DOI: 10.26387/bpa.2024.00012

Analysis of the rarity of differences in FSIQ in the Italian sample of the WISC-V

Luca Mandolesi[1], Claudio Vezzani[1], Chiara Busdraghi[1], Daniela Traficante[2], Margherita Lang[3], Lina Pezzuti[4]

[1] Giunti Psychometrics Holding S.r.l., Firenze, Italy
[2] Department of Psychology - Università Cattolica del Sacro Cuore di Milano, Italy
[3] Department of Psychology - Università degli Studi Milano Bicocca, Italy
[4] Department of Dynamic Psychology - Sapienza Università di Roma, Italy

luca.mandolesi@giuntipsy.com

ABSTRACT. L'obiettivo del presente lavoro è indagare, nell'adattamento italiano della *Wechsler Intelligence Scale for Children – V edizione (WISC-V)*, l'unitarietà del Quoziente Intellettivo Totale (QIT) e di tre scale composite (Indice di Abilità Generale - IAG, Indice di Competenza Cognitiva - ICC e Indice Non Verbale - INV), al fine di determinarne gli specifici valori soglia. A tale scopo sono stati analizzati i base rate delle discrepanze tra i punteggi dei subtest e sono stati eseguiti calcoli statistici per individuare le soglie delle differenze ampie e rare. I risultati confermano la validità dell'approccio statistico utilizzato e della sua integrazione con i base rate nel determinare le soglie per il QIT e per gli indici IAG, ICC e INV della WISC-V.

SUMMARY. The objective of this study is to investigate the unitarity of the Full Scale IQ (FSIQ) and three composites (General Ability Index - GAI, Cognitive Proficiency Index - CPI, Nonverbal Index - NVI) of the Italian adaptation of the WISC-V, aiming to determine their specific rarity thresholds. The importance of this aim is to determine if there is the possibility of using the FSIQ (or GAI, CPI, NVI) as a unique factor deviation quotient or not, by an accurate representation of the ability it is intended to assess. The distributions of the differences between maximum and minimum value (Max-Min discrepancies) were calculated using the Italian WISC-V standardization sample. The base rates of these discrepancies were analyzed, and statistical calculations of thresholds for large and rare differences were performed. The results confirm the validity of this statistical approach in determining the thresholds for the FSIQ and the indices of the WISC-V that corresponds to rare and unusual discrepancies. The obtained results combine the psychometric approach developed in previous versions of the Wechsler scales with the effective findings in the population as reflected by base rates (Flanagan & Kaufman, 2004, 2009; Orsini, Pezzuti & Hulbert, 2015). However, the FSIQ should not be classified as "uninterpretable" under any circumstances. Doing so would ignore its inherent predictive value, which remains intact regardless of score variability (Daniel, 2007).

Keywords: *WISC-V, Unitarity of intelligence, FSIQ, Rarity thresholds, Base rates*

INTRODUCTION

The FSIQ is the most reliable score on the WISC-V and is typically reported and interpreted as a summary of a child's intellectual abilities. However, significant variability among the scores that comprise the FSIQ can undermine its validity as a summary measure. (Flanagan & Alfonso, 2017)

The concept of the unitarity of composites and the FSIQ has been explored in previous editions of the Wechsler scales, where the issue arose of determining when to use the FSIQ or the corresponding factor deviation quotient. In addressing this question, Kaufman (1994) revisited the notion of the unitary construct of the FSIQ. According to Flanagan and Kaufman (2004, 2009), if a significant discrepancy is observed among the scaled scores that constitute a given index, that index does not provide an accurate representation of the ability it is intended to assess and, consequently, cannot be interpreted correctly, as it does not reflect a single ability. Conversely, an ability is considered unitary when it is formed by a cohesive set of scaled scores, each of which reflects unique or slightly different aspects of the ability itself.

