Whitaker, Simon and Gordon, Shirley

Floor effects on the WISC-IV

INTERNATIONAL JOURNAL OF DEVELOPMENTAL DISABILITIES, Vol. 58 No. 1, March, 2012, 000–000

Floor effects on the WISC-IV

SIMON WHITAKER[1] and SHIRLEY GORDON[2]

[1]*The School of Human and Health Science, University of Huddersfield, Huddersfield, UK,* [2]*Lime Trees Child, Adolescent & Family Unit, York, UK*

Seventeen 16-year-olds in special education were given both the Wechsler Adult Intelligence Scale (WAIS)-III and the Wechsler Intelligence Scale for Children (WISC)-IV in counterbalanced order. It was found that there was a significant floor effect on the WISC-IV due to low raw scores and raw scores of 0 being given a scaled score of 1. Extrapolating the relationship between raw scores and scaled scores down below raw score, one showed that the WISC-IV may be overestimating some low intelligence quotients (IQs) by several IQ point due to this floor effect.

KEYWORDS Floor effect, WAIS-III, WISC-IV, Low IQ

Introduction

Whitaker (2005) and Whitaker and Wood (2008) have pointed to a possible floor effect on the UK standardizations of the Wechsler Adult Intelligence Scale third edition (WAIS-III UK) (Wechsler, 1997a) and the Wechsler Intelligence Scale for Children third edition (WISC-III UK) (Wechsler, 1992) due to scaled scores of one being given for very low raw scores. The aim of this paper is to investigate if the effect also occurs on the WISC-IV (UK) (Wechsler, 2004) and to extend the analysis to estimate the degree to which this floor effect could affect Full Scale Intelligence Quotient (FS IQ).

Derivation of IQ on the Wechsler intellectual assessments

The Wechsler intellectual assessments assess IQ by giving the client a series of subtests measuring different aspects of intellectual ability. Ten subtests are used to calculate FS IQ on the WISC-IV and 11 on the WAIS-III. The subtest raw scores are converted to a normalized scaled score with a mean of 10, a standard deviation (SD) of 3, and a range between 1 and 19. This conversion of raw scores to scaled scores is based on the cumulative frequency of raw scores obtained by 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. The FS IQ is then derived from the sum of the scaled scores from all the subtests in a similar way.

 DOI 10.1179/2047387711Y.0000000012

The floor effects

This method of assessing IQ leads to three possible sources of floor effects. First, as has been pointed out by Whitaker (2005) and Whitaker and Wood (2008), allocating a scaled score of 1 to very low but finite raw scores may overestimate the client's true ability on a subtest. This can be illustrated by an example from WISC-IV (UK) Administrative Manual (Wechsler, 2004) raw score to scaled score conversion tables for the Digit Span subtest for age group from 16:00 to 16:30.

It can be seen that there is a linear relationship between raw scores and scaled scores from raw score 18 to raw score 10 with a reduction in a raw score by 1 corresponding to a reduction in a scaled score by 1. However, all raw scores of 9 and less are given a scaled score of 1. This implies that a 16-year-old who gets a raw score of 9 has the same ability as a 16-year-old who gets a raw score of 2, 1, or 0. This seems unlikely.

It may be possible to gain an estimate of the true ability of a client who gains a low raw score by extrapolating the relationship between raw scores and scaled scores down below scaled score 1. If it is assumed that the same relationship between raw scores and scaled scores continues for raw scores below 10, it would follow that a raw score of 9 was equivalent to a scaled score of 1, a raw score of 9 was equivalent to a scaled score of 0, and a raw score of 8 to a scaled score of −1 one and so on. This extrapolation would result in some clients being allocated negative scaled scores; however, a 0 or a negative scaled scored does not imply negative ability but simply a low ability more than 3·6 SDs below the mean.

The second source of a floor effect derives from having a raw score of 0. A raw score of 0 suggests that the client's level of ability is below the level the subtest was designed to test, however, it is not clear if it is just below the subtest's base level or several SDs below it. A scaled score of 1 allocated to a raw score of 0 may therefore be a considerable overestimate of the client's true ability on that subtest. However, unlike the case when a client gets a low but finite raw score, it is not possible to estimate their true ability by extrapolation.

