G Model
ERAP-100634; No. of Pages 8
ARTICLE IN PRESS
European Review of Applied Psychology xxx (xxxx) xxx
Available online at ScienceDirect www.sciencedirect.com
Elsevier Masson France EM|consulte www.em-consulte.com/en

Original article

Comparison of four short forms of the French adaptation of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV)

Comparaison de quatre formes abrégées de la version française de l'échelle d'intelligence de Wechsler pour adultes – Quatrième édition (WAIS-IV)

Jacques Grégoire[a,*], Anna Schmitt[b]

[a] *Faculté de psychologie et des Sciences de l'Éducation, université Catholique de Louvain, 10, place du Cardinal-Mercier, 1348 Louvain-la-Neuve, Belgium*
[b] *Faculté des Sciences Humaines, des Sciences de l'Éducation et des Sciences Sociales, université du Luxembourg, 2, avenue de l'université, 4365 Esch-sur-Alzette, Luxembourg*

A R T I C L E I N F O

Article history:
Received 15 December 2020
Received in revised form 8 March 2021
Accepted 9 March 2021

Keywords:
WAIS-IV
Intelligence
Short form
Assessment

A B S T R A C T

Introduction. – The Wechsler Intelligence Scale for Adults - 4th Edition (WAIS-IV) is widely used in France and French-speaking countries for clinical assessment and research. This scale has excellent metric qualities; there is ample evidence of the validity and reliability of its scores. However, it takes 60 to 90 min to complete the full test. That can be problematic in research where time is limited and when examining very disturbed patients. In such cases, a short form of the WAIS-IV may be useful. Unfortunately, no short form of the French version of WAIS-IV has yet been validated using the standardization sample.
Objective. – The aim of the present study was to identify an abbreviated form of WAIS-IV with sufficient validity and reliability to replace, in some cases, the full test.
Method. – Four short forms were developed taking into account the fidelity and the correlation with the FIQ of the subtests included in each of form. The metric qualities the four short forms were compared using the standardization data of the French WAIS-IV. The standardization was done using the results of a sample of 876 individuals representative of the French population from 16 to 79 years 11 months.
Results. – The analyzes showed that a short form including Information, Matrices, Arithmetic and Code provides a fairly good estimate of the FIQ. The fidelity of the estimated IQ was 0.94 and its correlation with the FIQ was 0.91. However, the average of the absolute differences between the IQ calculated with this short form and the IQ calculated with the full form was 4.24 points, with a standard deviation of 3.15 points. These differences could lead to misidentification of some individuals tested with the short form.
Conclusion. – The proposed short form provides a sufficient approximation of the FIQ to be used in research where the collective results are more important than those of individuals. This short form, however, does not provide an IQ measure as robust as the full form. It should therefore be used with caution only when using the full form is not possible.

Mots clés :
WAIS-IV
Intelligence
Forme abrégée
Évaluation

R É S U M É

Introduction. – L'échelle d'intelligence de Wechsler pour adultes – 4e édition (WAIS-IV) est largement utilisée en France et dans les pays de langue française pour les examens cliniques et pour la recherche. Cette échelle possède de solides qualités métriques; les preuves de la validité et de la fidélité de ses scores sont nombreuses. Toutefois, elle nécessite un temps de passation 60 à 90 min. Ce qui peut poser problème dans le cadre de recherches où le temps disponible est limité et lors d'examens de patients très fatigables. Dans de tels cas, une forme abrégée de la WAIS-IV peut se révéler utile. Malheureusement, aucune forme abrégée de la version française de la WAIS-IV n'a encore été validée sur la base des données de l'échantillon d'étalonnage.

* Corresponding author.
E-mail address: jacques.gregoire@uclouvain.be (J. Grégoire).

https://doi.org/10.1016/j.erap.2021.100634
1162-9088/

Please cite this article as: J. Grégoire and A Schmitt, Comparison of four short forms of the French adaptation of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV), *Eur. Rev. Appl. Psychol.*, https://doi.org/10.1016/j.erap.2021.100634

