The Unitary Ability of IQ and Indexes in WAIS-IV

Abstract

Lichtenberger and Kaufman (2009, p. 167) defined unitary ability as 'an ability […] that is represented by a cohesive set of scaled scores, each reflecting slightly different or unique aspects of the ability'. Flanagan and Kaufman (2009) and Lichtenberger and Kaufman (2012) used a difference of 23 IQ points between the highest score (Max) and the lowest score (Min) obtained by a subject in the four Indexes of the WAIS-IV to define unitarity of the total IQ score. A similar method has been used to assess the unitary ability of the four Indexes, with a threshold of 5 points. Such difference scores (of 23 for IQ and 5 for Indexes) are considered high and infrequent and the authors therefore conclude that the corresponding Full-Scale IQ score or Index score is uninterpretable. In this paper we argue that these thresholds are inappropriate because they are based on the wrong standard deviation. The main aim of this study was to establish variability thresholds for IQ and the WAIS-IV Indexes for the American standardization sample and to compare these thresholds with those for the Italian standardization sample. We also consider an alternative approach to determining whether an IQ score represents a unitary ability, based on the maximum difference score for the 10 core subtests that contribute to Full-Scale IQ scores.

Keywords: Intelligence; WAIS-IV; IQ unitary; Indexes unitary; Max-Min difference

The Unitary Ability of IQ and Indexes of WAIS-IV

The Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008a, b) is an individually administered clinical instrument which is used to assess intelligence in individuals aged from 16 years 0 months to 90 years 11 months. The WAIS-IV consists of 15 subtests, 10 core and 5 supplementary. All subtests use weighted scores with *M* = 10, *SD* =3. The WAIS-IV also provides composite scores representing intellectual functioning in four cognitive domains (Verbal Comprehension Index - VCI; Perceptual Reasoning Index - PRI; Working Memory Index - WMI; Processing Speed Index - PSI) and an overall composite score representing general intellectual ability (Full-Scale IQ). All composite scores (for the four main Indexes and Full-Scale IQ) are standardized with *M* = 100, *SD* = 15. The VCI comprises three core subtests (Similarities; Vocabulary; Comprehension) and one supplementary subtest (Information); the PRI comprises three core subtests (Block Design; Matrix Reasoning; Visual Puzzles) and two supplementary subtests (Figure Weights; Picture Completion); the WMI comprises two core subtests (Digit Span; Arithmetic) and one supplementary subtest (Letter-Number Sequencing); the PSI comprises two core subtests (Symbol Search; Coding) and one supplementary subtest (Cancellation). All 10 core subtests should be administered so that all the five composite scores (CVI, PRI, WMI, PSI, Full-Scale IQ) can be computed.

Lichtenberger and Kaufman (2009, p. 167) defined unitary ability as ‘an ability […] that is represented by a cohesive set of scaled scores, each reflecting slightly different or unique aspects of the ability’. We can therefore conclude that that if the variance in scores on the subtests which contribute to a given WAIS index is not unusually high, that WAIS index represents a unitary ability and the score can be interpreted accordingly.

The same logic can be applied to Full-Scale IQ. In individual cases one can assess whether Full-Scale IQ score represents a unitary ability by comparing the difference between an individual's highest (Max) and lowest (Min) scores for the four WAIS-IV Indexes (VCI, PRI, WMI and PSI) with population norms. If the difference score is high and infrequent in the general population, then the associated Full-Scale IQ score is considered uninterpretable. A similar method is then used to determine whether the scores on the four Indexes represent unitary abilities, in this case however the difference score is based on Max and Min scores for the subtests which contribute to that Index.

