difference in test score, and so less of the variance is likely to be accounted for by the Flynn effect.

Some studies suggest that the Flynn effect is not uniform across IQ levels, but is conditional on IQ. Spitz (1989) analysed the literature that had tested participants on the original WAIS and the WAIS-R (the WAIS-R was the first revision of the WAIS, which has now been superseded by the WAIS-III) and reported that people within the average intelligence range (an IQ of 90–110) are affected by the Flynn effect consistently with what Flynn (1984, 1987) described. It would be expected that an individual performing at the same level would not achieve the same score on the WAIS-R, but a lower score, as his/her performance is compared with normative data of a higher standard than those published with the WAIS. However, for participants who scored in the mental retardation range (an IQ of <70) the effect of the Flynn effect upon IQ scores was not consistent with what Flynn described; IQ scores did not drop with the introduction of the WAIS-R, but the WAIS-R produced higher scores than the previous test version (WAIS). Four of the five samples that tested participants with Full Scale IQ scores within the mental retardation range reported a negative difference between the WAIS and WAIS-R (i.e. the WAIS-R was bigger than the WAIS).

The samples in Spitz's (1989) study were taken from across 13 studies ($n = 784$, 15 samples) that looked at the effect of the Flynn effect on the WAIS and WAIS-R scores with varying IQ levels. Spitz (1989) analysed all the 15 samples in one regression and it was found that the Flynn effect varied as a function of IQ. The WAIS was 9.09 Full Scale IQ points higher than the WAIS-R in the average intelligence range and 4.75 Full Scale IQ points higher in the borderline intelligence range. However, the WAIS-R was 1.26 Full Scale IQ points higher in the mild mental retardation range and was 5.84 points higher in the moderate mental retardation range. The samples that were representative of the mental retardation population in the Spitz (1989) analysis tended to be those that administered the tests years apart. As with Truscott & Frank (2001), the large time gap between test administrations may have increased the error variance and hence the difference between IQ tests may be unduly influenced by factors other than the Flynn effect. Furthermore, the combination of samples in the analysis may have increased the amount of variance due to sample or design factors, rather than due to differences in IQ scores.

Spruill & Beck (1988) produced the same findings reported by Spitz (1989) in a study design that did not have the same methodological limitations. WAIS and WAIS-R were concurrently administered to a sample of students of varying IQ levels applying for special education services ($n = 108$). They found that the WAIS-R was significantly lower than the WAIS, by 1.90 Full Scale IQ points, 2.33 Verbal Scale IQ points and 1.20 Performance Scale IQ points, but again this was not consistent across IQ levels. The difference between the WAIS versions varied across the IQ range. In the average and borderline intelligence ranges, the WAIS was significantly higher than the WAIS-R (7.92 Full Scale IQ points, 9.13 Verbal Scale IQ points and 9.09 Performance Scale IQ points; and 3.46 Full Scale IQ points, 7.06 Verbal Scale IQ points and 4.75 Performance Scale IQ points respectively). However, in the mild and moderate mental retardation ranges the WAIS-R was significantly higher than the WAIS [−1.20 Full Scale IQ points, −0.50 Verbal Scale IQ points (not significant) and −1.26 Performance Scale IQ points and −5.72 Full Scale IQ points, −8.80 Verbal Scale IQ points and −5.84 Performance Scale IQ points respectively].

The conflicting evidence available for revision of the WISC and the WAIS in samples with intelligence in the lower IQ range means that the applicability of the Flynn effect specifically to those with learning disability or borderline learning disability remains unclear. Furthermore, no study to date has investigated any changes to individual IQ scores in those in the lower IQ range, as a result of the introduction of the WAIS-III in its revision from the WAIS-R.