Objectif. – La présente étude vise à identifier une forme abrégée de la WAIS-IV dont la validité et la fidélité soient suffisantes pour remplacer, dans certains cas, la forme complète du test.
Méthode. – Quatre formes abrégées ont été construites en prenant en compte la fidélité et la corrélation avec le QIT des subtests inclus dans chacune de ces formes. Leurs qualités métriques ont été comparées en utilisant les données d'étalonnage de la version française de la WAIS-IV. Cet étalonnage a été réalisé sur la base des résultats d'un échantillon de 876 personnes représentatives de la population française âgée de 16 à 79 ans et 11 mois.
Résultats. – Les analyses montrent qu'une forme abrégée comprenant: *Information, Matrices, Arithmétique et Code* permet une assez bonne estimation du QIT. La fidélité du QI estimé est de 0,94 et sa corrélation avec le QIT est de 0,91. Toutefois, la moyenne des différences absolues entre le QI calculé avec la forme abrégée et celui calculé avec la forme complète est de 4,24 points, avec un écart-type de 3,15 points. Ces différences peuvent conduire à des erreurs d'identification de certains individus testés avec la forme abrégée.
Conclusion. – La forme abrégée qui a été identifiée permet une approximation suffisante du QIT pour être utilisée dans le cadre de recherches où les résultats collectifs sont plus importants que ceux des individus. Cette forme abrégée ne permet toutefois pas une mesure aussi robuste du QI que la forme complète. Elle ne devrait dès lors être utilisée qu'avec précaution lorsque la passation de la forme complète n'est pas possible.

In French, the fourth edition of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2009) is undoubtedly the intelligence test most often used in clinical practice. This test measures a Full Scale IQ (FSIQ) and four index scales: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI) and Processing Speed (PSI) derived from ten subtest scores. These four index scales were developed to match the main group factors of the second stratum of the hierarchical model of intelligence named CHC, which is the acronym of the founding fathers of this model: Cattell, Horn and Cattell (McGrew, 2009). VCI is a measure of *Comprehension-knowledge* (Gc), PRI is a composite measure of *Fluid reasoning* (Gf) and *Visual processing* (Gv), WMI is a measure of *Short-term memory* (Gsm) and PSI is a measure of *Processing speed* (Gs).

The FSIQ is calculated on the basis of all ten subtests and thus reflects the main components of the CHC model. The indices are calculated on the basis of the following subsets of subtests:

- VCI = Similarities + Vocabulary + Information;
- PRI = Block Design + Matrix Reasoning + Visual Puzzles;
- WMI = Digit Span + Arithmetic;
- PSI = Symbol Search + Coding.

In addition to the ten core subtests used to calculate the FSIQ and the indices, the WAIS-IV includes five additional subtests that can be used to supplement the information provided by the core subtests or, on occasion, to replace one of them. Because the additional subtests are not normally used to calculate the FSIQ, they have not been taken into account in this research.

The French WAIS-IV is the adaptation of the American test of the same name (Wechsler, 2008). French norms have been established on the basis of a sample of 876 subjects representative of the entire population, from 16 years to 79 years and 11 months of age. Several evidences of validity and reliability ensure the relevance and accuracy of the FSIQ and those of the indices calculated using this scale.

Administration of the ten core subtests of the WAIS-IV takes 67 min on average (Wechsler, 2008). This time can vary from one hour to one and a half hours depending on the intellectual level of the subjects, their age and any pathologies likely to interfere with taking the test. The higher the intellectual level of the subjects, the greater the number of items presented, which prolongs the testing process accordingly. Despite the length of this administration process, it is recommended that all the subtests be taken in most clinical situations, as it provides a robust measurement of the FSIQ and the four indices, and allows a broad array of information useful in diagnosis to be collected.

In some cases, however, complete administration of the WAIS-IV can pose a problem. Some patients may be incapable of taking all ten of the core subtests due to their pathology. The most serious problem emerges in clinical or experimental research, where IQ measurement is only one element in a long evaluation protocol. Administration of the entire WAIS-IV can lead to tiredness, lassitude and a risk of discontinuation by the participants. Use of an abridged version of the test can then prove to be appropriate. Kaufman and Kaufman (2001) have, however, questioned the relevance of the short forms of the intelligence tests, because their reliability and validity are lower than those of the full scales. These authors stress that the norms of the tests have been established on the basis of the complete version, following a sequence of administering the subtests that is no longer followed in the short forms. These criticisms are valid, but nevertheless should not lead to rejection of all the short forms. If they are constructed with care and provide sufficient evidence of validity and reliability, and if the practitioners comply with the conditions for use, they can be highly useful.

