Running head: WAIS-III Short-Form

British Journal of Clinical Psychology, in press

An Index-Based Short-Form of the WAIS-III with Accompanying Analysis of Reliability and Abnormality of Differences

John R. Crawford, Samantha Allum, and Jess E. Kinion

School of Psychology

University of Aberdeen

Address for correspondence: Professor John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 3HN, United Kingdom. E-mail: j.crawford@abdn.ac.uk

Abstract

Objectives. To develop an Index-based, seven subtest, short-form of the WAIS-III that offers the same comprehensive range of analytic methods available for the full-length version

Design and Methods. Psychometric

Results. The short-form Indices had high reliability and criterion validity. Scores are expressed as Index scores and as percentiles. Methods are provided that allow setting of confidence limits on scores, and analysis of the reliability and abnormality of Index score differences. A computer program that automates scoring and implements all the analytical methods accompanies this paper and can be downloaded from the following web address: http://www.abdn.ac.uk/~psy086/dept/sf_wais3.htm.

Conclusions. The short-form will be useful when pressure of time or client fatigue precludes use of a full-length WAIS-III. The accompanying computer program scores and analyzes an individual's performance on the short-form instantaneously and minimizes the chance of clerical error

Like its predecessors, the Wechsler Adult Intelligence Scale third edition (WAIS-III; Wechsler, 1997; Wechsler, Wycherley, Benjamin, Crawford, & Mockler, 1998) continues to serve as the workhorse of cognitive assessment in clinical research and practice. Time constraints and potential problems with patient fatigue mean that a short-form version of the WAIS-III is often required. Four principal approaches to the development of short-forms can be delineated. In the Satz-Mogel approach, which is perhaps the most radical method, all subtests are administered but every second or third test item is omitted; see Ryan, Lopez and Werth (1999) for a Satz-Mogel short-form for the WAIS-III. It could be argued that this is a wasteful approach because all subtests are not created equal: some are more reliable and more valid indicators of the ability dimensions or factors that underlie WAIS-III performance. Thus, when time is limited, there is a case for focusing on these subtests rather than spreading effort widely but thinly.

The three remaining approaches all omit subtests but differ in how the short-form is constructed. Probably the most widely adopted approach is to prorate omitted subtests (i.e., substitute the mean score on those subtests administered for those omitted) and thereafter proceed as though the full-length version had been given. Yet another alternative is to build regression equations to predict full-length IQs or Index scores from a subset of the subtests (Crawford, Allan, & Jack, 1992; Reynolds, Willson, & Clark, 1983).

The fourth approach, and the one adopted here, is that originally proposed by Tellegen and Briggs (1967); see also Atkinson (1991) for an excellent example of its application to the WAIS-R. With this approach, the subtests selected for the short-form are combined into composites and the composite scores transformed to an

IQ metric (i.e., mean 100 and standard deviation 15). Thus the aim is not to predict full-length IQs or Indices but to treat the composites as free standing measures of ability. This does not mean that criterion validity is necessarily ignored: For example, subtests could be selected to maximize the correlation between the short-form IQs or Indices and their full-length counterparts.

This latter approach has a very significant advantage: it is relatively simple to provide all the additional information required to conduct the same forms of quantitative analysis on the short-form scores as are available for the full-length WAIS-III. This is in marked contrast to the other methods of forming short-forms. Taking prorating as an example: The reliability of the prorated IQs or Indices will differ from their full-length counterparts, thereby invalidating the use of confidence intervals on scores derived from the full-length version. The differences in reliabilities also invalidate the use of the tabled values in the WAIS-III manual (Table B.1) when attempting to test for reliable differences between an Individual's IQs or Index scores. Moreover, for the full-length WAIS-III, analysis of the abnormality of differences among an individual's IQs or Indices can be conducted using a table of the base rates for differences in the standardization sample (Table B.2). The use of this table with prorated scores is questionable because (a) the correlations between the prorated IQs or Indices will differ from their full-length counterparts, and (b) these correlations determine the level of abnormality of any differences (Crawford, Garthwaite, & Gault, 2007). With the approach used in the present study all of these problems are overcome by calculating the reliabilities and intercorrelations of the short-form IQs and Indices from the statistics of the subtests contributing to them.

Turning now to the selection of subtests for the short-form: The primary consideration was that the short-form should provide Index scores rather than Verbal and Performance IQs. Index scores reflect the underlying factor structure of the WAIS-III and therefore have superior construct validity to these latter measures. They are also only marginally less reliable (the differences in reliability largely stem from the use of fewer subtests for the Indices versus IQs). Furthermore, empirical studies have shown that factor based composites are superior to VIQ and PIQ at differentiating between healthy and impaired functioning (e.g., Crawford, Johnson, Mychalkiw, & Moore, 1997).

So that there would be significant time savings when using the short-form we limited it to seven subtests: there were two indicators each for three of the WAIS-III Indices and one for the Processing Speed (PS) Index. Vocabulary and Similarities were selected for the Verbal Comprehension (VC) Index: Vocabulary is highly reliable and has the highest loading on the verbal comprehension factor (Tulsky, Zhu, & Ledbetter, 1997), Similarities is a little less reliable and has a slightly lower loading on the verbal comprehension factor than Information but is a useful measure of the ability to engage in basic abstract verbal reasoning. Block Design and Matrix Reasoning were selected for the Perceptual Organisation (PO) Index: these subtests have higher reliabilities and higher loadings on the perceptual organisation factor than Picture Completion. Arithmetic and Digit Span were selected for the Working Memory (WM) Index: we considered Letter Number Sequencing as a possibility as it has a higher loading than Arithmetic on the working memory factor however, it is less reliable than Arithmetic and is not a core subtest in determining Full Scale IQ. Finally, Digit Symbol was selected for the Processing Speed Index: this subtest is more reliable and has a higher loading on the processing speed factor than Symbol Search.

A reasonable amount of technical detail on the methods used to build and analyze the short-form is provided. This was because we considered it important that potential users of the short-form should be fully informed of the methods that underlie the results it provides. Although details on the set of methods used could be found in various papers and textbooks, they are gathered together here in a systematic fashion. Therefore, the methods (which are mainly derived from classical test theory) could readily be adopted by others to create either alternative WAIS-III short-forms or short-form versions of other psychological instruments. Finally, this paper contains all the information required to score and analyze the short-form. However, we have also developed a computer program to automate this process (see later for details). The program provides a convenient alternative to hand scoring and reduces the chance of clerical error.

Building the Index-Based Short-form

The first step in developing short-forms of the Indices is to determine the means and standard deviation of the composites. The means are obtained simply by multiplying the number of subtests in each composite by 10 (the mean of an individual WAIS-III subtest); thus, for the Verbal Comprehension composite, the mean is 20, and for FSIQ, the mean is 70. The standard deviation of a composite is a function of the standard deviations of the individual components (i.e., the subtests) and their intercorrelations. The simplest way of obtaining this standard deviation is to form a variance-covariance matrix (by multiplying each correlation by the standard deviations of the relevant pairs of components; in the present case, because the subtests have a common standard deviation of 3, the correlation is simply multiplied by 9). For example, from the WAIS-III technical manual, the correlation between Vocabulary and Similarities is 0.76 and thus the covariance is 6.84. The sum of the elements in this covariance matrix is the variance of the composite and by taking the square root of this we obtain the standard deviation of the composite (5.628 in this case).

The means and standard deviations of the five composites are presented in Table 1 (note that the Processing Speed "composite" consists only of Digit Symbol and thus the mean and standard deviation are simply 10 and 3 respectively).