The Ability to Accurately Diagnose Learning Disability

In the diagnosis of mental retardation, criterion A of DSM-IV-TR (American Psychiatric Association 2004) states 'The essential feature … is significantly sub-average general intellectual functioning' (defined by an IQ of about 70 or below). Likewise, one of the ICD-10's (World Health Organisation 1992) criteria for learning disability is an IQ score below 70. The 'real world' importance of the affect of the Flynn effect is also highlighted by Kanaya *et al.* (2001) who conducted an extensive archival analysis of children tested for eligibility for special education services. They recorded WISC-R and WISC-III scores as well as psychologist recommendations for placement ($n = 8944$). They reported that the Full Scale IQ score from the WISC-R to the WISC-III significantly dropped by 5.6 points. Consequently, those children tested on the WISC-R followed by the WISC-III had a significantly increased percentage recommended for mental retardation classification (compared with those tested on the WISC-R at both administrations or the WISC-III at both administrations). Furthermore, children who had equivalent cognitive abilities (those who had a Full Scale score between 71–75 on the WISC-R and 66–70 on the WISC-III) were recommended for different classifications. Recommendation for a diagnosis of mental retardation was three times as likely if the WISC-III had been administered.

Kanaya *et al.* (2001) concluded that the diagnosis of mental retardation is unreliable as different psychologists use different versions of IQ tests and comparisons should not be made across test versions. Furthermore, if the same test version is administered at different evaluations, the reliability of the norms will vary depending on the age of the norms (as per the Flynn effect described above). It is important to note that the results were based on psychologists' recommendations for placements, and not actual classifications. In addition, the outcome measure of recommendation for classification is possibly confounded because the psychologist's judgement would also be based upon a judgement and/or a measure of adaptive behaviour. This was not measured by Kanaya *et al.* (2001). Nonetheless, the study suitably highlights the direct application of the Flynn effect to the ability of clinicians to accurately diagnose learning disability. Thus the re-norming of intelligence tests to alleviate the Flynn effect, and possible subsequent changes in IQ score could have significant clinical and resource implications for health services. The aim of the present study was to investigate any effect upon IQ scores because of the revision of the WAIS-R to the WAIS-III.

Method

Sample characteristics

A total of 66 service users from Craegmoor Healthcare Services Ltd (Craegmoor) were approached to take part in the study. Craegmoor is an independent sector facility that provides support to individuals within the social and health care sectors at various levels of security. The specialist learning disability services at Craegmoor were targeted in the attempt to test a sample with a range of intellectual abilities at the lower end of the IQ range (with learning disability or borderline learning disability). Those who were actively psychotic or risk assessed as inappropriate to take part at the time of testing, were not included in the study. Nine individuals were removed from the sample pool on this basis. In addition it was decided for six people that their intellectual impairment was so severe that it would not be possible for them to sustain the attention required to complete the IQ tests (each test took approximately $1\frac{1}{2}$ h to complete). Therefore the sample did not include those who were mentally ill or had severe impairment of intellectual functioning and so it is not entirely representative of all service users at Craegmoor. Furthermore, six service users declined to take part in the study and one withdrew after completion of one of the IQ tests.

The final sample consisted of 45 participants (29 men and 16 women). The mean age of participants was 32 years (SD = 9.14 years, range 18–50 years). Thirty-eight (84.4%) of the participants that took part in the study had a diagnosis of learning disability and five (11.1%) had a diagnosis of borderline learning disability. Diagnoses made by a qualified clinician were taken from patient files. Two of the participants (4.4%) had ability in the low average intelligence range, but their co-morbid diagnosis of Autism or Asperger's syndrome warranted institutionalization in a secure provision or in 24-h staffed accommodation.

Comparisons between the WAIS-R and the WAIS-III are limited by their different scoring criteria

The WAIS-III introduces a number of changes to the test compared with the previous version. First, the compulsory subtests required to calculate IQ score differs between the two test versions, with the Matrix Reasoning subtest replacing the Object Assembly subtest. For the remaining subtests, there is a different item structure and so it is not possible to directly compare raw scores or scaled scores across tests.

The scoring criteria also differ between the two IQ tests in terms of when age-based norms are used. The WAIS-R raw scores for each subtest are converted into scaled scores using non-age-related norms, and then these scaled scores are converted to IQ scores (using age-related norms). Conversely, the WAIS-III raw scores for each subtest are converted into scaled scores using age-related norms, and then these scaled scores are converted to IQ scores (using non-age-related norms). Therefore only final IQ scores can be compared in statistical analysis as raw scores and scaled scores are not equivalent.