Lichtenberger and Kaufman suggested that a threshold value for maximum difference score could be computed as 1.5**SD*, where 1.5 is the *z*-score associated with the proportion of area equal to 6.7% (on one tail) under the normal curve distribution (according to these authors the 6.7% could be defined as the value expressing the infrequency or rarity in the normal population), and SD is the standard deviation to be used. The latter would be (a) 15 (which is the standard deviation of standard scores with $M = 100$ and $SD = 15$) when dealing with Max-Min difference scores between the four Indexes which contribute to Full Scale IQ, or (b) 3 (which is the standard deviation of the weighted scores of the subtests, with $M = 10$ and $SD = 3$) when dealing with Max-Min difference scores for the subtests contributing to an Index.

It follows that if we are using Max-Min difference scores to assess whether the Full-Scale IQ represents a unitary ability the threshold is 1.5*15 = 22.5; this was rounded up to 23 by Lichtenberger and Kaufman. If, however, we are interested in using difference scores to assess the extent to which an Index score represents a unitary ability then the threshold is 1.5*3 = 4.5 (rounded up to 5 by Lichtenberger and Kaufman). As an example, consider a participant who obtains the following set of scores on the WAIS-IV: VCI = 125, PRI = 115, WMI = 113 and PSI = 98. In this case the highest score was obtained on the VCI (125) and the lowest on the PSI (98), giving a difference score of 27 (125 – 98) which is greater than the threshold (23), indicating that the Full-Scale IQ should be considered uninterpretable.

Similarly, if a participant obtains the following scores for the VCI subtests: Similarities = 12, Vocabulary = 10 and Comprehension = 6, the difference score would be 6 (12 – 6), which is, once again, greater than the threshold (5) and therefore indicates that the participant's VCI score should be considered uninterpretable.

However, the formula used by Lichtenberger and Kaufman (2009) to compute these thresholds is wrong. Their formula is easy to use and is supposed to be valid for all Wechsler scales regardless of the version and standardization sample, and it has been used accordingly, even with the WISC-IV (Flanagan & Kaufman, 2009). Such formula is independent of the number of Indexes or subtests on which the Max-Min difference score is based. The same threshold 23 is used to assess Max-Min differences between scores contributing to Full-Scale IQ (Lichtenberger & Kaufman, 2009, p. 153) and differences between two Indexes which contribute to the General Ability Index (p. 155) or the Cognitive Proficiency Index (p. 157).

The threshold of 5 used for assessing Max-Min differences between the subtests which contribute to the Indexes is treated in the same way. It is used to assess differences when three subtests contribute an Index score, as with the VCI and PRI (Lichtenberger & Kaufman, 2009, p. 163) but also when only two subtests are involved, as with the WMI and PSI (Lichtenberger & Kaufman, 2009, p. 168). Following this logic, it would be legitimate to use the same threshold, 5, to assess Max-Min difference scores for the WAIS-IV composite score, which is based on 10 core subtests.

Orsini, Pezzuti, and Hulbert (2014) criticised the use of these formulae and demonstrated that the thresholds proposed by Flanagan and Kaufman (2009) should not be applied to the WISC-IV. This is because the standard deviation value of 15, which Flanagan and Kaufman used to delimit maximum differences for which Full-Scale IQ represents a unitary ability, is based on standard scores (where M = 100 and SD = 15) and is therefore the standard deviation of the distribution standard score of Full-Scale IQ or of Indexes. However, it is more appropriate to evaluate the amplitude and frequency of a Max-Min difference scores in terms of the distribution of Max-Min difference scores, so a threshold should be based one the mean and standard deviation of Max-Min distribution, rather than the distribution of the standard scores.