Having obtained the means and standard deviations of the composites, we now have the constants required to be able to transform each of the composite scores to have a mean and standard deviation of 100 and 15 respectively. The generic formula is

$$X_{\text{new}} = \frac{s_{\text{new}}}{s_{\text{old}}}\left(X_{\text{old}} - \overline{X}_{\text{old}}\right) + \overline{X}_{\text{new}}, \quad (1)$$

where X_{new} = the transformed score, X_{old} = the original score, s_{old} = the standard deviation of the original scale, s_{new} = the standard deviation of the scale you wish to convert to, $\overline{X}_{\text{old}}$ = the mean of the original scale, and $\overline{X}_{new}$ = the mean of the scale you wish to convert to (Crawford, 2004).

Thus for example, if the sum of an individual's subtest scores on Vocabulary and Similarities is 15 then the short-form VC Index score is 87 after rounding. Formula (1) was used to generate the tables for conversion of the sums of subtest scores to short-form Index scores and FSIQ (Tables 2 to 6) and is also used in the computer program that accompanies this paper. For the full-length Indices, scores are also expressed as percentiles. Therefore, in keeping with the aim of providing equivalent information for the short-form Indices, percentile norms are also presented in Tables 2 to 6 and are provided by the computer program. To express the scores as percentiles, Index scores were expressed as *z*, and the probabilities corresponding to these quantiles multiplied by 100. Thus for example, the *z* for an Index score of 115 is +1.0 and the score is thus at the 84th percentile. In Tables 2 to 6 percentiles are expressed as integers unless the Index score is very extreme (i.e., below the 1st or above the 99th percentile) in which case they are presented to one decimal place.

Reliabilities and standard errors of measurement for the short-form Indices

In order to set confidence limits on an individual's score on the short-form Indices, and to test whether an individual exhibits reliable differences between her/his short-form Index scores, it is necessary to obtain the standard error of measurement for each short-form Index. To obtain this statistic we first need to obtain the reliability of the short-form Indices. Of course the reliability of the short-form is also an important piece of information in its own right; measures with low reliability should be avoided, particularly when the concern is with assessing an *individual's* performance (Crawford, 2004).

When, as in the present case, the components have equal means and standard deviations, and are given equal weights in determining the composite score, the reliability of a composite is a simple function of the reliabilities of the components and their intercorrelations (the higher the intercorrelations between components, the higher the reliability of the composite). The formula (Nunnally & Bernstein, 1994) is

$$r_{YY} = 1 - \frac{k - \Sigma r_{XX}}{\bar{R}_Y}, \qquad (2)$$

where k = the number of components, r_{XX} are the reliabilities of the components and $\bar{R}_Y$ is the sum of elements of the correlation matrix for the components (including the unities in the diagonal).

The reliabilities of the short-form Indices calculated by this method are presented in Table 1; the reliabilities of the corresponding full-length Indices are also presented for comparison purposes (these latter reliabilities are from the WAIS-III technical manual). It can be seen that the reliabilities of the short-form Indices are all very high and only marginally lower than the reliabilities of their full-length equivalents (the very modest reduction in reliability when moving from a full-length to short-form Index can be attributed to the fact that those subtests selected for inclusion in the short-form had, in most cases, higher reliabilities and higher intercorrelations than those omitted).

Having obtained the reliabilities of the short-form Indices, the next stage is to calculate their standard errors of measurement. The formula (Ley, 1972) for the standard error of measurement (SEM) is

$$\text{SEM}_X = s_X \sqrt{1 - r_{XX}}\ , \tag{3}$$

where s_X is the standard deviation of the scale in question, and r_{XX} is its reliability coefficient. The standard errors of measurement for the short-from Indices are presented in Table 1. In the present case we also compute the standard errors of measurement for scores expressed on a true score metric: these latter standard errors are obtained by multiplying the standard error of measurement for obtained scores by the reliability coefficient for the relevant composite (Glutting, Mcdermott, & Stanley, 1987; Stanley, 1971); i.e.,

$$\text{SEM}_{X_t} = r_{XX}\left(s_X \sqrt{1 - r_{XX}}\right), \tag{4}$$

where all terms have been previously defined. These two forms of standard errors will be used to provide alternative means of (a) setting confidence limits on Index scores, and (b) testing for reliable differences between Index scores (see later).

Intercorrelations of the short-form Indices and correlations with their full-length equivalents

When attempting to detect acquired impairments, it is important to quantify the degree of abnormality of any differences in an individual's Index score profile. Quantifying the abnormality of differences requires the standard deviation of the differences between each of the Indices, which in turn, requires knowledge of the correlations between the Indices. These correlations can be calculated from the matrix of correlations between the subtests contributing to the Indices (Nunnally & Bernstein, 1994) using the formula

$$r_{XY} = \frac{\bar{\mathbf{R}}_{XY}}{\sqrt{\bar{\mathbf{R}}_X}\sqrt{\bar{\mathbf{R}}_Y}}, \tag{5}$$

where $\bar{\mathbf{R}}_{XY}$ is the sum of the correlations of each variable in composite *X* (e.g., the VC short-form) with each variable in composite *Y* (e.g., the PO short-form), and $\bar{\mathbf{R}}_X$ and $\bar{\mathbf{R}}_Y$ are the sums of the full correlation matrices for each composite. Applying this formula, the correlations between the short-form Indices were as follows: VC with PO = 0.63; VC with WM = 0.62; VC with PS = 0.45; PO with WM = 0.61; PO with PS =0.45; WM with PS = 0.45.

The formula for the correlation between composites is flexible in that it can be used to calculate the correlation between two composites when they have components in common; the components common to both are entered into the within-composite matrices ($\mathbf{R}_X$ and $\mathbf{R}_Y$) for both composites. This means that the formula can also be used to calculate the correlation between each short-form Index and its full-length equivalent; such correlations are criterion validity coefficients. The correlations are presented in Table 1, from which it can be seen that all correlations are very high. They range from 0.91 for Processing Speed to 0.97 for Verbal Comprehension; note also that the correlation between the short-form FSIQ and full-length FSIQ is also very high (0.97).

Confidence intervals on short-form Index scores

Confidence limits on test scores are useful because they serve the general purpose of reminding users that test scores are fallible (they counter any tendencies to reify the score obtained) and serve the very specific purpose of quantifying this fallibility (Crawford, 2004). For the full-length WAIS-III, confidence intervals for Index scores are true score confidence intervals and are centered on estimated true scores rather than on individuals' obtained scores (Glutting, Mcdermott, & Stanley, 1987). For consistency the same approach to setting confidence intervals is made available for the short-form Indices. Estimated true score are obtained using the following formula

$$\text{True score} = r_{XX}\left(X - \overline{X}\right) + \overline{X}\,, \tag{6}$$

where X is the obtained score and $\overline{X}$ is the mean for the scale (Crawford, Henry, Ward, & Blake, 2006). In words, an obtained score is expressed as a deviation score by subtracting the mean (100 in this case) and then multiplying the deviation score by the reliability of the test. This will pull in scores towards the mean (as reliability coefficients are always less than 1). The mean is then added back on to obtain the estimated true score. So, if an individual obtained a score of 90 on a test with a mean of 100 and reliability of 0.8, the estimated true score would be 92.

The estimated true score can be seen as striking a compromise between predicting the individual is average (the best guess in the absence of any information) and predicting that they are as extreme as their obtained score indicates (Crawford, Smith, Maylor, Della Sala, & Logie, 2003). Note that, if the test has high reliability (as is the case in the present context), then there will only be modest differences between obtained scores and estimated true scores (particularly if scores are near to the mean in the first place).