Design and analysis

The study was a within-subject repeated-measures design. The two dependent variables were the IQ test scores for the WAIS-R and the WAIS-III. All participants completed both IQ tests, which were completed as per the standardized instructions. Only the compulsory subtests were administered. The order of test completion was counterbalanced to control for practice effects. Twenty-two participants completed the WAIS-R first and 23 completed the WAIS-III first. There was no significant difference between these two groups in age (*t*-test; $P = 0.94$) or gender (*t*-test: $P = 0.47$). For 11 of the participants, IQ scores were available from file records at Craegmoor for either the WAIS-R or the WAIS-III. For these participants only the second test was administered during the 2-month data collection period of the present study, giving an average gap between test administrations of 987 days (median 449, range 32–4255). The remaining 34 participants were administered both tests within the data collection period, the average gap between testing sessions being 5.7 days (median 7, range 1–15). Although the manuals (Wechsler 1981; Wechsler 1997) states that there should be at least 6 months between testing sessions, the short gap between IQ tests was necessary for practical reasons in running the study. This difference in duration between test administrations was controlled for by using the order of administration as a factor in the analysis and it did not significantly affect performance on the IQ tests (see Results).

Results

The average WAIS-R Full Scale IQ score was 67.1 (SD = 7.4, range 54–87) and the average WAIS-III Full Scale IQ score was 63.0 (SD = 8.3, range 51–88). Based upon the WAIS-R (the higher of the two IQ scores) nine of the participants' IQ score fell in the 70–79 range and three participants in the 80–89 IQ range. Based upon the WAIS-III Full Scale IQ scores four participants scored within the 70–79 IQ range and three participants scored within the 80–89 IQ range. Therefore, the majority of Full Scale IQ scores are within the learning disability intellectual ability range and the remainder within the borderline intellectual disability range. This mirrors the percentage of those with a diagnosis of learning disability and borderline learning disability outlined in the section Method.

Analyses of variance (ANOVA) were run with IQ test version (WAIS-R versus the WAIS-III; within-subject factor) the and administration order (WAIS-R or the WAIS-III being administered first; between-subjects factor) as factors, for each of the IQ scales (Full, Verbal and Performance Scale). The findings are summarized in Table 1.

Examination of the Full Scale IQ test version by administration order interaction suggests that the difference in WAIS-R versus WAIS-III scores was greater for those who had the WAIS-III administered first ($t = -6.68$, d.f. = 22, $P < 0.001$), compared with those

Table 1 Summary of ANOVA for each of the IQ scales

	WAIS-R FS IQ	*WAIS-III FS IQ*	*WAIS-R VS IQ*	*WAIS-III VS IQ*	*WAIS-R PS IQ*	*WAIS-III PS IQ*
Mean (SD)	67.1 (7.4)	63.0 (8.3)	67.6 (8.3)	64.9 (8.4)	69.5 (7.8)	67.5 (8.3)
main effect of test version[1] (*F*-value)	45.58***		12.55***		5.43**	
main effect of test administration order (*F*-value)	0.00		0.26		0.01	
test version by administration order interaction (*F*-value)	4.07*		1.32		2.51	

*$P < 0.05$, **$P < 0.025$ ***$P < 0.001$.

[1]d.f. = (1. 43) for each of the comparisons. WAIS-III, Wechsler Adult Intelligence Scale-III; WAIS-R,Wechsler Adult Intelligence Scale-Revised. FS, Full Scale; VS, Verbal Scale; PS, Performance Scale.

who had the WAIS-R administered first, though the difference remained significant ($t = -3.12$, d.f. = 21, $P < 0.005$). The reason for this interaction is probably due to practice effects causing an additional enhancement of WAIS-R scores if it is administered second (see Discussion).

Power calculations

Despite the significant results it was felt prudent to measure the effect size for the main effect of test version. Cohen's d was calculated for each IQ scale (Cohen 1992; Dunlap *et al.* 1996). Cohen's d revealed a medium effect size for the difference in Full Scale IQ score ($d = 0.51$), a small effect size for the Verbal Scale IQ score ($d = 0.32$) and a small effect size for the Performance Scale IQ score ($d = 0.25$).