Several researchers have created short forms of the WAIS-IV, ensuring that their validity and reliability are as high as possible. These short forms have been developed on the basis of the American version of the WAIS-IV or its adaptation into other languages. To our knowledge, a single short form has been created on the basis of the French version of the WAIS-IV (Bulzacka et al., 2016). Unfortunately, the utility of this short form remains limited due to its length (seven subtests), its complex calculation procedure and insufficient evidences of validity, based on a sample of 70 patients who are schizophrenic or have a schizoaffective disorder. In this research, we wanted to make available to practitioners and researchers a short form of the French WAIS-IV that is shorter, uses a simple calculation formula and presents strong evidences of validity and reliability based on the data of the French standardization sample ($n = 876$).

We first examined the various short forms of the WAIS-IV developed in other languages. They are composed of two to seven subtests, with a preponderance of combinations of four subtests. The shortest form, proposed by Girard, Axelrod, Patel and Crawford (2015), includes only two subtests, *Coding* and *Information*, and requires less than 15 min of testing time. It was constructed on the basis of a sample of 482 patients examined in neuropsychological consultation. Its correlation with the FSIQ is 0.86. On the basis of their score on this short form, 83% of the patients lie in an interval of plus or minus 10 points around the FSIQ obtained with the complete form.

The longest form, proposed by Meyers, Zellinger, Kockler, Wagner, & Miller (2013), includes seven subtests: *Block Design, Similarities, Digit Span, Arithmetic, Information, Coding* and *Picture Completion*. The composition of this form retains that proposed by Ward (1990) for the WAIS-R. The same composition of subtests was used for a short form of the WAIS-III by Pilgrim, Meyers, Bayless, & Whetstone (1999), and by Axelrod, Rayan, and Ward (2001) in the context of neuropsychological evaluations. Meyers et al. (2013) presented evidence of the validity of this short form of the WAIS-IV from the results for 70 patients divided into two groups depending on their pathology. The authors report a correlation of 0.99 between the score on this abridged version and the FSIQ. Despite this high correlation, this short form raises several problems. The first is its length. Reducing test administration to seven subtests does not save a great deal of time, but has the disadvantage of precluding calculation of the indices. The second problem is the use of a subtest, *Picture Completion* that is not part of the ten core subtests. Finally, the third problem is the omission of the subtest *Matrix Reasoning*, which measures an important component of intelligence, i.e. fluid reasoning (Gf).

The other short forms of the WAIS-IV proposed by researchers include four subtests. This number of subtests seems to allow the best balance between a substantial reduction in testing time and a sufficiently accurate estimate of the FSIQ.

With the Dutch version of the WAIS-IV, van Ool et al. (2018) have evaluated the metric properties of five short forms including four subtests on a group of 324 adults with various neurological disorders. The form including the subtests *Visual Puzzles, Symbol Search, Vocabulary* and *Arithmetic* appears to be the most satisfactory. Its correlation with the FSIQ is 0.73 and 0.82, depending on whether the FSIQ is below 80 or greater than or equal to this value. Merz, Lace, Eisenstein and Grant (2019) used data from 318 patients seen in outpatient consultation for psychological problems to identify the best short forms among thirty-six possible combinations of four subtests. Using regression analyses, they highlighted the three combinations that allow the best prediction of the FSIQ. Among these, the short form including *Vocabulary, Block Design, Arithmetic* and *Coding* stands out slightly. With this short form, 97.9% of the patients obtain an estimated IQ lying within an interval of plus or minus ten points around the FSIQ obtained with the complete form. Fan et al. (2019) used the standardization sample of the Chinese WAIS-IV ($n = 1.757$) and a sample of 239 patients presenting various psychological disorders to identify the best short form of four subtests. They identified the combination of *Block Design, Information, Arithmetic* and *Coding* as the best, as it is the best correlated with the FSIQ ($r = .95$). Moreover, the Bland-Altman plot showed that the differences between the scores on this short form and the FSIQ are in a rather narrow interval. Chen and Hua (2020) used the Taiwanese standardization sample of the WAIS-IV ($n = 1.105$) to determine the best combination of four subtests for a short form. They took account of the test administration time, the correlation between the score on the short form and the FSIQ, the magnitude of the deviations from the FSIQ and the percentage of correct rankings in five categories of FSIQ. On the basis of these criteria, several combinations turn out to be satisfactory. The form including *Information, Visual Puzzles, Digit Span* and *Symbol Search* allows the best balance between the time saved and the accuracy of the estimate of the FSIQ.