To form 95% confidence intervals for scores expressed on a true score metric (centered on the estimated true score) the standard error of measurement of true scores (formula 4) for each Index is multiplied by 1.96. Subtracting this quantity from the estimated true score yields the lower limit and adding it yields the upper limit. 90% confidence limits are formed in the same way but substituting 1.645 for 1.96; the accompanying computer program offers a choice between these two sets of limits; for reasons of space the tabled values are limited to 95% limits. To reiterate, these limits are calculated using the same method as was used to report limits for the full-length Indices in the WAIS-III manual. These 95% limits on true scores appear in brackets in Tables 2 to 6; the limits without brackets in these tables are based on the traditional approach described next.

The traditional approach (Charter & Feldt, 2001) to obtaining confidence limits for true scores expresses the limits on an obtained score metric and are centered on the individual's obtained score rather than estimated true score. The limits are obtained by multiplying the standard error of measurement of obtained scores (formula 3) by the appropriate value of *z* (1.96 for 95% two-sided limits, 1.645 for 90% two-sided limits). That is

$$\mathrm{CI} = X_0 \pm z\left(\mathrm{SEM}_X\right). \tag{7}$$

The 95% confidence limits calculated using formula (7) are presented in Tables 2 to 6. We decided to offer these alternative confidence limits because of criticisms of the Glutting et al method offered by Charter and Feldt (2001). The arguments are technical but centre around the mixing of parameter estimates from different theories of measurement. Moreover, as Charter and Feldt (2001) point out, JC Stanley, the principal psychometric theorist on the Glutting et al. (1987) paper, would appear to have reverted to the "traditional" approach in subsequent writings (Hopkins, Stanley, & Hopkins, 1990). Note also that true score limits are potentially misleading for users. It is important to be aware that the standard deviation of true scores is not 15: rather it is $r_{XX}15$ so that the true score standard deviations for the Indices are necessarily less than 15 and are not constant across the four Indices (either for the full-length or short-form versions) because the Indices differ in their reliabilities.

Percentile confidence intervals on short-form Index scores

All authorities on psychological measurement agree that confidence intervals should accompany test scores. However, it remains the case that some psychologists do not routinely record confidence limits. There is also the danger that others will dutifully record the confidence limits but that, thereafter, these limits play no further part in test interpretation. Thus it could be argued that anything that serves to increase the perceived relevance of confidence limits should be encouraged. Crawford and Garthwaite (Submitted) have recently argued that expressing confidence limits as percentile ranks will help to achieve this aim (they also provided such limits for the full-length WAIS-III).

Expressing confidence limits on a score as percentile ranks is very easily achieved: the standard score limits need only be converted to *z* and the probability of *z* (obtained from a table of areas under the normal curve or algorithmic equivalent) multiplied by 100. For example, suppose an individual obtains a score of 84 on the short-form Verbal Comprehension Index (the score is therefore at the 14^{th} percentile): using the traditional method of setting confidence limits on the lower and upper limits on this score (77 and 91) correspond to *z*s of -1.53 and -0.60. Thus the 95% confidence interval, with the endpoints expressed as percentile ranks, is from the 6th percentile to the 27th percentile.

The WAIS-III manual does not report confidence intervals of this form (neither to our knowledge is this practice currently adopted for any other psychological test). However, as Crawford and Garthwaite (Submitted) argue, such limits are more directly meaningful than standard score limits and offer what is, perhaps, a more stark reminder of the uncertainties involved in attempting to quantify an individual's level of cognitive functioning. The lower limit on the percentile rank in the foregoing example (the lower limit is at the 6th percentile) is clearly more tangible than the Index score equivalent (77) since this latter quantity only becomes meaningful when we know that 6% of the normative population is expected to obtain a lower score.

In view of the foregoing arguments, the computer program that accompanies this paper provides conventional confidence intervals but supplements these with confidence intervals expressed as percentile ranks. Because of pressure of space, the conversion tables (Tables 2 to 6) do not record these latter intervals.

Testing for reliable differences among an individual's Index scores

Individuals will usually exhibit differences between their Index scores on the short-form. A basic issue is whether such differences are reliable; that is, are they large enough to render it unlikely that they simply reflect measurement error. The standard error of measurement of the difference (SEM_D) is used to test for reliable differences between scores (Anastasi, 1990). The formula is

$$\text{SEM}_D = \sqrt{\text{SEM}_X^2 + \text{SEM}_Y^2} \tag{8}$$

where SEM_X and SEM_Y are the standard errors of measurement obtained using formula (3). The standard errors for each of the six pairwise comparisons between Indices are presented in Table 7. To obtain critical values for significance at various *p* values, the SEM_D is multiplied by the corresponding values of *z* (a standard normal deviate); for example, the SEM_D is multiplied by 1.96 to obtain the critical value for significance at the 0.05 level (two-tailed). The differences observed in an individual are then compared to these critical values. Critical values for significance at the .15, .10, .05, and .01 level (two-tailed) are recorded in Table 8 for each of the six possible pairwise comparisons between short-form Indices. For example, suppose that an individual obtained a subtest score of 10 on Vocabulary and a score of 11 on Similarities (yielding an Index score of 103) and scores of 9 and 8 on Block Design and Matrix Reasoning (yielding a PO Index score of 92). Thus there is a difference of 11 points between VC and PO. From Table 8 it can be seen that this is a reliable difference at the 0.05 level, two-tailed (the critical value is 10.80). Note that this result is also a testament to the reliabilities of the short-form Indices: the difference in raw scores is relatively modest but the difference is reliable even on a two-tailed test.

A closely related alternative to the use of these critical values is to divide an observed difference by the relevant SEM_D (5.511 in the present case; see Table 7), the resultant value is treated as a standard normal deviate and the precise probability of this *z* can be obtained (e.g., from tables of areas under the normal curve or a statistics package). To continue with the previous example: for a difference of 11 points, *z* is 1.996 and the corresponding two-tailed probability is approximately 0.045. This latter approach is implemented in the computer program that accompanies this paper (these data are not presented in the present paper because they would require voluminous tables).

Note that the critical values in Table 8 are two-tailed. If a clinician has, a priori, a directional hypothesis concerning a specific pair of Indices they may prefer to perform a one-tailed test. The computer program provides one- and two-tailed values; those who choose to work from the tables should note that the critical values for the 0.10 level of significance two-tailed also serve as critical values or a one-tailed test at the 0.5 level.

Both of these foregoing methods test for a reliable difference between *obtained* scores. Some authorities on test theory (Silverstein, 1989; Stanley, 1971) have argued that such an analysis should instead be conducted using estimated true scores (see Crawford, Henry, Ward, & Blake, 2006 for a recent example). The general approach is the same as that outlined above for observed scores, except that interest is in the difference between an individual's estimated true scores (these can be found in Tables 2 to 6) and it is the standard error of measurement of the difference between true scores that used to test if this difference is reliable. The formula (Silverstein, 1989) for this latter standard error is

$$\mathrm{SEM}_{Dt} = \sqrt{\mathrm{SEM}_{Xt}^2 + \mathrm{SEM}_{Yt}^2}\,. \tag{9}$$

These standard errors are reported in Table 7 and critical values for the difference between estimated true Index scores are presented in Table 8. Just as is the case for differences between obtained scores, an alternative is to divide the difference between estimated true scores by the relevant SEM_{Dt} and calculate a probability for the z thereby obtained (this is the method used by the computer program that accompanies this paper).