Research conducted after the publication of *Essentials of WISC-IV Assessment (*Flanagan & Kaufman, 2004, 2009) showed the importance of the evaluation of the proportion of subjects (i.e., base rates) in the standardization sample that occur in a psychometrically defined threshold (Orsini, Pezzuti & Hulbert, 2014). Score differences that occur in <10% of the population are considered rare (Flanagan & Alfonso, 2017)

Some authors suggest that the interpretation of the FSIQ must consider the variability among the subtest scores that compose it. When this variability is minimal, the FSIQ can be interpreted as a cohesive measure of overall intellectual ability. However, when variability is large and rare, the FSIQ may not be a valid summary, and a more detailed analysis of the index scores is necessary to provide a comprehensive understanding of the individual's cognitive abilities (Flanagan & Alfonso, 2017). However, it's crucial to understand that the FSIQ maintains its predictive validity even when there is significant variability among the subtest scores that compose it. Daniel (2007) emphasized that the FSIQ's construct, and predictive validity are independent of the variability in the component scores. This means that despite large discrepancies among subtest scores, the FSIQ can still provide valuable predictive information. Given this, the FSIQ should not be classified as "uninterpretable" under any circumstances. Doing so would ignore its inherent predictive value, which remains intact regardless of score variability.

METHOD

Participants

The sample used for the following study is the Italian standardization sample of the WISC-V test, composed of 1,410 subjects aged between 6 years and 0 months and 16 years and 11 months, balanced for gender (M = 50.2%, F = 49.8%) and representative of the Italian population (see Wechsler, 2023).

Data analysis

Unlike in the WISC-IV, where the FSIQ was composed of the sum of the scores of four composites (Verbal Comprehension Index - VCI, Perceptual Reasoning Index - PRI, Working Memory Index - WMI, Processing Speed Index - PSI), in the fifth version of the battery, the calculation of the FSIQ score is derived from the sum of the scaled scores of seven primary subtests. Therefore, the calculation of the unitarity of the FSIQ follows the typical approach for composites, namely the Max-Min difference of the scaled scores of the subtests that compose it (see Wechsler, 2023).

To study the unitarity of the FSIQ and the composites, the distributions of the Max-Min values of the scaled scores that compose the FSIQ and the composites of the WISC-V were analyzed. To do so for each index, the difference between the maximum and minimum values of the scaled scores for the subtests that compose them was calculated. These differences are thus always positive.

This analysis was conducted for the composites composed of more than two subtests, as for the composites composed of two subtests, pairwise comparisons between the subtests have already been analyzed in the Italian validation study.

Subsequently, an analysis of variance (ANOVA) was conducted to evaluate whether the Max-Min differences between the subtests that compose the FSIQ are independent of the age and education level of the mother or the level of the FSIQ.

For each of these differences, absolute frequency, percentage frequency, and the base rate (%Ss) have been calculated, which represents the percentage of subjects who obtained a value of difference equal to or greater than a specified threshold. The trend of the Max-Min distribution allows clinicians to assess how much and in what manner any threshold affects the reference sample of the test.

The statistical calculation of the rarity threshold values for the FSIQ and the composites of the WISC-V was performed using the method proposed by Flanagan and Kaufman (2004, 2009) and the subsequent modifications by Orsini, Pezzuti and Hulbert (2014). This method uses the formula:

Difference Threshold = $M_{\text{Max-Min}} + z^{*}SD_{\text{Max-Min}}$
where $M_{\text{Max-Min}}$ represents the mean of the distribution of the Max-Min differences (range) of the scaled scores of the subtests that compose the FSIQ or the index, $SD_{\text{Max-Min}}$ is the standard deviation of this distribution, and z is the normal distribution value (one-tailed) associated with the chosen percentage of subjects.

The lower this percentage, the greater the differences needed to be defined as rare. Conversely, the higher the chosen percentage, the smaller the differences needed to be defined as rare. While there is no universally accepted percentage of subjects considered rare, following Flanagan and Kaufman (2004, 2009), a value of 6.7% of the population (corresponding to $z = 1.5$) can be considered an adequate rarity criterion. For the sake of completeness, this text will also illustrate the threshold values for both lower and higher percentages.