Studies conducted on the short forms of the WAIS-IV show that the form including only two subtests is less well correlated with the FSIQ and provides less accurate measurements than the forms including a larger number of subtests. However, taking into account too large a number of subtests does not allow substantial time savings, which is the first objective of a short form. Most of the studies show that a short form of four subtests allows the best balance between time savings and the accuracy of the measurements. The four studies presented that propose a short form of four subtests use four different versions of the WAIS-IV: the original American version (Merz et al., 2019), the Dutch adaptation (van Ool et al., 2018), the Chinese adaptation (Fan et al., 2019) and the Taiwanese adaptation (Chen & Hua, 2020). All these studies recommend different combinations of subtests. The *Arithmetic* subtest is the only one found in three of the four short forms. The other subtests appear only in one or two of these short forms. Clearly, there is no short form that would be a universal standard and could be adopted throughout the world for all the versions of the WAIS-IV. The metric properties of the subtests vary significantly depending on the languages and cultures into which the WAIS-IV has been adapted. A single subtest does not necessarily provide the same estimate of the FSIQ regardless of the adaptation of the WAIS-IV used. It is therefore necessary to identify the most valid and reliable short form for each version of the WAIS-IV. To do so, it is preferable to rely on the standardization data for this version, as they are based on a large sample representative of the reference population in terms of age, gender and other sociological characteristics.

1. Method

1.1. Participants

All the analyses were conducted on the basis of the standardization data of the French version of the WAIS-IV[1]. This standardization was conducted starting from the results of a stratified random sample of 876 individuals representative of the French population. This sample was constituted on the basis of the 2007 general census of the population conducted by the National Institute of Statistics and Economic Studies (*Institut national de statistiques et d'études économiques*, INSEE). Age, gender, socio-professional category and population density of the place of residence were taken into account in constituting the various strata. The participants ranged from 16 years to 79 years and 11 months of age and were divided into eleven age groups: 16–17 years, 18–19 years, 20–24 years, 25–29 years, 30–34 years, 35–44 years, 45–54 years, 55–64 years, 65–69 years, 70–74 years and 75–79 years of age.

1.2. Construction of the short forms

A short form should provide an estimated IQ as close as possible to the FSIQ calculated with the WAIS-IV full scale. To do this, we used the criteria selected in three previous studies of short forms of the French versions of the Wechsler scales with proven reliability and validity, namely those of the WISC-III, WISC-IV and WAIS-III (Grégoire, 2000, 2006; Grégoire & Wierzbicki, 2009). Four criteria were taken into account:

- the subtests should be part of the ten core subtests used for calculation of the FSIQ;
- the reliability of the subtests and their correlation with the FSIQ should be as high as possible;
- the various indices should be represented if possible; in any case VCI and PRI, measures of Gc and Gf, should be taken into account since they are considered the most important group factors in the CHC model (Grégoire, 2013);
- the short form should include a maximum of four subtests in order to ensure a substantial savings in test administration time.

Table 1 presents the two metric properties, the associated indices and the administration duration of the ten core subtests

[1] The authors thank *ECPA by Pearson* for allowing them to use the standardization data for the WAIS-IV in the context of this research.

Table 1
Metric properties, index and administration duration of the ten core subtests of the WAIS-IV.

Subtest	Correlation with FSIQ	Reliability	Index	Duration[1]
Similarities	0.74	0.83	VCI	9
Vocabulary	0.69	0.90	VCI	13
Information	0.73	0.90	VCI	6
Block Design	0.72	0.87	PRI	11
Matrix Reasoning	0.74	0.90	PRI	8
Visual Puzzles	0.72	0.91	PRI	9
Digit Span *Mémoire de chiffres*	0.68	0.91	WMI	9
Arithmetic	0.71	0.88	WMI	8
Coding	0.60	0.83	PSI	3
Symbol Search	0.56	0.78	PSI	3

[1] Mean duration, in minutes, for the Taiwanese adaptation of the WAIS-IV (Chen & Hua, 2020).

taken into account in selecting the subtests to be included in a short form of the WAIS-IV. The correlation of the subtest with the FSIQ is a very important number, as it tells us the degree of covariation of the scores on the subtest with those of the FSIQ. A subtest showing a high correlation with the FSIQ is a good candidate for inclusion in a short form. The reliability coefficient cited in the third column tells us the degree of accuracy of the scores provided by each subtest. The higher the coefficient, the lower the measurement error of these scores. It is necessary to calculate the short form IQ on the basis of scores with a low error of measurement so that the reliability of this IQ is as high as possible. The fourth column lists the associated indices of the subtests. The subtests included in the short form should take into account the various indices so that the estimated IQ is as representative as possible of the composition of the FSIQ. Finally, the fifth column provides information on the mean administration duration of each subtest. Unfortunately, these durations do not exist for the subtests of the French version of the WAIS-IV. They are only available for the Taiwanese version of WAIS-IV, and were published in the article by Chen and Hua (2020). Although it is possible that the administration durations differ slightly from those of the French version, the Taiwanese data give us a useful estimate of the relative administration duration of the various subtests.