Orsini, Pezzuti, and Hulbert (2014) proposed that the threshold for determining whether Full-Scale IQ scores, or scores on the main four Indexes, represent unitary abilities should be defined as $M_{\text{Max-Min}} + z*SD_{\text{Max-Min}}$, where mean and standard deviation refer to the distribution of Max-Min difference scores. The *z*-score used will depend on the choice of rarity criterion and will therefore vary from case to case. For instance, when $z = 1.5$, the threshold will exclude the bottom 6.7% of the population, whereas using $z = 1.28$ would exclude the bottom 10% of the population, and so on. Using the characteristics of the Studentised range, Orsini, Pezzuti, and Hulbert (2014) estimated the *M* and *SD* of the Max-Min difference scores for the American standardization sample for the WISC-IV and showed that a threshold of 23, as suggested by Flanagan and Kaufman (2009), excluded 44% of the sample (and not 6.7%) and was therefore inappropriate. Similarly, applying a threshold of 23 to the Italian standardization sample for the WAIS-IV (Orsini & Pezzuti, 2013) would have resulted in exclusion of 50.4% of the participants, again far too many for the threshold to be considered acceptable in terms of the rarity of non-unitary IQ scores.

The main aim of this research was therefore to establish appropriate thresholds of infrequency for the Full-Scale IQ and the main four Indexes of the WAIS-IV in the American standardization sample and to compare such thresholds with those for the Italian standardization sample.

It was, however, considered appropriate to include the above-mentioned computations when dealing with the 10 core subtests that contribute to the Full Scale IQ of the WAIS-IV because an IQ score is calculated as the sum of individual subtest scores, not as the sum of the four index scores. Although we use Max-Min difference scores for the four Indexes to assess the extent to which an IQ score represents a unitary ability there are alternative approaches. For example one could use the Max-Min difference scores for the 10 core subtests, as they are already used in the assessment of the unitary ability of the various Indexes.

Method

Sample and Instruments

We used the WAIS-IV Italian standardization sample (Orsini & Pezzuti, 2013; Wechsler, 2013), which consisted of 1424 subjects subdivided into 9 age groups ranging from 16 to 69 years old. Again, we used data on the American sample (n = 1800) published in the Technical and Interpretive Manual of the WAIS-IV (Wechsler, 2003b).

Statistical Analyses

To achieve the main aim of this paper we implemented two methodological points.

Point 1. Our starting point was the following statement by Silverstein (1987, p. 410): ‘If x_1 and x_2 are the lowest (Min) and the highest (Max) values, respectively, in a sample of n independent observations, the range of the sample in standardized form is $W = (x_n – x_1)/\sigma$, where σ is the standard deviation of X’. Miller (1981) suggested that in cases where individual scores are correlated a correction can be implemented by multiplying the denominator by $\sqrt{1-\rho}$; the scaled-score range of the standardization sample is then estimated by computing the product of the critical value at the 1-p probability level (Owen, 1962, Table 6.1) and $\sigma\sqrt{1-\rho}$. It should be noted that σ = 15 for the Indexes of the WAIS-IV, and ρ is the average correlation between Indexes.

Point 2. The product of E(W) and $\sqrt{\sigma^2(W)}$ (Owen, 1962, Table 6.2) is used to provide estimates of the mean and standard deviation of the distribution Max-Min of scores respectively.

Points 1 and 2 should address two different but related issues in both the American and Italian standardization samples: 1) they allow estimation of the frequency with which a particular Max-Min difference score occurs, and 2) they allow estimation of the means and standard deviations of the distribution of Max-Min difference scores. In the case of the Italian standardization sample we also present both statistical analyses for the observed data.

More specifically:

1) In order to estimate which Max-Min difference score for the four Indexes is associated with a specific frequency of occurrence, we used the following formula, suggested by Silverstein (1989):

$$R = q * \sigma \sqrt{1 - 2\sum r_{ij} / k(k-1)}$$

where:

- R is the threshold Max-Min score for a given α level;
- q is the critical value of the Studentised range (for $\alpha = 0.01 \ldots 0.90$ in increments of 0.01); the values of q for each α value are taken from tables provided by Harter and Balakrishnan (1998) with $k = 4$ (where k is the number of Indexes being compared e.g. $k = 4$ for the WAIS-IV (VCI, PRI, WMI, and PSI)) and ∞ is the number of degrees of freedom (df) for large samples;
- σ is the common standard deviation of the Indexes and is equal to 15;
- r_{ij} are intercorrelations reported in Table A.1 (Wechsler, 2003b) for the American sample and in Table 3-12 (Orsini & Pezzuti, 2013) for the Italian sample.