RESULTS

The analysis of variance conducted confirmed that, similarly to what found for the WISC-IV, in the WISC-V as well, the distribution of Max-Min differences for the subtests composing the FSIQ is independent of age, maternal education level, and FSIQ level (<80, 80-89, 90-109, 110-119, >119). All main effects and interactions were found to be statistically non-significant (p>.05) (see Table 1).

Table 1 – ANOVA Max-Min FSIQ difference as dependent variable by age, mother's education level, FSIQ level and their interactions as factors (independent variables)

Factor	F	*df*	*p*
Age	1.59	10;1217	.104
Mother's education level	1.38	4;1217	.238
FSIQ level	1.42	4;1217	.226
Age * Mother's education level	1.02	32;1217	.443
Age * FSIQ level	.94	40;1217	.574
Mother's education level * FSIQ level	.33	14;1217	.990

Legenda. df = degree of freedom; FSIQ = *Full Scale IQ*.

Therefore, were calculated the percentage frequencies, and base rates of subjects (%Ss) for the different composites (see Table 2).

From these distributions of Max-Min differences, it is therefore possible to calculate the parameters of mean and standard deviation, which allow deriving the difference thresholds for different percentages of the population (5%, 6.7%, 8%, 10%, 13%) (see Table 3).

A Max-Min difference among the weighted scores that compose a specific composite greater than the threshold value at a certain percentage of subjects will indicate the presence of a rare difference in that composite. As expected, the difference threshold value decreases as the selected percentage of subjects increases. This implies that a higher percentage will identify smaller weighted score differences as rare and unusual. However, since these differences are always

Table 2 – Descriptives of the discrepancies of FSIQ, GAI, CPI and NVI

FSIQ			GAI			CPI			NVI		
Disc	*%Freq*	*%Ss*	*Disc*	*%Freq*	*%Ss*	*Disc*	*%Freq*	*%Ss*	*Disc*	*%Freq*	*%Ss*
15	.07	.07	13	.35	.35	13	.14	.14	15	.07	.07
14	.28	.36	12	.50	.85	12	.36	.50	14	.14	.21
13	.50	.85	11	1.35	2.20	11	1.85	2.35	13	.36	.57
12	1.64	2.49	10	2.20	4.40	10	1.85	4.20	12	.85	1.42
11	3.06	5.55	9	4.04	8.44	9	4.70	8.90	11	2.06	3.49
10	4.70	10.25	8	7.59	16.03	8	7.19	16.09	10	4.27	7.76
9	9.18	19.43	7	11.49	27.52	7	9.61	25.69	9	7.76	15.52
8	13.24	32.67	6	16.38	43.90	6	13.31	39.00	8	9.54	25.05
7	16.94	49.61	5	17.73	61.63	5	14.23	53.24	7	15.80	40.85
6	18.58	68.19	4	16.67	78.30	4	17.30	70.53	6	17.58	58.43
5	14.38	82.56	3	14.04	92.34	3	15.52	86.05	5	16.44	74.88
4	11.53	94.09	2	6.52	98.87	2	10.04	96.09	4	13.59	88.47
3	4.56	98.65	1	1.06	99.93	1	3.49	99.57	3	8.90	97.37
2	1.35	100.00	0	.07	100.00	0	.43	100.00	2	2.42	99.79
									1	.21	100.00

Legenda. FSIQ = *Full Scale IQ*; GAI = General Ability Index; CPI = Cognitive Proficiency Index; NVI = Nonverbal Index; Disc = discrepancy; %Freq = percentage frequencies; %Ss = base rates.