The third source of a floor effect is due to the relationship between the sum of scaled scores (SSS) and IQs in the low ability range. On the WISC-IV for IQ 70 and below, there is an almost linear relationship between SSS and FS IQ, in that a reduction in SSS of two points results in a reduction in IQ by one point. The minimum FS IQ on the WAIS-III of 45 is given for a SSS of 16 and all SSS between 16 and the minimum possible SSS of 11. On the WISC-IV, the minimum FS IQ of 40 is given for all SSS between 14 and the lowest SSS of 10. Given the linear relationship that preceded this point, it seems likely that FS IQs based on SSS below 16 on the WAIS-III and below 14 on the WISC-IV are overestimates. This error could be corrected for by extrapolating the relationship between sums of scaled scores and FS IQ down below the lowest IQs given in the tables.

These floor effects will clearly affect IQs in the 40s where inevitably there will be scaled scores of 1. The degree to which it is likely to affect IQs in the 50s, 60s and

Raw score	18	17	16	15	14	13	12	11	10	0–9
Scaled score	10	9	8	7	6	5	4	3	2	1

70s was investigated by Whitaker and Wood (2008). They plotted the distribution of scaled scores in all the WISC-III (UK) and WAIS-III (UK) that had been given as part of clinical practice by a psychology service provided for people with intellectual disabilities. It was found that the distribution of scaled scores for the WAIS-III (UK) was approximately normal with very few scaled scores of one, suggesting that this hidden floor effect would only be a potential problem for IQs in the 40s and 50s. However, with the WISC-III (UK), there was a skewed distribution of scaled scores with more scaled scores of one than any other scaled score. Scaled scores of 1 were shown in all IQ levels up to those in the 70s where they accounted for 10% of the scaled scores.

It is therefore likely that IQs in the 40s and 50s on the WAIS-III (UK) up to those in the 70s on the WISC-III (UK) have been artificially increased by these floor effects. However, this analysis is based on the Whitaker and Wood (2008) study which used unmatched samples on the WISC-III and WAIS-III. There is therefore the possibility that the difference in the proportion of scaled scores of 1 between the WISC-III and WAIS-III could have been due to differences in the clients who took both assessments. Also, it is not known if the latest version of the WISC, the WISC-IV, has a similar distribution of scaled scores as was found in the WISC-III.

The aims of this study are to:

- to assess if the skewed distributions of scaled score, in low IQs, found in the WISC-III also occurs in the WISC-IV;
- to correct scaled scores of one to lower scaled scores and IQs of 40 on the WISC-IV and 45 on the WAIS-III to lower IQs in order to estimate the degree to which the floor effects described above could affect FS IQ.

The data are taken from a previous study (Gordon *et al.*, 2010) in which both the WISC-IV (UK) and WAIS-III (UK) were given to 17 adolescents attending special school, in order to compare the two assessments in the low ability range. Although this study found a highly significant correlation between the two assessments of 0·933, it also showed the WISC-IV to systematically measure FS IQs lower than the WAIS-III by an average of just less than 12 points.

Method

Seventeen 16-year-old adolescents who attended three special schools in the local area were given both the WISC-IV (UK) and the WAIS-III (UK) in counterbalanced order. The assessments were done by the second author and an assistant psychologist, both of whom were experienced at administering the assessments. The study was done in 2006 and 2007.

Data analysis

For each subtest used to measure FS IQ on the WISC-IV (UK) and WAIS-III (UK), a best fit and equations were found between raw scores and scaled scores using data from the raw score to scaled score conversion tables in the manuals (Wechsler, 1997b, 2004) for scaled scores of 10 down to one, using SPSS. In all subtests, except Digit Symbol and Digit Span on the WAIS-III where the best fit was a power function, the best fit was linear equation. From these equations, tables were produced for each subtest giving a scaled score for each raw score between the raw score that achieved as scaled score of one in the manual tables and a raw score of 0. The lowest scaled score, corresponding to a raw score of 0, varied between subtests, with some subtests having possible scaled scores as low as −10 and others not being able to be extrapolated down below scaled score 1. On the WAIS-III subtests, these tables showed possible scaled scores of 0 on the Picture Completion, Vocabulary, Coding, and Similarities subtests and scaled scores as low as −2 on the Arithmetic subtest. On the WISC-IV, it was possible to extrapolate down to much lower scaled scores with all subtests used to calculate FS IQ having a possible scaled score of 0, with possible scaled scores down as low as −10 for Picture Concepts and Letter Number Sequencing subtests.

All raw scores that obtained a scaled score of 1 were then re-examined and if necessary an adjusted lower scale score was given. In the case of a raw score of 0, a scaled score equivalent to a raw score 0 was used, though it is possible that this still may be an overestimate of the client's true ability.