We have constructed four short forms, taking account of the criteria cited above and the values presented in Table 1: two forms including four subtests, one including three subtests and one including two subtests.

1.2.1. Short forms with 4 subtests

Form 1 (IQ_{sf1}) includes the subtests *Information, Matrix Reasoning, Arithmetic* and *Coding,* which come from the VCI, PRI, WMI and PSI respectively. Their correlation with the FSIQ and/or their reliability coefficient are the highest among the subtests of the corresponding index. Although the correlation of *Information* with the FSIQ is slightly lower than that of *Similarities* (0.73 versus 0.74), we have included *Information* in this short form, as its reliability is distinctly superior and its administration duration is shorter. The mean estimated administration duration for this short form is 25 min.

Form 2 (IQ_{sf2}) includes the subtests *Similarities, Vocabulary, Matrix Reasoning* and *Block Design,* which come from the VCI and PRI respectively. This form emphasizes the two indices best correlated with the FSIQ (Wechsler, 2009), which correspond to the two facets of intelligence (Gc and Gf) considered most important (Grégoire, 2013). This short form had been shown to be the best for estimating the FSIQ of the previous version of the Wechsler scale, the WAIS-III (Grégoire & Wierzbicki, 2009). The mean estimated administration time for this short form is 41 min.

1.2.2. Short form with 3 subtests

Form 3 (IQ_{sf3}) includes the subtests *Information, Matrix Reasoning* and *Arithmetic,* which come from the VCI, PRI and WMI respectively. This shorter form excludes the subtests of the PSI that are least well correlated with the FSIQ. The mean estimated administration time for this short form is 22 min.

1.2.3. Short form with 2 subtests

Form 4 (IQ_{sf4}) includes the subtests *Information* and *Matrix Reasoning,* which come from VCI and PRI, respectively. The first is a standard measure of Gc, and the second of Gf. The mean estimated administration time for this short form is 14 min.

The sum of the standard scores for the subtests of each short form has been calculated for each subject in the standardization sample. Each of the four sums has then been converted into a standard IQ of mean 100 and standard deviation 15 to obtain a short-form intellectual quotient (IQ_{sf}). The following formula has been used for this conversion:

$$IQ_{sf} = 15\left(\frac{\sum x_i - \bar{x}_l}{s_i}\right) + 100$$

$\sum x_i$ = sum of the standard scores for the subtests of the short form,

$\bar{x}_l$ = mean of the sums of the standard scores for the whole standardization sample,

s_i = standard deviation of the sums of the standard scores for the whole standardization sample.

To facilitate conversion of the sum of the standard scores into IQ_{sf} by users, the conversion formulas have been simplified for each of the four short forms. The simplified conversion formulas are presented in Table 2.

1.3. Analyses

What is the value of the four short forms of the WAIS-IV? Do they allow a sufficiently accurate estimate of the FSIQ obtained with the full version of the test? Does one of the simplified formulas stand out from the others in this regard? Several statistical analyses have been performed to answer these questions. The first is the calculation of the correlations between each of the IQ_{sf} and the FSIQ. These correlations were calculated using Levy's formula (1967), modified by Girard and Christensen (2008), which corrects the inflated correlation between a test and a part of the same test, and takes into account the high correlations between the subtests of the short form. The formula used to calculate the corrected correlations $\left(r'_{sf}\right)$ is the following:

$$r'_{sf} = r_{sf} - \left((1 - x_{xx'}) \times \frac{p}{w} \times \frac{SD_{sf}}{SD_{fs}}\right)$$

r_{fa} = uncorrected correlation between the IQ_{sf} and the FSIQ,
$x_{xx'}$ = FSIQ reliability coefficient,
p = number of subtests in the short form,
w = number of subtests in the full scale,

Table 2
Conversion formula of the sum of the standard scores of each short form to get the corresponding IQ_{sf}.