A similar procedure was used to estimate the values of *R* for Max-Min difference scores for the subtests contributing to the VCI and PRI. The values of *q* for each α value are taken from the tables provided by Harter and Balakrishnan (1998), with common standard deviation equal to 3, and ∞ df for large samples. The intercorrelations between subtests (r_{ij}) are taken from Table A.1 (Wechsler, 2003b) for the American sample and Table 3-12 (Orsini & Pezzuti, 2013) for the Italian sample.

2) In order to estimate the mean and standard deviation of the Max-Min difference scores for the four Indexes, which are used to assess the extent to which an IQ score represents a unitary ability, we used the method described by Silverstein (1987, 1988) and Owen (1962), which computes the product of $\sigma\sqrt{1-\rho}$ (where $\sigma = 15$ is the standard deviation of the Indexes and ρ is the average correlation between Indexes) as E(W) and $\sqrt{\sigma^2(W)}$ (Owen, 1962, Table 6.2).

A similar procedure was used to estimate the mean and standard deviation of the distribution of the Max-Min difference scores for the subtests contributing to VCI and PRI. In these analyses the standard deviation of the subtests was $\sigma = 3$, and ρ was the average intercorrelation between subtests.

We also used an alternative approach to assess the extent to which Full-Scale IQ scores represent a unitary ability, namely applying a threshold to the Max-Min difference scores for the 10 core subtests which contribute to the IQ.

Results

Table 1 shows the *R* values (or the estimated threshold Max-Min values that are in third and fourth columns), that is, the estimated Max-Min difference scores for the four Indexes in the cases of both the American and Italian samples for each α level. For the Italian sample we also show, in the fifth column, the observed values of Max-Min distribution.

According to Table 1 more than 30% of the participants in the American standardization sample (see α in the first column corresponding to estimated *R* values between 22.0 and 24.6) and more than 40% in the Italian standardization sample (see α in the first column corresponding *R* values between 22.2 and 24.8) scored above threshold suggested by Lichtenberger and Kaufman (2009), and could therefore be considered as having a non-interpretable IQ. For the Italian sample the result is confirmed by the analysis of the observed data (see fifth column).

INSERT TABLE 1 ABOUT HERE

Tables 2, 3 and 4 show the results of similar computations for the three subtests of the VCI and the PRI and for the 10 core subtests respectively. Considering, for example, the VCI (see Table 2), it would look as if more than 5% of the participants in the American standardization sample (see α in the first column, corresponding estimated *R* values were between 4.8 and 5.5) and more than 10% in the Italian standardization sample (see α in the first column, corresponding *R* values were between 4.6 and 5.5) scored above the threshold (5) and could therefore be considered as having a non-interpretable VCI.

INSERT TABLES 2, 3 and 4 ABOUT HERE

Table 5 reports *M*s and *SD*s of the estimated difference scores for the four WAIS-IV Indexes, together with the observed values from the Italian standardization sample (Orsini, Pezzuti, 2014). From these it was possible to estimate the exact percentage of participants for whom the maximum difference score is at or above the 23-point threshold suggested by Lichtenberger and Kaufman (2009) by transforming this score into a *z*-score using the *M* and

SD estimated for the American sample: $z = (23 - 20.49)/8.76 = 0.29$. The percentage of participants with a *z*-score above this *z* threshold can then be obtained from tables of the normal distribution; in this case it is 38.6% of the population.

INSERT TABLE 5 ABOUT HERE

Table 6 summarises the results obtained so far and also includes reference to the thresholds for different frequency criteria. Users can choose what rarity criterion to apply when deciding whether or not a score, be it a Full-Scale IQ score or a VCI or PRI index score, represents a unitary ability.