The SSS to FS IQ tables were also projected down below IQ 40 on the WISC-IV and IQ 45 on the WAIS-III using a similar method. New FS IQs were calculated based on sums of corrected scaled scores. If the SSS were less than 15 on the WAIS-III or 13 on the WISC-IV, then IQs less than 45 and 40 were derived using an extrapolation of the SSS to IQ tables.

Results

Table 1 shows the both the uncorrected and corrected FS-IQ scores, on the WISC-IV and WAIS-III, and indicates which test was taken first for each of the 17 adolescents.

Figure 1 shows the distribution of obtained uncorrected scaled scores on the WAIS-III and Fig. 2 the distribution of scaled scores corrected to include scaled scores of less than one. Both distributions appear to be approximately normal. Of the 10 scaled scores of one on the WAIS-III, four of them, shown by four separate participants, had raw scores low enough to be corrected to a scaled score of 0. As can be seen from Table 1, the correction resulted in two subjects (8 and 14) having their FS IQs reduced by one point, from 55 to 54 and from 49 to 48. None of the subjects had either corrected or uncorrected FS IQs of 45, so no correction was needed to be made for SSS of less than 15. There were, however, four raw scores of 0, two of which were scored by the same subject and two were from subtests that on the WAIS-III could not be extrapolated down below scaled score 0 at this age level. This leaves open the possibility that the IQs of these subjects may have been artificially increased.

The uncorrected distribution of scaled scores for the WISC-IV is shown in Fig. 3. The distribution is very different from the uncorrected distribution of scaled scores in the WAIS-III in Fig. 1: it is highly skewed with significantly more scaled scores of one than the WAIS-III ($P<0{\cdot}001$; Wilcoxon's signed rank test).

Of the 66 raw scores corresponding to the scaled scores of 1, 45 were so low that a correction to a lower scaled score could be made.

TABLE 1

THE UNCORRECTED AND CORRECTED FULL SCALE IQ SCORES ON BOTH THE WISC-IV (UK) AND THE WAIS-III (UK): THE DIFFERENCE BETWEEN THE CORRECTED AND UNCORRECTED FS IQS AND TEST WAS GIVEN, FIRST FOR EACH CLIENT

Client	Uncorrected WAIS-III FS IQ	Corrected WAIS-III FS IQ	Difference	Uncorrected WISC-IV FS IQ	Corrected WISC-IV FS IQ	Difference
1*	61	61	0	41	25	16
2*	72	72	0	58	56	2
3	69	69	0	57	54	3
4	49	49	0	40	13	27
5	67	67	0	54	54	0
6	70	70	0	60	60	0
7*	70	70	0	55	54	1
8	55	54	1	40	26	14
9	78	78	0	72	72	0
10	70	70	0	60	60	0
11*	69	69	0	58	58	0
12	67	67	0	52	52	0
13*	54	54	0	40	25	15
14*	49	48	1	40	9	31
15*	62	62	0	48	42	6
16*	67	67	0	58	58	0
17*	73	73	0	68	68	0
Mean	64·82	64·71		53·00	46·24	
SD	8·5	8·7		10·1	19·2	

Note: *WAIS-III, given first.

Figure 4 shows the distribution of scaled scores corrected for the floor effect. The frequency of corrected scaled scores does not gradually decease as one would expect if this was the low end of a normal distribution, rather there appears to be a second mode at scaled score -4 followed by a decrease in frequency down to scaled score -10.

Both the corrected and uncorrected FS IQs on the WISC-IV are given in Table 1. Nine of the 17 subjects have reduced FS IQs as a result of these corrections. It can be seen that four of the subjects (2, 3, 7, and 15) have corrected IQs of within six points of their original measured IQ. The other five subjects have much greater reduction in FS IQ to well below 40. In spite of these very low corrected scores on the WISC-IV, there were only four raw scores of 0, three of them scored by subject 14 and one by subject 4.

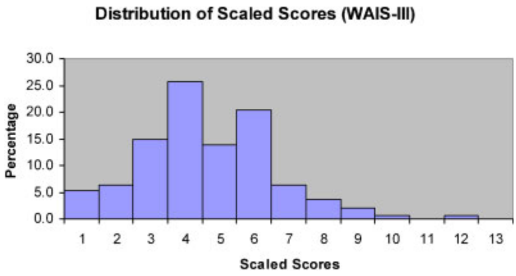

FIGURE 1 The distribution of scaled scores on the WAIS-III uncorrected for the floor effect.