Secondary analyses

For each IQ scale, a discrepancy score (WAIS-R − WAIS-III) was calculated to simplify comparisons. A mean discrepancy score of zero signifies that the tests are equivalent as there is no consistent difference between the two IQ scores. If the mean discrepancy score is positive this indicates that the WAIS-III score is consistently lower than the WAIS-R score. If it is negative then the WAIS-R score is consistently lower than the WAIS-III score. Using this convention we found a discrepancy score of 4.1 for the Full Scale; 2.7 for the Verbal Scale and 2.0 for the Performance Scale, indicating consistently lower IQ scores on the WAIS-III for all IQ scales (as hypothesized).

Duration between test administrations

The average gap between testing sessions was 5.7 days for the majority of participants ($n = 34$), but for those participants ($n = 11$) for who IQ scores were available from Craegmoor, the gap was larger between test administration, averaged at 987 days. No significant relationship was found between the duration between test administration (days) and Full Scale discrepancy score ($\rho = -0.21$, $n = 45$, $P = 0.17$), Verbal Scale discrepancy score ($\rho = -0.10$, $n = 45$,$P = 5.3$), or Performance Scale IQ discrepancy score ($\rho = -0.02$, $n = 45$, $P = 0.88$).

In addition, it was felt prudent to repeat the main analysis without those 11 participants for whom Craegmoor provided the IQ scores. In this secondary analysis, the results for the Full Scale IQ remained the same. The main effect of test version was significant, the main effect of test administration order remained non-significant and an interaction was found between the two effects (see, Discussion for interpretation of this result). The results for the Verbal Scale IQ remained largely the same. The main effect of test version was significant and the main effect of test administration order remained non-significant. However, an interaction was found between the two effects (again, see Discussion for interpretation of this result). The results for the Performance Scale IQ changed slightly as the main effect of test version became non-significant. The main effect of test administration order remained non-significant and no interaction was found between the two effects, as in the main analysis. This finding suggests that the main effect is not removed as the effect size is the same but due to the smaller sample size resulting in a loss of power the main effect of test version became non-significant. (Cohen's d being 0.25 in the first analysis and 0.27 in the second analysis, a small effect size in each case.)

Relationship between IQ score and discrepancy score

It remains unclear from the literature if the Flynn effect varies as a function of IQ and so it is possible that discrepancy scores between the tests may change as IQ becomes lower. To test this it was investigated if the discrepancy score is related to IQ score by correlating IQ score (WAIS-R IQ score) against the discrepancy score. For the Full Scale IQ there was no significant relationship ($r = 0.05$, $n = 45$, $P = 0.77$). There was a small significant relationship between Verbal Scale score and discrepancy score ($r = 0.30$, $n = 45$, $P < 0.05$) and a small but not statistically significant relationship between Performance Scale IQ score and discrepancy score ($r = 0.29$, $n = 45$, $P = 0.06$).

The correlation between the Full Scale IQ score and the discrepancy score reveals a small effect (Cohen 1992). To detect an effect of this size a much larger sample would be needed ($n = 783$, Cohen 1992). The relationship between the Performance Scale IQ score and discrepancy score is a medium effect size and so would require 85 participants to detect the effect, if the effect is present.

Discussion

Our results are consistent with the Flynn effect affecting IQ scores (Flynn 2000). According to Flynn (1984), the populations' IQ is increasing at roughly 0.3 IQ points per year. Therefore, with the 16 years between the norming of the WAIS-R and the WAIS-III, the decrease in

IQ score would be expected to be 4.8 Full Scale IQ points. We found a decrease of 4.1 IQ points.

Limitations of the design

Truscott & Frank (2001) noted that a limitation of their study was the 3-year duration between test administrations. Consequently the participant's life experience in these intervening years could have differentially affected performance on the two IQ tests, thus increasing the error variance. Therefore, in the present study, the participants were tested in close proximity for practical reasons, also ensuring that their level of challenging behaviour and life experience did not differentially affect performance on the two IQ tests. However, against this rationale it could be argued that this short duration between test administrations increased the chance of a practice effect.

