University of HUDDERSFIELD

Error in the Measurement of Low IQ

Dr Simon Whitaker - Consultant Clinical Psychologist

HUMAN& HEALTH SCIENCES

Errors are Both Chance and Systematic

Chance Error

The 95% confidence interval of an IQ test score is a function of the reliability of the test given by the following formula:

95% confidence interval = (SD √(1-r))*1.96

Where: r is the reliability coefficient and SD is the standard deviation of the test (which is 15 in the case of the WISC-IV and the WAIS-III).

The 95% confidence intervals in the test manuals for WISC-IV and WAIS-III for FS IQ 70 are 67-75 and 66-76 respectively.

However, this is only based on the chance error due to a lack of internal consistency, which is estimated from the split-half reliability of the tests from the standardisation sample.

The 95% confidence interval in the manuals is not based on a low IQ sample and does not take into account other chance errors due to variation between the conditions under which the test was administered, which is given by the test re-test reliability.

Whitaker (in press) did a meta analysis of test re-test reliability for IQs less than 80 and found: a weighted mean stability coefficient for FS IQ of .82 for a mean test re-test interval of 2.8 years. As there was no statistically significant relationship between inter-test interval and stability, it is likely that little of the other variance was due to change in actual intellectual ability over the year. If this stability coefficient is used to calculate a 95% confidence interval, it results in an interval 13 points either side of the measured FS IQ.

There is one study that gives data on internal consistency in the low IQ range: Davis (1966) found split-half reliabilities of .90 for children with moderate ID (mean IQ 48) and .97 for those with borderline mental ID (mean IQ 76), the weighted mean reliability being .92.

Adding these errors together gives a 95% confidence interval of 15 points either side of the measured IQ.

Systematic Error

The Floor Effect

Zero and low raw score on subtests are given scaled scores of 1. For example: the relationship between scaled score and raw scores for Digit Span on the WISC-IV for age group 16:00 to 16:30 is:-

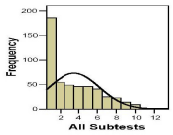
Scaled Score: 10 9 8 7 6 5 4 3 2 1
Raw Score: 18 17 16 15 14 13 12 11 10 0-9

It is likely that a scaled score of 1 is an overestimate of ability for clients who get raw scores less than 8.

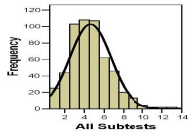
Whitaker and Wood (in press): Distribution of scaled scores on 50 WISC-IIIs and 49 WAIS-IIIs done as part of clinical practice with people with LD.

Frequency of Scaled Scores

WISC-III



WAIS-III



10% of scaled scores on the WISC-III for IQs in the 70s were scaled score 1.

The Flynn Effect

Flynn (1984, 1985) has shown that measured intellectual ability has been going up by about 3 points a decade since about 1900. This has resulted in older tests producing higher IQ scores. The rate at which this systematic error has occurred is about 3 points a decade.

If we were confident that the Flynn Effect was continuing at a rate of 3 points per decade, then this could be compensated for. However, recent evidence suggests that the effect may have slowed down or gone into reverse in the low IQ range. Therefore, as we do not know if the Flynn Effect is currently causing tests to measure too high or too low, there is an additional error of the order of 3 points per decade since a test was standardised.

Disagreement Between Tests

Both the WISC-IV and WAIS-III are considered to be "gold standard" IQ tests, against which other tests should be compared. It is therefore important that they produce very close results.

Both Flynn (1985) and Spitz (1986; 1989) reported that the WISC-R gives IQ scores up to 15 points lower than the WAIS-R for IQs of 70 and below.

Gordon (2007) gave the WISC-IV and the WAIS-III to seventeen 16-year-olds in special education. The tests were administered in counter-balanced order.

	WISC-IV	WAIS-III	Diff	r
Full Scale IQ	53.00	64.82	11.82	.93

About two points of the nearly 12 point difference will be accounted for by the Flynn Effect. So either the WISC-IV is measuring 10 points too low or the WAIS-III is measuring 10 points too high or both tests are in error by between 1 and 9 points.

Summing Errors and Deriving a Confidence Interval

WISC-IV
There is a chance error of 15 points, to which two points must be added due to uncertainty as to how the Flynn Effect has affected the intellectual ability of people with low IQs since the WISC-IV was standardised giving an effective 95% confidence interval of 17 points. It may also measure 10 points too low due to other systematic errors demonstrated by the difference with the WAIS-III, but possibly measure one or two points too high due to the floor effect. This suggests that in addition to the confidence interval it may measure eight points too low. If these sources of error are added together then the effective confidence interval extends 25 points above the measured IQ and 17 points below.

WAIS-III
There is a chance error of 15 points, to which three points must be added due to uncertainty as to the degree to which the Flynn Effect has affected low IQs since the WAIS-III was standardised giving an effective 95% confidence interval of 18 points. It may also measure 10 points too high due other systematic error demonstrated by the difference with the WISC-IV and possibly measure one point too high due to the floor effect. This suggests that in addition to the confidence interval it may measure 11 points too high. If these sources of error are added together then the effective confidence interval extends 18 points above the measurec IQ and 29 points below.