<u>Bonferroni correction when testing for reliable differences between Index scores</u>

Multiple comparisons are usually involved when testing if there are reliable differences between an individual's Index scores (as noted, there are six possible pairwise comparisons). Thus, if all comparisons are made, there will be a marked inflation of the Type I error rate. Although clinicians will often have an a priori hypothesis concerning a difference between two or more particular Index scores, it is also the case that often there is insufficient prior information to form firm hypotheses. Moreover, should a clinician wish to attend to a large, unexpected, difference in a client's profile then, for all intents and purposes, they should be considered to have made all possible comparisons.

One possible solution to the multiple comparison problem is to apply a standard Bonferroni correction to the *p* values. That is, if the family wise (i.e., overall) Type I error rate (α) is set at 0.05 then the *p* value obtained for an individual pairwise difference between two Indices would have to be less than 0.05/6 = to be considered significant at the specified value of alpha. This, however, is a conservative approach that will lead to many genuine differences being missed.

A better option is to apply a *sequential* Bonferroni correction (Larzelere & Mulaik, 1977). The first stage of this correction is identical to a standard Bonferroni correction. Thereafter, any pairwise comparisons that were significant are set aside and the procedure is repeated with $k - l$ in the denominator rather than *k*, where *l* = the number of comparisons recorded as significant at any previous stage. The process is stopped when none of the remaining comparisons achieve significance. This method is less conservative than a standard Bonferroni correction but ensures that the overall Type I error rate is maintained at, or below, the specified rate.

This sequential procedure can easily be performed by hand but, for convenience, the computer program that accompanies this paper offers a sequential Bonferroni correction as an option. Note that, when this option is selected, the program does not produce exact *p* values but simply records whether the discrepancies between Indices are significant at the .05 level after correction.

Abnormality of differences between Indices

In order to estimate the abnormality of a difference between Index scores it is necessary to calculate the standard deviation of the difference between each pair of Indices. When, as in the present case, the measures being compared have a common standard deviation, the formula for the standard deviation of the difference (Ley, 1972; Payne & Jones, 1957) is

$$\mathrm{SD}_D = s\sqrt{2-2r_{XY}}\,, \tag{10}$$

where *s* is the common standard deviation (i.e., 15 in the present case) and r_{XY} is the correlation between the two measures[1].

The standard deviations of the difference for the six pairings of Index scores are presented in Table 7. To calculate the size of difference between Index scores required for a specified level of abnormality the standard deviation of the difference for each pair of Indices was multiplied by values of *z* (standard normal deviates). The differences required to exceed the differences exhibited by various percentages of the healthy population are presented in Table 9. Two sets of percentages are listed – the first column records the size of difference required regardless of sign, the second column records difference required for a directional difference. To illustrate, suppose an individual obtains scores of 116 and 92 on the VC and PO Indices respectively; the difference between the Index scores is therefore 24 points. Ignoring the sign of the difference, it can be seen from Table 9 that this difference is larger than that required

[1] Note that this is an asymptotic method. That is, it does not consider the uncertainties involved in estimating the population mean and SD from normative sample data. Given the large size of the

(22) to exceed all but 10% of the population but is not large enough to exceed all but 5% of the population (difference required = 26 points). If the concern is with the percentage of the population expected to exhibit a difference *in favour* of VC, it can be seen that this difference is larger than that required (22) to exceed all but 5% of the population but is not large enough to exceed all but 1% (difference required = 31 points).

A closely related alternative to the approach outlined to is to divide an individual's difference by the standard deviation of the difference and refer the resultant z (z_D) to a table of areas under the normal curve (or algorithmic equivalent) to obtain a precise estimate of the percentage of the population expected to exhibit this large a difference. To continue with the current example, it is estimated that approximately 6% of the population would exhibit a difference of 24 points between VC and PO regardless of the sign of the difference and that approximately 3% would exhibit a difference of 24 points in favour of VC. This latter approach is that used in the computer program that accompanies the present paper (as was the case for reliable differences, these data are not presented in the present paper because they would require voluminous tables).

<u>Percentage of the population expected to exhibit *j* or more abnormally low Index scores and *j* or more abnormally large Index score differences</u>

Information on the rarity or abnormality of test scores (or test score differences) is fundamental in interpreting the results of a cognitive assessment (Crawford, 2004; Strauss, Sherman, & Spreen, 2006). When attention is limited to a single test, this information is immediately available; if an abnormally low score is

---

WAIS-III standardization sample its use is justifiable here. See Crawford and Garthwaite (2007) for a defined as one that falls below the 5th percentile then, by definition, 5% of the population is expected to obtain a score that is lower (in the case of Wechsler Indices, scores of 75 or lower are below the 5th percentile). However, the WAIS-III has four Indices and thus it would be useful to estimate what percentage of the healthy population would be expected to exhibit at least one abnormally low Index score. This percentage will be higher than for any single Index and knowledge of it is liable to guard against over inference; that is, concluding impairment is present on the basis of one "abnormally" low Index score if such a result is not at all uncommon in the general, healthy population. It is also useful to know what percentage of the population would be expected to obtain two or more, or three or more abnormally low scores; in general, it is important to know what percentage of the population would be expected to exhibit *j* or more abnormally low scores.

One approach to this issue would be to tabulate the percentages of the WAIS-III standardization sample exhibiting *j* or more abnormal Index scores. However, such empirical base rate data have not been provided for the full-length WAIS-III Indices, far less for short-forms. Crawford, Garthwaite and Gault (2007) have recently developed a generic Monte Carlo method to tackle problems of this type and have applied it to full-length WAIS-III Index scores. That is, they produced estimates of the percentage of the population expected to exhibit *j* or more abnormally low Index scores for a variety of different definitions of abnormality. We used this method (which requires the matrix of correlations between the short-form Index scores) to generate equivalent base rate data for the present WAIS-III short-form: three alternative definitions of what constitutes an abnormally low score were employed: a score below the 15th, 10th or 5th percentile. The results are presented in

---

full discussion of these issues and for optimal methods for normative samples with more modest *N*s

Table 10. If an abnormally low Index score is defined as a score falling below the 5th percentile (this is our preferred criterion and hence appears in bold) it can be seen that it will not be uncommon for members of the general population to exhibit one or more abnormally low scores from among their four Index scores (the base rate is estimated at 14.5% of the population); relatively few however are expected to exhibit two or more abnormally low scores (4.11%), and three or more abnormally low scores will be rare.

A similar issue arises when the interest is in the abnormality of pairwise differences between Indices; i.e. if an abnormally large difference between a pair of Indices is defined as, say, a difference exhibited by less than 5% of the population, then what percentage of the population would be expected to exhibit one or more of such differences from among the six possible pairwise comparisons? The base rates for this problem can also be obtained using Crawford et al's. (2007) Monte Carlo method and are presented in Table 11. To use these two tables the user should select their preferred definition of abnormality, note how many Index scores and /or Index score differences are exhibited by their client and refer to Tables 10 and/ or 11 to establish the base rate for the occurrence of these numbers of abnormal scores and score differences. The computer program accompanying this paper makes light work of this process: the user need only select a criterion for abnormality. The number of abnormally low scores and abnormally large differences exhibited by the case is then provided, along with the percentages of the general population expected to exhibit these numbers.