Even though no significant practice effect was found, we did find an interaction between test version and administration order for the Full Scale IQ score. The practice effects on the WAIS-R (following administration of the WAIS-III) causing a further enhancement (increase) of WAIS-R score. As the practice effect served to increase the Flynn effect in this order, whereas when the WAIS-III was administered second the practice effect was working against the Flynn effect. The interaction may have been due to the WAIS-R having a higher proportion of shared items. The WAIS-III contains 212 of the original WAIS-R items plus 130 new items. The WAIS-R contains the same 212 items plus another 49 original items. Consequently, the WAIS-III consists of 61% of the shared items and the WAIS-R consists of 81% of the shared items. Therefore, if the WAIS-III was administered second the participant had previously completed 61% of the test, compared with 81% of the test if the WAIS-R was administered second. A similar finding has also been previously noted for revisions of the WISC (see Davis 1977; Wheaton *et al.* 1980). Nevertheless, these practice effects were minor in both directions and did not substantially alter the Flynn effect in either order condition.

It was decided for six people that their intellectual impairment was so severe that it would not be possible for them to sustain the attention required to complete the IQ tests. Therefore, the sample did not include those who had severe impairment of intellectual functioning.
As a limited range of IQ scores were represented by the sample, in order to investigate the relationship between the Flynn effect as a function of IQ score a greater spread of IQ scores should be tested. However, it would be very difficult to include those with a severe learning disability as it is unlikely that they could provide informed consent or complete the lengthy testing.

Limitations of the comparisons of various scales across studies

The main findings do not support those of Spitz (1989) or Spruill & Beck (1988) who found that the effect of the Flynn effect upon IQ scores reversed (producing a negative difference between IQ scores due to a higher, not a lower IQ score achieved on the most recent version of the IQ test) for those with a Full Scale IQ of <70. These authors tested differences between the WAIS and the WAIS-R. Spitz (1989) refers to Flynn (1985) for an explanation of these findings, suggesting that the disparity is due to a low number of individuals in the retarded range in the stratified samples of the WAIS tests causing discrepancies when converting raw scores into IQ scores. In addition, Spitz conducted statistical analysis on a number of samples combined and so this would have had increased variance compared with the sample recruited for the present study. Spruill and Beck suggest that the reversal of the Flynn effect is due to the scoring procedure at the lower end of the scale being different between the WAIS and the WAIS-R. Furthermore, the stratified sample on which the WAIS was based contained individuals institutionalized as a result of their retardation, whereas the WAIS-R sample did not contain any institutionalized individuals. It is not possible to compare the IQ scores on a subscale basis as the subscales consist of different subtests. In addition, the different subtests contain different items and different sliding scales of difficulty and so even a simple percentage correct comparison of performance would not be accurate.

We also found that the Verbal Scale IQ score decreased by 2.7 points and the Performance Scale IQ score decreased by 2.0 points. These findings are not in line with the suggestion of Flynn (1984, 1987) and Truscott & Frank (2001) that the populations' performance IQ is rising more rapidly than verbal IQ. This may be due to our sample being of a lower IQ range than those used in the previous studies, or that the previous studies examined different tests (e.g. WISC-R versus WISC-III). It could be that the children's performance IQ is rising faster than verbal IQ because of the advent of computers, mobile phones, etc.

Clinical and resource implications of the Flynn effect

The most widespread implication of the Flynn effect is that as the populations' IQ score steadily increases less people should fall below an IQ score of 70 and thus will no longer meet one of the diagnostic features of learning disability. Upon re-norming of the test, the proportion of people meeting this criterion should suddenly rise once more, before beginning the gradual decline. In line with this we show that the shift from WAIS-R to WAIS-III has produced a four-point decrease in obtained IQ for the population at or near the learning disability cut-off. A four-point decrease in Full Scale IQ could shift an individual's IQ score from the borderline learning disability range (Full Scale IQ score 70–80) to the learning disability range (Full Scale IQ score 60–69) and thus they may become eligible for learning disability services [if they also have deficits in adaptive functioning; see *Diagnostic and Statistical Manual of Mental Disorders* Fourth Edition, Text-Revised (DSM-IV-TR), American Psychiatric Association 2004; International Classification of Diseases-10 (ICD-10), World Health Organisation 1992].