An alternative approach to evaluating whether an IQ score represents a unitary ability would be to use the Max-Min difference scores for the 10 core subtests. We analysed the observed data for the Italian sample and found that in order to classify 6.7% of subjects as having IQ scores which do not represent a unitary ability (this requires a rarity criterion of $z = 1.5$) we need a difference score threshold of 11 (see eighth column in Table 6). This means that a person whose maximum score for the 10 core subtests is 11 or more points different from his or her minimum score for the 10 core subtests shows a pattern of performance which is too variable for the Full-Scale IQ score to be considered as representing a unitary ability; such level of variability across the core subtests occurs in only 6.7% Italians.

INSERT TABLE 6 ABOUT HERE

The above analyses were used to assess the extent to which Full-Scale IQ scores and scores on two Indexes (VCI and PRI) represent unitary abilities, but we have not considered the WMI and PSI as both comprise only two subtests, and individual users should easily be able to compute the mean and standard deviation of the difference score for the two contributing subtests by referring to the means, standard deviations and pairwise correlations between subtests given in Table A-1 (Wechsler, 2003b). For instance, in the American standardization sample both the Digit Span and Arithmetic subtests have *M* = 10.0, *SD* = 3.0, and the correlation between them is *r* = 0.60. The necessary computations will regard the mean and standard deviation of difference between Digit Span and Arithmetic:

M_{dif} = 10.0 – 10.0 = 0

SD_{dif} = square root ($3^2 + 3^2$ – 2 x 0.60 x 3 x 3) = 2.68

It is then easy to calculate the threshold at which a composite score ceases to represent a unitary ability. The difference score for Digit Span and Arithmetic which corresponds to *z* = 1.5 (6.7%) is computed as: M_{dif} + 1.5(SD_{dif}) = 0 + 1.5 (2.68) = 4.18, which can be rounded to 5. Thus the WMI score for a person whose scores on the Digit Span and Arithmetic subtests differ by 5 or more would be considered uninterpretable.

Conclusions

If we decide to use the amplitude and rarity of the maximum difference score to determine whether composite scores, such as those for Full-Scale IQ and the four main WAIS-IV Indexes, represent unitary abilities, it seems obvious that the method proposed Lichtenberger and Kaufman (2012) is not fit for purpose. In the formula 1.5*SD*, the *z*-value of 1.5 is an acceptable rarity criterion, but the standard deviation of the standard scores does not constitute an acceptable way of defining an amplitude criterion. If we enter the realm of the standard score distribution, we get stuck in it and the only conclusions we can draw from the results of the formula (100+1.5*15 = 122.5, rounded up to 123) are limited to the IQ score of 123, which has a percentile rank of 93.3%, and which is the only one that cuts off 6.7% of the population; no conclusions can be drawn about the distribution of Max-Min difference scores. It is worth remembering that because we are dealing with the distribution of the Max-Min difference scores we should refer to the standard deviation of this distribution, rather than the distribution of the standard scores. If we set a rarity criterion of 10% (i.e. we decide to consider that composite scores for which the maximum difference score is in the top 10% of the population do not represent a unitary ability), then the threshold for the American standardization sample becomes 32, which is slightly lowest than the corresponding threshold 36 for the Italian standardization sample (see Table 1). In the case of the Italian sample, the thresholds identified using Silverman's method (1989) are very similar to those derived from the observed data.

As already noted by Orsini, Pezzuti, and Hulbert (2014, p. 175), it is instantly clear that the use of an appropriate threshold 'has an immediate effect on both the individual diagnostic as well as on the study of clinical groups.' Research by Liratni and Pry (2007, 2012) supports our conclusions; these authors reported that when the Max-Min difference threshold of 23 was applied to a sample of gifted participants, approximately 87% of the sample had an IQ that was classified as uninterpretable. More specifically, their percentage (87%) demonstrates that the cut-off value (23 points) was not adequate, but not that gifted participants are more heterogeneous (as it is frequently assumed).