A global measure of the abnormality of an individual's Index score profile

Although not available for the full-length version of the WAIS-III, it would be useful to have a single measure of the overall abnormality of an individual's profile of scores; i.e., a multivariate index that quantifies how unusual a particular combination of Index scores is. One such measure was proposed by Huba (1985) based on the Mahalanobis distance statistic. Huba's Mahalanobis distance index (MDI) of abnormality of a case's profile of scores on *k* tests is

$$\mathbf{x}'\mathbf{W}^{-1}\mathbf{x}, \tag{11}$$

where **x** is a vector of *z* scores for the case on each of the *k* tests of a battery and $\mathbf{W}^{-1}$ is the inverse of the correlation matrix for the battery's standardization sample (the method requires the covariance matrix but the correlation matrix is the covariance matrix when scores are expressed as *z* scores). When this index is calculated for an individual's profile it is evaluated against a chi-square distribution on *k* df. The probability obtained is an estimate of the proportion of the population that would exhibit a more unusual combination of scores.

This method has been used to examine the overall abnormality of an individual's profile of *subtest* scores on the WAIS-R (Burgess, 1991; Crawford, 1994). However, it can equally be applied to an individual's profile of *Index* scores. Indeed we consider this usage preferable given that research indicates that analysis at the level of Wechsler factors (i.e., Indices) achieves better differentiation between healthy and impaired populations than analysis of subtest profiles (Crawford, Johnson, Mychalkiw, & Moore, 1997). The Mahalanobis Distance Index was therefore implemented for the WAIS-III short-form: This index estimates the extent to which a case's combination of Index scores, i.e., the profile of relative strengths and weaknesses, is unusual (abnormal). Note that it is not a practical proposition to calculate the MDI by hand, nor is it all practical to provide tabled values as there is a huge range of possible combinations of Index scores. Therefore the MDI for a case's profile of Index scores is provided only by the computer program that accompanies this paper.

A Computer Program for Scoring and Analysing the Index-Based Short-form

As noted, a computer program for PCs (SF_WAIS3.EXE) accompanies this paper. A compiled version of the program can be downloaded (as a zip file) from the following website address: http://www.abdn.ac.uk/~psy086/dept/sf_wais3.htm. The program implements all the procedures described in earlier sections. Although the paper contains all the necessary information to score and interpret an individual's short-form Index scores (with the exception of the MDI), the program provides a very convenient alternative for busy clinicians as it performs all the transformations and calculations (it requires only entry of the scaled scores on the subtests). The computer program has the additional advantage that it will markedly reduce the likelihood of clerical error. Research shows that clinicians make many more simple clerical errors than we like to imagine (e.g., see Faust, 1998; Sherrets, Gard, & Langner, 1979; Sullivan, 2000).

The program prompts for the scores on the seven subtests used in the short-form and allows the user to select analysis options. There is also an optional field for entry of user notes (e.g., date of testing, client details etc) for future reference.

The output first reproduces the subtest scores used to obtain the short-form Index scores, the analysis options selected, and user notes, if entered. Thereafter it reports the short-form Index scores with accompanying confidence limits and the scores expressed as percentiles (plus percentile confidence limits), followed by results from the analysis of the reliability and abnormality of differences between the individual's Index scores (including the base rates for the number of abnormal scores and score differences and the MDI of the abnormality of the Index score profile as covered in the two preceding sections). If the default options are not overridden the program generates 95% confidence limits on obtained scores, and tests for a reliable difference between observed scores without applying a Bonferroni correction. The results can be viewed on screen, edited, printed, and saved as a text file.

<u>Worked example of the use of the short-form</u>

To illustrate the use of the foregoing methods and the accompanying computer program, suppose that a patient (of high premorbid ability) who has suffered a traumatic brain injury obtains the following scaled scores on the seven subtests that comprise the short-form: Vocabulary = 13, Similarities = 12, Block Design = 12, Matrix Reasoning = 12, Arithmetic = 5, Digit Span = 6, Digit Symbol = 4. Suppose also that the psychologist opts for 95% confidence limits on obtained Index scores, chooses to examine the reliability of differences between observed (rather than estimated true scores), opts not to apply a Bonferroni correction (as would be appropriate when they have an a priori hypotheses concerning the pattern of strengths and weaknesses), and chooses to define an abnormally low Index score (and abnormally large difference between Index scores) as a difference exhibited by less than 5% of the normative population (these are the default options for the computer program).

The short-form Index scores, accompanying confidence limits and percentiles for this case (obtained either by using Tables 2 to 6 or the computer program) are presented in Figure 1a; this figure presents the results much as they appear in the output of the accompanying computer program. Note that, in addition to the 95%

limits on obtained scores, confidence limits are also expressed as percentile ranks. Examination of the Index scores reveals that the patient's Index scores on Processing Speed and Working Memory are abnormally low (they are at the 2^{nd} and 4^{th} percentile respectively). It can also be seen from Figure 1b that these two Indices are significantly (i.e., reliably) poorer than the patient's scores on both the Verbal Comprehension and Perceptual Organisation Indices. Thus, in this case, it is very unlikely that the differences between these Indices are solely the result of measurement error; that is, there are genuine strengths and weaknesses in the patient's profile.

This pattern is consistent with the effects of a severe head injury in an individual of high premorbid ability (Crawford, Johnson, Mychalkiw, & Moore, 1997). However, low scores and reliable differences on their own are insufficient grounds for inferring the presence of *acquired* impairments: a patient of modest premorbid ability might be expected to obtain abnormally low scores, and many healthy individuals will exhibit reliable differences between their Index scores. Therefore it is also important to examine the abnormality of any differences in the patient's Index score profile. In this case it can be seen from Figure 1c that the differences between the patient's PS and WM Index scores and his VC and PO scores are abnormal: that is, it is estimated that few healthy individuals would exhibit differences of this magnitude.

It can also be seen from Figure 1c that, applying the criterion that a difference exhibited by less than 5% of the population is abnormal, four of the patient's differences are abnormal (i.e., VC vs. PS, VC vs. WM, PO vs. PS, and PO vs. WM). Application of Crawford and Garthwaite's (2007) Monte Carlo method reveals that *very* few individuals in the normative population will exhibit this number of abnormal differences (0.16%). Moreover, the MDI, which provides a global measure of the abnormality of the patient's Index score profile, is highly significant (Chi square = 16.376, p = 0.00255). That is, the patient's overall profile is highly unusual. The results of analyzing this patient's scores converge to provide convincing evidence of marked acquired impairments in processing speed and working memory consistent with a severe head injury. As the inputs for this example (i.e. the subtest scores) and outputs (Figure 1) are all provided, it may be useful for clinicians to work through this example (using either the tables or the accompanying program) prior to using the short-form with their own cases.

Conclusion

In conclusion, we believe the WAIS-III short-form developed in the present paper has a number of positive features: it yields short-from Index scores (rather than IQs), it has good psychometric properties (i.e., high reliabilities and high validity), and offers the same useful methods of analysis as those available for the full-length version. The provision of an accompanying computer program means that (a) the short-form can be scored and analyzed very rapidly, and (b) the risk of clerical error is minimized. As clinicians working with children and adolescents have the same need for sound short-forms as those working with adult populations, it would be useful to develop an equivalent short-form for the recently released WISC-IV (Wechsler, 2003).

Finally, some clinicians or researchers will no doubt take issue with the particular subtests selected for the WAIS-III short-form. As the methods used to form, evaluate, score and analyze the short-form are stated explicitly this should allow others to develop alternative short-forms based on the same approach.

References

Anastasi, A. (1990). *Psychological testing* (6th ed.). New York: Macmillan.

Atkinson, L. (1991). Some tables for statistically based interpretation of WAIS-R factor scores. *Psychological Assessment, 3*, 288-291.

Burgess, A. (1991). Profile analysis of the Wechsler Intelligence Scales: A new index of subtest scatter. *British Journal of Clinical Psychology, 30*, 257-263.