Table 3 – Mean, standard deviation and discrepancy thresholds

Index	Parameter		Threshold				
	Mean	*SD*	*5%*	*6.7%*	*8%*	*10%*	*13%*
FSIQ	6.65	2.22	10.3	10.0	9.8	9.5	9.2
GAI	5.35	2.20	9.0	8.7	8.5	8.2	7.8
CPI	5.02	2.38	8.9	8.6	8.4	8.1	7.7
NVI	6.14	2.24	9.8	9.5	9.3	9.0	8.7

Legenda. FSIQ = *Full Scale IQ*; GAI = General Ability Index; CPI = Cognitive Proficiency Index; NVI = Nonverbal Index.

integer values, this change does not always translate into appreciable differences in the test's practical use.

To confirm the results of this study, comparing the threshold values relative to the 6.7% of the population with the %Ss found in the standardization sample shows that this criterion for rarity of differences effectively isolates a percentage of subjects always less than 10% across all composites. The data derived from parametric statistics is thus consistent with the analysis of empirical distributions, supporting its use in practical test applications.

CONCLUSION

This study has allowed us to calculate, similar to previous versions of the Wechsler scales, the rarity criterion for differences within the FSIQ and composites of the WISC-V. The analyses conducted have demonstrated that the distribution of these differences does not vary within the test's standardization sample. Therefore, a single threshold value can be considered for all different age ranges within the sample. Through appropriate statistical procedures, we have calculated this threshold value for various percentages of the population.

If the difference between the weighted scores of the subtests that compose a composite/FSIQ exceeds the identified threshold value, it indicates that this difference is identifiable as "rare", suggesting that the index in question does not represent a cohesive summary of the child's functioning. For example, considering the FSIQ, Table 3 identifies a threshold value for rarity of differences in 6.7% of cases as 10.0. If the difference between the highest and lowest weighted scores among the 7 subtests that compose the FSIQ is greater than 10.0, the ability cannot be considered a unitary measure. Conversely, if the difference is equal to or less than 10.0, it indicates that the FSIQ score provides a cohesive summary of the child's functioning.

In evaluating the rarity of differences, different percentages of subjects can be considered simply by applying the corresponding difference threshold for the desired population percentage.

The obtained results combine the psychometric approach developed in previous versions of the Wechsler scales with the effective findings in the population as reflected by base rates. It is always prudent to integrate both the psychometric approach and the empirical findings in the population, as reflected by base rates. This approach ensures a comprehensive understanding of the actual rarity of observed differences within the population (Flanagan & Kaufman, 2004, 2009; Orsini et al., 2014).

It's very important consider that, however, coherently with Daniel (2007) and FSIQ maintains its predictive validity even when there is significant variability among the subtest scores that compose it. This means that despite large discrepancies among subtest scores, the FSIQ can still provide valuable predictive information.

References

DANIEL, M.H. (2007). “Scatter” and the construct validity of FSIQ: Comment on Fiorello et al. *Applied Neuropsychology*, *14* (4), 291-295.

FLANAGAN, D.P., & ALFONSO, V.C. (2017). *Essentials of WISC-V assessment*. John Wiley & Sons.

FLANAGAN, D.P., & KAUFMAN, A.S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: John Wiley & Sons.

FLANAGAN, D.P., & KAUFMAN, A.S. (2009). *Essentials of WISC-IV assessment (2nd ed.)*. Hoboken, NJ: John Wiley & Sons.

KAUFMAN, A.S. (1994). *Intelligent testing with the WISC-III*. New York, NY: Wiley.

ORSINI, A., PEZZUTI, L., & HULBERT, S. (2015). Beyond the floor effect on the Wechsler Intelligence Scale for Children – 4th ed. (WISC-IV): Calculating IQ and indexes of subjects presenting a floored pattern of results. *Journal of Intellectual Disability Research*, *59* (5), 468-473.

WECHSLER, D. (2023). *Wechsler Intelligence Scale for Children® – Fifth Edition: Manuale*. Adattamento italiano di Lina Pezzuti, Daniela Traficante e Margherita Lang. Firenze: Giunti Psychometrics.