Finally, we agree with an anonymous reviewer of this manuscript who asserted that even when the Max-Min difference is large and infrequent, the associated IQ score is to be considered interpretable. We consider that a composite score is always interpretable, whether it represents a cohesive, unitary set of abilities or not. In fact because a composite score includes both common and specific variance, and because discrepant performance in one subtest could be due mainly to specific variance, a composite score always represents common variance.

A lack of cohesion among the subtests contributing to a composite score means only that practitioners should extend their interpretation so as to explain the lack of cohesion.

This study's contribution to the literature is the confirmation, at international level, that we can not use maximum difference score thresholds of 23 and 5 (for IQ and Indexes), and that the appropriate thresholds will be different in different populations, because there will be different Max-Min difference distributions. This article provides Italian practitioners with thresholds which can be applied to difference scores for the subtests contributing to the main index scores and to Full-Scale IQ scores. We have also proposed a method for evaluating the variability of scores on the 10 core subtests which contribute to the Full-Scale IQ. This can be used by practitioners to detect high variability between cognitive abilities and may prompt investigations into the cause or causes of this variability. Occasionally, practitioners might consider using other Indexes such as the General Ability Index (GAI; Prifitera, Weiss & Saklofske, 1998) and the Cognitive Proficiency Index (CPI; Dumont, Willis, 2001). The GAI is a measure of crystallized and fluid intelligence, derived from the VCI and the PRI, whereas the CPI is a measure of memory and processing speed and is derived from the sum of the WMI and the PSI scores. Clinicians can use low performance scores on one or both of these indexes and the comparison between them to gain important insights into the cognitive deficit of participants.

References

Flanagan, D. P., Kaufman, A. S. (2009).*Essentials of WISC-IV Assessment*-2nd Ed., John Wiley & Sons. Hoboken, NJ: (Trad. It., *Fondamenti per l'assessment con la WISC-IV*. Giunti-OS. Firenze, 2012).

Harter, H. L., Balakrishnan, N. (1998). *Tables for the Use of Range and Studentized Range in Tests of Hypotheses*. CRC Press. Boca Raton. New York

Lichtenberger, E. O., Kaufman, A. S. (2009). *Essentials of WAIS-IV Assessment*. John Wiley & Sons, Inc. Hoboken, New Jersey.

Lichtenberger, E. O., Kaufman, A. S. (2012). *Essentials of WAIS-IV Assessment. Second Edition.* John Wiley & Sons, Inc. Hoboken, New Jersey.

Liratni, M., Pry, R. (2007). Psychometrie et WISC-IV: quel avenir pour l'identification des enfants à haut potentiel intellectuel? *Neuropsychiatrie de l'Enfance et de l'Adolescence, 55*, 214-219.

Liratni, M., Pry, R. (2012). Profils psicométriques de 60 enfants à haut potentiel au WISC-IV. *Pratiques Psychologiques, 18,* 63-74.

Miller, R. G. (1966). Simultaneous Statistical Inferente. New York. Mc Graw & Hill.

Orsini, A., Pezzuti, L. (2013). *WAIS-IV. Contributo alla taratura italiana*. Firenze. Giunti OS.

Orsini, A., Pezzuti, L., Hulbert, S. (2014). The unitary ability on IQ in the WISC-IV and its computation. *Personality and Individual Differences*, 69, 173-175. DOI: 10.1016/j.paid.2014.05.023.

Owen, D. B. (1962). *Handbook of statistical tables,* Reading. MA. Addison-Wesley.

Prifitera, A., Weiss, L.G., & Saklofske, D.H. (1998). The WISC-III in context. In A. Prifitera & D. H. Saklofske (Eds.), WISC-III clinical use and interpretation: Scientist-practitioner perspectives (pp. 1-38). San Diego, CA: Academic Press.