The WAIS-III manual (Wechsler 1997) stipulates that the re-norming of the WAIS ensures that 2% of the population falls below a Full Scale IQ score of 70, 50% below 100, etc. According to the discrepancy score found in the present study, at the time that the WAIS-III was published, if testing Full Scale IQ with the WAIS-R, a higher cut off score of 74 would be required to ensure that 2% of the population falls within the learning disability range. This calculation also shows that when using the standard cut-off score of 70 for the learning disability range, this change of test version from the WAIS-R to the WAIS-III changes the proportion of the current population falling below an IQ score of 70 from 1.2% to 2%, an increase of around 66%. Thus, for the UK population (estimated to be 60 million) this translates to 720 000 people meeting criterion A of the learning disability diagnostic criteria if assessed with the WAIS-R, compared with 1.2 million if assessed with the WAIS-III, an extra 480 000 people.

It should be noted that a low IQ score alone does not qualify somebody for a diagnosis of learning disability or borderline learning disability (or mental retardation in America). It is also necessary for a properly qualified clinician to assess the individuals' social and adaptive functioning according to the ICD-10 (World Health Organisation 1992) or DSM-IV-TR (American Psychiatric Association 2004) criteria. However, as noted in the 'Introduction', IQ score is a important for a diagnosis of learning disability. The DSM-IV-TR goes as far as to call IQ score the essential feature of mental retardation.

Intelligence quotient score alone, and as part of a formal diagnosis of learning disability, is often used by professionals to make decisions about the provision of services (e.g. whether someone be admitted to mental health services or specialized learning disability services). Similarly, they are used to determine issues such as 'fitness to plead' and subsequent disposal into criminal justice or mental health services. The finding of a four-point shift in IQ scores between these tests must be taken into account when considering any such decision.

Acknowledgments

We are very grateful to Christine Lovell and all staff at Craegmoor Healthcare Services Ltd, Cardiff, for their help and enthusiasm in running this study. We also thank all of our participants for giving up their valuable time. In addition, many thanks to Rebecca Fone for her support and assistance in organizing the project.

Correspondence

Any correspondence should be directed to Professor R. J. Snowden, School of Psychology, Cardiff University, Cardiff CF10 3AT, UK (e-mail: snowden@Cardiff.ac.uk).

References

American Psychiatric Association (2004) *Diagnostic and Statistical Manual of Mental Disorders* 4th edn, text revision (DSM-IV-TR). APA, Washington, DC.

Cohen J. (1992) A power primer. *Psychological Bulletin* **112**, 155–159.

Davis E. E. (1977) Matched pair comparisons of WISC and WISC-R scores. *Psychology in Schools* **14**, 161–166.

Dunlap W. P., Cortina J. M., Vaslow J. B., & Burke M. J. (1996) Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods* **1**, 170–177.

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 for mental retardation? *American Journal of Mental Deficiency* **90**, 236–244.

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

Flynn J.R. (2000) The hidden history of IQ and Special Education: Can the problem be solved? *Psychology, Public Policy and Law* **6**, 191–198.

Kanaya T., Scullin M.H. & Ceci S.J. (2001) The Flynn Effect and U.S. Policies: the impact of rising IQ scores on American Society via Mental Retardation diagnoses. *American Psychologist* **58**, 778–790.

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

Spruill J. & Beck B.L. (1988) Comparison of the WAIS and the WAIS-R: different results for different IQ groups. *Professional Psychology: Research and Practice* **19**, 31–34.

Truscott S.D. & Frank A.J. (2001) Does the Flynn effect affect IQ scores of students classified as LD? *Journal of School Psychology* **39**, 319–334.

Wechsler D. (1974) *The Manual for the Wechsler Intelligence Scale for Children-Revised*. The Psychological Corporation, New York.

Wechsler D. (1981) *The Manual for the Wechsler Adult Intelligence Scale-Revised*. The Psychological Corporation, New York.

Wechsler D. (1991) *The Manual for the Wechsler Intelligence Scale for Children* 3rd edn. The Psychological Corporation, San Antonio, TX.

Wechsler D. (1997) *The Manual for the Wechsler Adult Intelligence Scale*, 3rd edn. The Psychological Corporation, San Antonio, TX.

Wheaton P.J., Vanderdriff A.F., & Nelson W.H. (1980) Comparability of the WISC and WISC-R with bright elementary school students. *Journal of School Psychology* **18**, 271–275.

World Health Organisation (1992) *ICD-10 Classification of Mental and Behavioural Disorders. Clinical Conditions and Diagnostic Guidelines*. WHO, Geneva.