Charter, R. A., & Feldt, L. S. (2001). Confidence intervals for true scores: Is there a correct approach? *Journal of Psychoeducational Assessment, 19*(4), 350-364.

Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 121-140). Chichester: Wiley.

Crawford, J. R., & Allan, K. M. (1994). The Mahalanobis distance index of WAIS-R subtest scatter: Psychometric properties in a healthy UK sample. *British Journal of Clinical Psychology, 33*, 65-69.

Crawford, J. R., Allan, K. M., & Jack, A. M. (1992). Short-forms of the UK WAIS-R: Regression equations and their predictive accuracy in a general population sample. *British Journal of Clinical Psychology, 31*, 191-202.

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology, 24*, 343-372.

Crawford, J. R., & Garthwaite, P. H. (Submitted). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks.

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21*, 419-430.

Crawford, J. R., Henry, J. D., Ward, A. L., & Blake, J. (2006). The Prospective and Retrospective Memory Questionnaire (PRMQ): Latent structure, normative data and discrepancy analysis for proxy-ratings. *British Journal of Clinical Psychology, 45*, 83-104.

Crawford, J. R., Johnson, D. A., Mychalkiw, B., & Moore, J. W. (1997). WAIS-R performance following closed head injury: A comparison of the clinical utility of summary IQs, factor scores and subtest scatter indices. *The Clinical Neuropsychologist, 11*, 345-355.

Crawford, J. R., Smith, G. V., Maylor, E. A. M., Della Sala, S., & Logie, R. H. (2003). The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample. *Memory, 11*, 261-275.

Faust, D. (1998). Forensic assessment. In *Comprehensive Clinical Psychology Volume 4: Assessment* (pp. 563-599). Amsterdam: Elsevier.

Glutting, J. J., Mcdermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement, 47*, 607-614.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.

Huba, G. J. (1985). How unusual is a profile of test scores? *Journal of*

Psychoeducational Assessment, 4, 321-325.

Larzelere, R. E., & Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychological Bulletin, 84,* 557-569.

Ley, P. (1972). *Quantitative aspects of psychological assessment*. London: Duckworth.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Payne, R. W., & Jones, G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13,* 115-121.

Reynolds, C. R., Willson, V. L., & Clark, P. L. (1983). A four-test short form of the WAIS-R for clinical screening. *Clinical Neuropsychology, 5,* 111-116.

Ryan, J. J., Lopez, S. J., & Werth, T. R. (1999). Development and preliminary validation of a Satz-Mogel short form of the WAIS-III in a sample of persons with substance abuse disorders. *International Journal of Neuroscience, 98*(1-2), 131-140.

Sherrets, F., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. *Psychology in the Schools, 16,* 495-496.

Silverstein, A. B. (1989). Confidence intervals for test scores and significance tests for test score differences: A comparison of methods. *Journal of Clinical Psychology, 45*, 828-832.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356-442). Washington D.C.: American Council on Education.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York: Oxford University Press.

Sullivan, K. (2000). Examiners' errors on the Wechsler Memory Scale-Revised. *Psychological Reports, 87*(1), 234-240.

Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting and Clinical Psychology, 31*, 499-506.

Tulsky, D., Zhu, J., & Ledbetter, M. F. (1997). *WAIS-III WMS-III technical manual*. San Antonio TX: Psychological Corporation.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale -Third Edition*. San Antonio TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (4th ed.)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D., Wycherley, R. J., Benjamin, L., Crawford, J. R., & Mockler, D. (1998). *Manual for the Wechsler Adult Intelligence Scale -Third Edition (U.K.)*. London: The Psychological Corporation.

Table 1. Summary statistics and basic psychometric properties of the Index-based short-form of the WAIS-III

Composite	Mean prior to transformation	SD prior to transformation	SEM of short-form Indices (and FSIQ)	SEM_t for true scores	Reliability Short-Form	Reliability Full-Length	*r* with full-length Indices
VC	20	5.628	3.674	3.454	0.94	0.96	0.97
PO	20	5.367	4.108	3.800	0.93	0.93	0.94
WM	20	5.231	4.025	3.735	0.93	0.94	0.95
PS	10	3.000	6.000	5.040	0.84	0.88	0.91
FSIQ	70	15.784	2.598	2.520	0.97	0.98	0.97

Note. VC = Verbal Comprehension Index; PO = Perceptual Organisation Index; WM = Working Memory Index; PS = Processing Speed Index; FSIQ = Full Scale IQ.

Table 2. Conversion of the sum of subtest scores (SSS) on the Vocabulary and Similarities subtests to **Verbal Comprehension (VC)** short-form Index scores.

				95% CLs	
SSS	Index Score	Est True score	Percentile	Lower limit	Upper limit
2	52	55	<0.1	45 (48)	59 (62)
3	55	57	0.1	47 (50)	62 (65)
4	57	60	0.2	50 (53)	65 (67)
5	60	62	0.4	53 (55)	67 (70)
6	63	65	0.7	55 (58)	70 (72)
7	65	67	1	58 (60)	73 (75)
8	68	70	2	61 (63)	75 (77)
9	71	72	3	63 (65)	78 (80)
10	73	75	4	66 (68)	81 (82)
11	76	77	5	69 (70)	83 (85)
12	79	80	8	71 (73)	86 (87)
13	81	82	10	74 (75)	89 (90)
14	84	85	14	77 (78)	91 (92)
15	87	87	19	79 (80)	94 (95)
16	89	90	23	82 (83)	97 (97)
17	92	92	30	85 (85)	99 (100)
18	95	95	37	87 (88)	102 (102)
19	97	97	42	90 (90)	105 (105)
20	100	100	50	93 (93)	107 (107)
21	103	103	58	95 (95)	110 (110)
22	105	105	63	98 (98)	113 (112)
23	108	108	70	101 (100)	115 (115)
24	111	110	77	103 (103)	118 (117)
25	113	113	81	106 (105)	121 (120)
26	116	115	86	109 (108)	123 (122)
27	119	118	90	111 (110)	126 (125)
28	121	120	92	114 (113)	129 (127)
29	124	123	95	117 (115)	131 (130)
30	127	125	96	119 (118)	134 (132)
31	129	128	97	122 (120)	137 (135)
32	132	130	98	125 (123)	139 (137)
33	135	133	99	127 (125)	142 (140)
34	137	135	99.3	130 (128)	145 (142)
35	140	138	99.6	133 (130)	147 (145)
36	143	140	99.8	135 (133)	150 (147)
37	145	143	99.9	138 (135)	153 (150)
38	148	145	>99.9	141 (138)	155 (152)

Note: Estimated true scores and 95% confidence limits on obtained Index scores are also provided (limits on true scores are in brackets) as is the percentile corresponding to each Index score.

Table 3. Table for converting the sum of subtest scores (SSS) on the Block Design and Matrix Reasoning subtests to **Perceptual Organisation (PO)** short-form Index scores.