Silverstein, A. B. (1987). Two Indexes of subtest scatter on Wechsler's intelligence scales: estimated vs. empirical values. *Journal of Clinical Psychology, 43*, 409-414.

Silverstein, A. B. (1988). Estimated vs. empirical values of scaled score range on Wechsler's intelligence scales: a correction. *Journal of Clinical Psychology, 44*, 259-260.

Silverstein, A. B. (1989). Reliability and Abnormality of Scaled-Score Ranges. *Journal of Clinical Psychology*, *45*, 926-929.

Wechsler, D. (2008a). *WAIS-IV administration and scoring manual*. San Antonio. TX: The Psychological Co0rporation.

Wechsler, D. (2008b). *WAIS-IV. Technical and Interpretive Manual*.San Antonio. TX.: The Psychological Corporation.

Wechsler, D. (2013). *WAIS-IV. Manuale di somministrazione e scoring*. Firenze. Giunti OS.

Table 1 - Computational steps using Silverstein' formula (1989) on the data from the USA and Italian standardization sample of the WAIS-IV for IQ scores.

		R (4 Indexes)		
α	$q_{4,\infty}$	USA (est)	ITALY (est)	ITALY (emp)
0.01	4.403	43.8	49.4	50
0.05	3.633	36.1	40.8	41
0.10	3.240	32.2	36.4	36
0.20	2.784	27.2	31.3	31
0.30	2.469	24.6	27.7	27
0.40	2.210	22.0	24.8	24
0.50	1.978	19.7	22.2	22
0.60	1.757	17.5	19.7	19
0.70	1.531	15.2	17.2	17
0.80	1.286	12.8	14.4	14
0.90	0.979	9.7	11.0	11

Legenda. q is the critical value of the studentized range (for α = 0.01 – 0.05 – 0.10 - 0.20 – 0.30 – 0.40 – 0.50 – 0.60 – 0.70 – 0.80 – 0.90); the values of q for each α value are taken from the tables provided by Harter and Balakrishnan (1998) with k=4 (4 Indexes) and ∞ degree of freedom (df); USA (est) are the estimate thresholds for USA sample; ITALY (est) are the estimate thresholds for ITALY sample; ITALY (emp) are the empiric thresholds for ITALY sample.

Table 2 - Computational steps using Silverstein' formula (1989) on the data from the USA and Italian standardization sample of the WAIS-IV for VCI scores.

		R (3 subtests VCI)		
α	$q_{3,\infty}$	USA (est)	ITALY (est)	ITALY (emp)
0.01	4.120	6.8	7.8	8
0.05	3.314	5.5	6.3	6
0.10	2.902	4.8	5.5	5
0.20	2.424	4.0	4.6	4
0.30	2.095	3.5	4.0	4
0.40	1.826	3.0	3.5	3
0.50	1.588	2.6	3.0	3
0.60	1.363	2.2	2.6	2
0.70	1.138	1.9	2.2	2
0.80	0.900	1.5	1.7	2
0.90	0.618	1.0	1.2	1

Legenda. q is the critical value of the studentized range (for α = 0.01 – 0.05 – 0.10 - 0.20 – 0.30 – 0.40 – 0.50 – 0.60 – 0.70 – 0.80 – 0.90); the values of q for each α value are taken from the tables provided by Harter and Balakrishnan (1998) with k=3 (3 subtest) and ∞ degree of freedom (df); USA (est) are the estimate thresholds for USA sample; ITALY (est) are the estimate thresholds for ITALY sample; ITALY (emp) are the empiric thresholds for ITALY sample.

Table 3 - Computational steps using Silverstein' formula (1989) on the data from the USA and Italian standardization sample of the WAIS-IV for PRI scores.