Discussion

There is a significant difference between the distribution of scales between the WISC-IV (UK) and WAIS-III (UK), similar to those reported by Whitaker and Wood (2008) for the WISC-III (UK) and WAIS-III (UK). The WAIS-III (UK) showed a distribution of scaled scores that appeared approximately normal. Of the 10 scaled scores of 1, four, shown by four separate participants, had raw scores low enough to be corrected to a scaled score of 0. This correction resulted in two subjects having their FS IQs reduced by one point, from 49 to 48 and from 55 to 54. Although these corrections on the WAIS-III (UK) are made without empirical evidence, there are compelling logical reasons to suppose that they are appropriate.

The WISC-IV (UK), on the other hand, showed a highly skewed distribution of scaled scores with considerably more scaled scores of 1 than any other scaled score. Correcting the scaled scores of 1 resulted in 45 of the 66 scaled scores of 1 being reduced, which then resulted in nine of the 17 FS IQ being reduced. This demonstrates that there is a substantial floor effect on the WISC-IV (UK), though how great this is in terms of IQ points is difficult to quantify. Although the reduction in FS IQ for four of these nine subjects was less than seven points, the reductions were much greater for the other five clients, between 14 and 31 points, resulting in FS IQs between 26 and 9. Such FS IQs in these subjects do not seem to have face validity as the adolescents were attending schools for children with moderate learning disabilities. It does not seem credible that children at such a school would have an IQ as low as this. However, what does seem clear is that in the low range, the floor effects on the WISC-IV can result in an increased score of several IQ points.

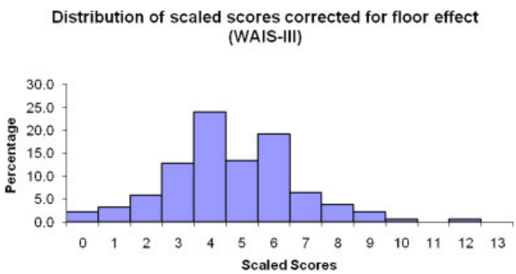

FIGURE 2 The distribution of scaled scores on the WAIS-III corrected for the floor effect.

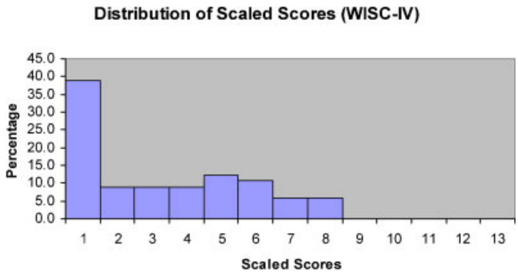

FIGURE 3 The distribution of scaled scores on the WISC-IV uncorrected for the floor effect.

The increase in scores resulting from the floor effect could have real world significance, particularly with the WISC-IV, were the effect was much greater. Although in the current study, it was only FS IQs in the 40s and 50s that were corrected for the floor effect, the sample was small and it was found by Whitaker and Wood (2008) that for the WISC-III 10% of scaled scores were scaled score 1, suggesting that the floor effect may well increase scores in the critical range for a diagnosis of intellectual disabilities of IQs in the 60s and 70s. For IQs in the 40s and 50s, it may also be important to establish a correct FS IQ for clinical reasons, for example, IQ has a predictive value with regard to a child's academic performance and could play a part in a decision re school placement. The need to get accurate scores for low IQ in research is also an issue. Laird and Whitaker (2011) found that IQ is used as one of the main description of clients in academic studies of intellectual disability, with IQs in the 30s and 40s being give as well as IQs in the 50s and 60s. IQ is also used as both a dependent and independent variable; it was argued that the inaccuracy of assessment, due to the floor effect as well as other sources of error in the low range (cf. Whitaker, 2010), would reduce correlations and statistical significance levels, possibly resulting in significant results being missed.