				95% CLs	
SSS	Index Score	Est True score	Percentile	Lower limit	Upper limit
2	50	53	<0.1	42 (46)	58 (61)
3	52	56	<0.1	44 (49)	61 (63)
4	55	59	0.1	47 (51)	63 (66)
5	58	61	0.3	50 (54)	66 (69)
6	61	64	0.5	53 (56)	69 (71)
7	64	66	0.8	56 (59)	72 (74)
8	66	69	1	58 (62)	75 (76)
9	69	72	2	61 (64)	77 (79)
10	72	74	3	64 (67)	80 (82)
11	75	77	5	67 (69)	83 (84)
12	78	79	7	70 (72)	86 (87)
13	80	82	9	72 (74)	88 (89)
14	83	84	13	75 (77)	91 (92)
15	86	87	18	78 (80)	94 (95)
16	89	90	23	81 (82)	97 (97)
17	92	92	30	84 (85)	100 (100)
18	94	95	34	86 (87)	102 (102)
19	97	97	42	89 (90)	105 (105)
20	100	100	50	92 (93)	108 (107)
21	103	103	58	95 (95)	111 (110)
22	106	105	66	98 (98)	114 (113)
23	108	108	70	100 (100)	116 (115)
24	111	110	77	103 (103)	119 (118)
25	114	113	82	106 (105)	122 (120)
26	117	116	87	109 (108)	125 (123)
27	120	118	91	112 (111)	128 (126)
28	122	121	93	114 (113)	130 (128)
29	125	123	95	117 (116)	133 (131)
30	128	126	97	120 (118)	136 (133)
31	131	128	98	123 (121)	139 (136)
32	134	131	99	125 (124)	142 (138)
33	136	134	99.2	128 (126)	144 (141)
34	139	136	99.5	131 (129)	147 (144)
35	142	139	99.7	134 (131)	150 (146)
36	145	141	99.9	137 (134)	153 (149)
37	148	144	>99.9	139 (137)	156 (151)
38	150	147	>99.9	142 (139)	158 (154)

Note: Estimated true scores and 95% confidence limits on obtained Index scores are also provided (limits on true scores are in brackets) as is the percentile corresponding to each Index score.

Table 4. Table for converting the sum of subtest scores (SSS) on the Arithmetic and Digit Span subtests to **Working Memory (WM)** short-form Index scores.

				95% CLs	
SSS	Index Score	True score	Percentiles	Lower limit	Upper limit
2	48	52	<0.1	40 (45)	56 (59)
3	51	55	<0.1	43 (47)	59 (62)
4	54	57	0.1	46 (50)	62 (65)
5	57	60	0.2	49 (53)	65 (67)
6	60	63	0.4	52 (55)	68 (70)
7	63	65	0.7	55 (58)	71 (73)
8	66	68	1	58 (61)	73 (75)
9	68	71	2	61 (63)	76 (78)
10	71	73	3	63 (66)	79 (81)
11	74	76	4	66 (69)	82 (83)
12	77	79	6	69 (71)	85 (86)
13	80	81	9	72 (74)	88 (89)
14	83	84	13	75 (77)	91 (91)
15	86	87	18	78 (79)	94 (94)
16	89	89	23	81 (82)	96 (97)
17	91	92	27	84 (85)	99 (99)
18	94	95	34	86 (87)	102 (102)
19	97	97	42	89 (90)	105 (105)
20	100	100	50	92 (93)	108 (107)
21	103	103	58	95 (95)	111 (110)
22	106	105	66	98 (98)	114 (113)
23	109	108	73	101 (101)	116 (115)
24	111	111	77	104 (103)	119 (118)
25	114	113	82	106 (106)	122 (121)
26	117	116	87	109 (109)	125 (123)
27	120	119	91	112 (111)	128 (126)
28	123	121	94	115 (114)	131 (129)
29	126	124	96	118 (117)	134 (131)
30	129	127	97	121 (119)	137 (134)
31	132	129	98	124 (122)	139 (137)
32	134	132	99	127 (125)	142 (139)
33	137	135	99.3	129 (127)	145 (142)
34	140	137	99.6	132 (130)	148 (145)
35	143	140	99.8	135 (133)	151 (147)
36	146	143	99.9	138 (135)	154 (150)
37	149	145	>99.9	141 (138)	157 (153)
38	152	148	>99.9	144 (141)	160 (155)

Note: Estimated true scores and 95% confidence limits on obtained Index scores are also provided (limits on true scores are in brackets) as is the percentile corresponding to each Index score

Table 5. Table for converting scores on the Digit Symbol subtest to **Processing Speed (PS)** short-form Index scores.

				95% CLs	
SSS	Index Score	True score	Percentiles	Lower limit	Upper limit
1	55	62	0.1	43 (52)	67 (72)
2	60	66	0.4	48 (57)	72 (76)
3	65	71	1	53 (61)	77 (80)
4	70	75	2	58 (65)	82 (85)
5	75	79	5	63 (69)	87 (89)
6	80	83	9	68 (73)	92 (93)
7	85	87	16	73 (78)	97 (97)
8	90	92	25	78 (82)	102 (101)
9	95	96	37	83 (86)	107 (106)
10	100	100	50	88 (90)	112 (110)
11	105	104	63	93 (94)	117 (114)
12	110	108	75	98 (99)	122 (118)
13	115	113	84	103 (103)	127 (122)
14	120	117	91	108 (107)	132 (127)
15	125	121	95	113 (111)	137 (131)
16	130	125	98	118 (115)	142 (135)
17	135	129	99.0	123 (120)	147 (139)
18	140	134	99.6	128 (124)	152 (143)
19	145	138	99.9	133 (128)	157 (148)

Note: Estimated true scores and 95% confidence limits on obtained Index scores are also provided (limits on true scores are in brackets) as is the percentile corresponding to each Index score

Table 6a. Table for converting the sum of subtest scores (SSS) on all seven subtests to short-form **FSIQ** scores – Part 1

				95% CLs						95% CLs	
SSS	IQ	ETS	Pcile	L	U	SSS	IQ	ETS	Pcile	L	U
7	40	42	<0.1	35	45	43	74	75	4	69	79
8	41	43	<0.1	36	46	44	75	76	5	70	80
9	42	44	<0.1	37	47	45	76	77	5	71	81
10	43	45	<0.1	38	48	46	77	78	6	72	82
11	44	46	<0.1	39	49	47	78	79	7	73	83
12	45	47	<0.1	40	50	48	79	80	8	74	84
13	46	47	<0.1	41	51	49	80	81	9	75	85
14	47	48	<0.1	42	52	50	81	82	10	76	86
15	48	49	<0.1	43	53	51	82	82	12	77	87
16	49	50	<0.1	44	54	52	83	83	13	78	88
17	50	51	<0.1	45	55	53	84	84	14	79	89
18	51	52	<0.1	45	56	54	85	85	16	80	90
19	52	53	<0.1	46	57	55	86	86	18	81	91
20	52	54	<0.1	47	58	56	87	87	19	82	92
21	53	55	<0.1	48	59	57	88	88	21	83	93
22	54	56	0.1	49	59	58	89	89	23	84	94
23	55	57	0.1	50	60	59	90	90	25	84	95
24	56	58	0.2	51	61	60	90	91	25	85	96
25	57	59	0.2	52	62	61	91	92	27	86	97
26	58	59	0.3	53	63	62	92	93	30	87	97
27	59	60	0.3	54	64	63	93	94	32	88	98
28	60	61	0.4	55	65	64	94	94	34	89	99
29	61	62	0.5	56	66	65	95	95	37	90	100
30	62	63	0.6	57	67	66	96	96	39	91	101
31	63	64	0.7	58	68	67	97	97	42	92	102
32	64	65	0.8	59	69	68	98	98	45	93	103
33	65	66	1.0	60	70	69	99	99	47	94	104
34	66	67	1	61	71	70	100	100	50	95	105
35	67	68	1	62	72	71	101	101	53	96	106
36	68	69	2	63	73	72	102	102	55	97	107
37	69	70	2	64	74	73	103	103	58	98	108
38	70	71	2	64	75	74	104	104	61	99	109
39	71	71	3	65	76	75	105	105	63	100	110
40	71	72	3	66	77	76	106	106	66	101	111
41	72	73	3	67	78	77	107	106	68	102	112
42	73	74	4	68	78	78	108	107	70	103	113

Note: Estimated true scores (ETS) and 95% confidence limits on obtained FSIQ scores are also provided, as is the percentile corresponding to each score.