		R (3 subtests PRI)		
α	$q_{3,\infty}$	USA (est)	ITALY (est)	ITALY (emp)
0.01	4.120	6.8	7.8	10
0.05	3.314	5.5	6.3	7
0.10	2.902	4.8	5.5	6
0.20	2.424	4.0	4.6	5
0.30	2.095	3.5	4.0	4
0.40	1.826	3.0	3.5	4
0.50	1.588	2.6	3.0	3
0.60	1.363	2.2	2.6	3
0.70	1.138	1.9	2.2	2
0.80	0.900	1.5	1.7	2
0.90	0.618	1.0	1.2	1

Legenda. q is the critical value of the studentized range (for α = 0.01 – 0.05 – 0.10 - 0.20 – 0.30 – 0.40 – 0.50 – 0.60 – 0.70 – 0.80 – 0.90); the values of q for each α value are taken from the tables provided by Harter and Balakrishnan (1998) with k=3 (3 subtest) and ∞ degree of freedom (df); USA (est) are the estimate thresholds for USA sample; ITALY (est) are the estimate thresholds for ITALY sample; ITALY (emp) are the empiric thresholds for ITALY sample.

Table 4 - Computational steps using Silverstein' formula (1989) on the data from the USA and Italian standardization sample of the WAIS-IV for the 10 core subtests.

		R (10 core subtest)		
α	$Q_{10,\infty}$	USA (est)	ITALY (est)	ITALY (emp)
0.01	5.157	11.3	12.2	13
0.05	4.474	9.8	10.6	11
0.10	4.129	9.0	9.8	10
0.20	3.730	8.2	8.8	9
0.30	3.455	7.6	8.2	8
0.40	3.229	7.1	7.7	7
0.50	3.024	6.6	7.2	7
0.60	2.826	6.2	6.7	6
0.70	2.623	5.7	6.2	6
0.80	2.394	5.2	5.7	5
0.90	2.094	4.6	5.0	5

Legenda. q is the critical value of the studentized range (for α = 0.01 – 0.05 – 0.10 - 0.20 – 0.30 – 0.40 – 0.50 – 0.60 – 0.70 – 0.80 – 0.90); the values of q for each α value are taken from the tables provided by Harter and Balakrishnan (1998) with k=10 (10 core subtest) and ∞ degree of freedom (df); USA (est) are the estimate thresholds for USA sample; ITALY (est) are the estimate thresholds for ITALY sample; ITALY (emp) are the empiric thresholds for ITALY sample.

Table 5 - Estimates of means and standard deviations of the Max-Min differences between the four Indexes of the WAIS-IV, the three subtests of VCI, the three subtests of PRI, and the 10 core subtests for the USA and Italian standardization sample of the WISC-IV.

	USA (est)		ITA (est)		ITA (emp)	
	Mean	SD	Mean	SD	Mean	SD
Max-Min 4 Indexes	20.49	8.76	23.10	9.87	22.62	10.01
Max-Min 3 subtests VCI	2.78	1.46	3.22	1.69	3.08	1.78
Max-Min 3 subtests PRI	3.34	1.75	3.66	1.92	3.53	2.07
Max-Min 10 subtest	6.71	1.74	7.26	1.88	7.06	2.11

Table 6 – Thresholds of the Max-Min differences between the four Indexes of the WAIS-IV, the 10 core subtests, the 3 subtests of the VCI and the 3 subtests of PRI for the USA standardization sample and the Italian standardization sample of the WAIS-IV for different levels of probability (%).													
		4 Indexes			10 subtest			3 subtest VCI			3 subtest PRI		
%	z	USA (est)	ITA (est)	ITA (emp)	USA (est)	ITA (est)	ITA (emp)	USA (est)	ITA (est)	ITA (emp)	USA (est)	ITA (est)	ITA (emp)
5	1.64	35	40	40	10	11	11	6	6	6	7	7	7
6.7	1.50	34	38	38	10	11	11	5	6	6	6	7	7
8	1.41	33	37	37	10	10	11	5	6	6	6	7	7
10	1.28	32	36	36	9	10	10	5	6	6	6	7	7