The implications of the difference in the scores between the WAIS-III (UK) and the WISC-IV (UK) have already been discussed by Gordon and colleagues (2010); however, it is worth noting that there is a difference in the number of clients who reached the criterion of having an IQ less than 70 for a diagnosis of intellectual disability on the WISC-IV (UK) and the WAIS-III (UK). It can be seen from Table 1 that on the WISC-IV (UK, uncorrected), 16 of the clients reach this criterion, whereas on the WAIS-IV (UK, uncorrected), only 11 of them reach it. This gives further weight to the argument by Whitaker (2008) that having a measured IQ

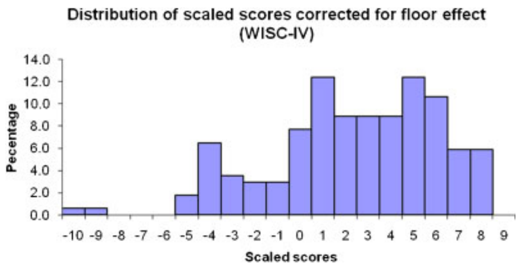

FIGURE 4 The distribution of scaled scores on the WISC-IV corrected for the floor effect.

below a critical point should not be a necessary criterion for obtaining a diagnosis of intellectual disabilities.

Although it has been demonstrated that there is a major floor effect on the WISC-IV and a much smaller floor effect on the WAIS-III, there are a number of outstanding issues:

The WAIS-III has now been superseded by the WAIS-IV (UK) (Wechsler, 2009). Although analysis of the criteria for gaining a scaled score of 2 (Whitaker, 2012) suggests that the distribution of scaled scores on the WAIS-IV would differ from the WAIS-III, this needs to be empirically demonstrated.

It is unclear to what extent the floor effect can be corrected for. The corrections done in this study on the WAIS-III, where scaled score was extrapolated down to a scaled score of 0, seem logical and reasonable. It is, therefore, the recommendation of the current authors that if a low raw score merits a scaled score of 0, this should be done and that corrected IQ scores should be presented together with uncorrected scores with a clear explanation as to why this was done.

It is not recommended that scaled scores should be corrected below scaled score 0 as it is not clear how accurately this could be done; however, where there are raw scores that seem to correspond to scaled scores of less than 0, no IQ should be calculated or it should be stated in the report that the derived IQ is likely to be an overestimate of the client's true IQ due to a floor effect.

Finally, the authors acknowledge that the current study is limited in the small sample size of 17 subjects. The floor effects need to be examined in much larger samples of WISC-IVs and WAIS-IVs. This could not only confirm the current findings for the WISC-IV and show that the floor effect on the WAIS-IV is similar to the WAIS-III, but also the larger sample would enable us to test if the distribution of corrected scaled scores tended to a normal distribution, as would be expected if the corrections are valid.

References

Gordon, S., Duff, S., Davison, T. and Whitaker, S. 2010. Comparison of the WAIS-III and WISC-IV in 16 year old special education students, *J. Appl. Res. Intell. Disabil.*, 23, 197–200.

Laird, C. and Whitaker, S. 2011. The use of IQ and descriptions of people with intellectual disabilities in the scientific literature, *Br. J. Dev. Disabil.*, 57, 181–189.

Wechsler, D. 1992. Wechsler Intelligence Scale for Children — Third UK Edition and Wechsler Memory Scale — Third Edition: Technical manual; London, The Psychological Corporation.

Wechsler, D. 1997a. Wechsler Adult Intelligence Scale — Third UK Edition and Wechsler Memory Scale — Third Edition: technical manual; San Antonio, TX, The Psychological Corporation.

Wechsler, D. 1997b. Wechsler Adult Intelligence Scale — Third UK Edition: administrative and scoring manual; London, The Psychological Corporation.

Wechsler, D. 2004. Wechsler Intelligence Scale for Children — Fourth UK Edition: administrative and scoring manual; London, The Psychological Corporation.

Wechsler, D. 2009. Wechsler Adult Intelligence Scale – Fourth UK Edition: Administrative and Scoring Manual; London, The Psychological Corporation.

Whitaker, S. 2005. The use of the WISC-III and the WAIS-III with people with a learning disability: three concerns, *Clin. Psychol.*, 50, 37–40.

Whitaker, S. 2008. Intellectual disability: a concept in need of revision, *Br. J. Dev. Disabil.*, 54, 3–9.

Whitaker, S. 2010. Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III, *Personal. Indiv. Diff.*, 48, 517–521.

Whitaker, S. 2012. The measurement of low IQ with the WAIS-IV: a critical review, *Clinical Psychology Forum*, (In Press).

Whitaker, S. and Wood, C. 2008. The distribution of scale score and possible floor effects on the WISC-III and WAIS-III, *J. Appl. Res. Intell. Disabil.*, **21**, 136–141.

Notes on contributors

Correspondence to: S Whitaker, Room HW 2/08, University of Huddersfield, Queen's Gate, Huddersfield HD1 3DH, UK. Email: s.whitaker@hud.ac.uk