Table 6b. Table for converting the sum of subtest scores (SSS) on all seven subtests to short-form **FSIQ** scores – Part 2.

				95% CLs	
SSS	IQ	ETS	Pcile	L	U
79	109	108	73	103	114
80	110	109	75	104	115
81	110	110	75	105	116
82	111	111	77	106	116
83	112	112	79	107	117
84	113	113	81	108	118
85	114	114	82	109	119
86	115	115	84	110	120
87	116	116	86	111	121
88	117	117	87	112	122
89	118	118	88	113	123
90	119	118	90	114	124
91	120	119	91	115	125
92	121	120	92	116	126
93	122	121	93	117	127
94	123	122	94	118	128
95	124	123	95	119	129
96	125	124	95	120	130
97	126	125	96	121	131
98	127	126	96	122	132
99	128	127	97	122	133
100	129	128	97	123	134
101	129	129	97	124	135
102	130	129	98	125	136
103	131	130	98	126	136
104	132	131	98	127	137
105	133	132	99	128	138
106	134	133	99	129	139
107	135	134	99.0	130	140
108	136	135	99.2	131	141
109	137	136	99.3	132	142
110	138	137	99.4	133	143
111	139	138	99.5	134	144
112	140	139	99.6	135	145
113	141	140	99.7	136	146
114	142	141	99.7	137	147
115	143	141	99.8	138	148
116	144	142	99.8	139	149
117	145	143	99.9	140	150
118	146	144	99.9	141	151
119	147	145	>99.9	141	152
120	148	146	>99.9	142	153
121	148	147	>99.9	143	154
122	149	148	>99.9	144	155
123	150	149	>99.9	145	155
124	151	150	>99.9	146	156
125	152	151	>99.9	147	157
126	153	152	>99.9	148	158
127	154	153	>99.9	149	159
128	155	153	>99.9	150	160
129	156	154	>99.9	151	161
130	157	155	>99.9	152	162
131	158	156	>99.9	153	163
132	159	157	>99.9	154	164
133	160	158	>99.9	155	165

Note. Estimated true scores (ETS) and 95% confidence limits on obtained FSIQ scores are also provided, as is the percentile corresponding to each score.

Table 7. Standard errors of measurement of the difference for observed scores and true scores, and standard deviations of the difference between short-form Indices.

Indices	SEM_D for observed scores	SEM_D for true scores	SD of the difference
VC and PO	5.511	5.135	12.97
VC and WM	5.450	5.087	13.11
VC and PS	7.036	6.110	15.76
PO and WM	5.751	5.328	13.26
PO and PS	7.272	6.312	14.28
WM and PS	7.225	6.273	14.28

Table 8. Critical values (two-tailed) for determining the reliability of differences between short-form Indices using either observed scores or estimated true scores.

	Critical values for **observed** scores				Critical values for estimated **true** scores			
p	.15	.10	.05	.01	.15	.10	.05	.01
VC and PO	7.94	9.07	10.80	14.20	7.39	8.45	10.06	13.23
VC and WM	7.85	8.97	10.68	14.04	7.33	8.37	9.97	13.10
VC and PS	10.13	11.57	13.79	18.12	8.80	10.05	11.98	15.74
PO and WM	8.28	9.46	11.27	14.81	7.67	8.76	10.44	13.72
PO and PS	10.47	11.96	14.25	18.73	9.09	10.38	12.37	16.26
WM and PS	10.40	11.89	14.16	18.61	9.03	10.32	12.30	16.16

Table 9. Difference between short-form Indices required to exceed various percentage of the healthy population.

	Difference required to exceed specified percentage of population –**absolute difference**				Difference required to exceed specified percentage of population –**directional difference**			
	15%	10%	5%	1%	15%	10%	5%	1%
VC and PO	19	22	26	34	14	17	22	31
VC and WM	19	22	26	34	14	17	22	31
VC and PS	23	26	31	41	17	21	26	37
PO and WM	20	22	26	35	14	17	22	31
PO and PS	23	26	31	41	17	21	26	37
WM and PS	23	26	31	41	17	21	26	37

Table 10. Percentage of the normal population expected to exhibit at least *j* abnormally low Index scores on the short-form WAIS-III; three definitions of abnormality are used ranging from below the 15th percentile to below the 5th percentile

	Percentage exhibiting *j* or more abnormally low WAIS-III short-form Index scores			
Criterion for abnormality	1	2	3	4
<15th	34.77	16.01	7.13	2.21
<10th	25.10	9.96	3.95	1.06
<5th	**13.98**	**4.34**	**1.43**	**0.31**

Table 11. Percentage of the normal population expected to exhibit *j* or more abnormal pairwise differences, regardless of sign, between short-form Index scores on the WAIS-III; three definitions of abnormality are used ranging from a difference exhibited by less than 15% of the population to a difference exhibited by less than 5%.

	Percentage exhibiting *j* or more abnormal pairwise differences (regardless of sign) between WAIS-III short-form Indices					
Criterion for abnormality	1	2	3	4	5	6
<15%	47.22	28.08	12.37	2.14	0.15	0.00
<10%	35.10	17.74	6.33	0.79	0.04	0.00
<5%	**20.15**	**7.72**	**1.99**	**0.16**	**0.00**	**0.00**

Figure Legends

Figure 1

Illustrative example of results from applying the WAIS-III short-form

(a)

Short-Form Index scores plus confidence limits (score is also expressed as a percentile):

Index	Score	(95% CI Traditional)	Percentile	(95% CI)
Verbal Comprehension	113	106 to 121	81.3	65.8 to 91.4
Perceptual Organization	111	103 to 119	77.2	58.3 to 90.0
Working Memory	74	66 to 82	4.3	1.2 to 11.6
Processing Speed	70	58 to 82	2.3	0.3 to 11.2
FSIQ	94	89 to 99	35.2	23.6 to 48.4

NUMBER of case's Index scores classified as abnormally low = 2

PERCENTAGE of normal population expected to exhibit this number or more of abnormally low scores: Percentage = 4.34%

(b)

RELIABLITY of differences between Short-Form Indices:

Index Pair	Difference	Two-tailed *p*	One-tailed *p*
VC versus PO	2	0.697	0.348
VC versus WM	39	0.000	0.000
VC versus PS	43	0.000	0.000
PO versus WM	37	0.000	0.000
PO versus PS	41	0.000	0.000
WM versus PS	4	0.562	0281

(c)

ABNORMALITY of differences between Short-Form Indices, i.e., percentage of population estimated to obtain a larger difference in same direction (figure in brackets is percentage regardless of sign) :

Index Pair	Difference	%age of populaion	(%age regardless of sign)
VC versus PO	2	43.975%	(86.957%)
VC versus WM	39	0.142%	(0.284%)
VC versus PS	43	0.299%	(0.598%)
PO versus WM	37	0.265%	(0.530%)
PO versus PS	41	0.434%	(0.867%)
WM versus PS	4	39.465%	(78.931%)

NUMBER of case's pairwise differences (regardless of sign) that meet criterion for abnormality = 4

PERCENTAGE of normal population expected to exhibit this number or more of abnormal differences=0.16%

MAHALANOBIS DISTANCE Index of the overall abnormality of the case's Index score profile: Chi-square = 16.317, *p* value = 0.00262