IQ_{sf}	Sum of the subtest scores	Conversion formula
Form 1	$\sum x_i$ = Information + Matrix Reasoning + Arithmetic + Coding	$IQ_{sf1} = \left(1.70 \sum x_i\right) + 31.73$
Form 2	$\sum x_i$ = Similarities + Vocabulary + Matrix Reasoning + Block Design	$IQ_{sf2} = \left(1.63 \sum x_i\right) + 34.98$
Form 3	$\sum x_i$ = Information + Matrix Reasoning + Arithmetic	$IQ_{sf3} = \left(2.08 \sum x_i\right) + 37.50$
Form 4	$\sum x_i$ = Information + Matrix Reasoning	$IQ_{sf4} = \left(2.96 \sum x_i\right) + 40.68$

SD_{sf} = standard deviation of the short form,
SD_{fs} = standard deviation of the full scale.

The difference between the mean of each IQ_{sf} and the mean of the FSIQs was then calculated. Student's t test for a paired sample was used to assess the statistical significance of the differences between means. The effect size was appraised using Cohen 's d for comparison of paired data coming from one sample (Cohen, 1988; Rosenthal, 1991).

It must be borne in mind that the differences can be positive ($FSIQ > IQ_{sf}$) or negative ($FSIQ < IQ_{sf}$). Because they are distributed approximately normally around the mean, they tend to cancel out. Therefore, the means of the differences and the t tests can give us a misrepresentation of the real discrepancies between the FSIQ and the various IQ_{sf}. To have a correct representation of these, the means and the standard deviations of the absolute differences between the FSIQ and each of the IQ_{sf} have been calculated. The percentage of subjects in the sample with a specific absolute difference (from 0 to 10 and more) has also been calculated.

The extent of agreement between the FSIQ and the IQ_{sf} has been evaluated for the entire standardization sample with the Bland-Altman plot (1995, 1999). Since Bland and Altman's seminal article (1986), use of this plot has become a standard for determining whether the results obtained with two methods are in sufficient agreement that these methods can be used interchangeably. This plot allows the extent of agreement between two measurements to be visualized using their mean and their difference. The extent of agreement is determined on the basis of the interval including 95% of the differences, the bounds of which lie on either side of the mean difference. These bounds correspond to $\pm$ 1.96 standard deviations of the distribution of the differences around the mean. The narrower the interval, the greater the agreement between the two measurements. In addition, the point cloud thus represented allows any special features within the distribution of the differences between the two measurements to be identified.

In clinical practice, some values of the FSIQ play an important role in identification of certain intellectual characteristics. This is the case for $FSIQ \leq 70$, which is one of the diagnostic criteria for mental retardation, and for $FSIQ \geq 130$, which is used as a criterion for identification of high intellectual potential. It is therefore essential that a short form of the WAIS-IV should allow subjects with mental retardation or a high intellectual potential to be properly classified. ROC curves have been produced to compare the efficacy of the four IQ_{sf} in identifying individuals with mental retardation or high intellectual potential previously identified on the basis of their FSIQ. For each IQ_{sf}, the ROC curve represents the relationship between sensitivity, that is, the percentage of subjects properly identified as having the target characteristic (true positive rate) and "1 - specificity", that is, the percentage of subjects erroneously identified as having this same characteristic (false positive rate). The curve thus plotted allows evaluation of the degree to which use of the IQ_{sf} leads to a classification of the subjects better than that of chance. It also allows the efficacy of the four IQ_{sf} to be compared based on the area under the curve. The greater this area is, the more the IQ_{sf} allows an accurate classification (a maximum of true positives and a minimum of false positives).

A final piece of information important in evaluating the four short forms is their reliability coefficient, which tells us the degree of accuracy of the measurements they provide. The developers of the WAIS-IV calculated two measures of reliability of the FSIQ, the test/retest correlation and the alpha coefficient. Given the development methods of the short forms, only an estimate of the alpha coefficient can be calculated for each of the IQ_{sf}. The four reliability coefficients $\left(r_{yy'}\right)$ have been calculated using the formula of Tellegen and Briggs (1967), which uses the reliability coefficients of the subtests included in the short form and the correlations between these subtests.

$$r_{yy'} = \frac{\sum r_{xx'_i} + 2\sum r_i}{k + 2\sum r_i}$$

$r_{xx'_i}$ = reliability coefficients of the subtests in the short form,
r_i = correlations between the subtests of the short form,
k = number of subtests in the short form.

2. Results

Table 3 shows the mean and the standard deviation of the distribution of IQs estimated according to the four short forms. We note that the means and standard deviations are very close to the theoretical mean and standard deviation of 100 and 15, respectively. The correlations are all significant and large. The highest correlation coefficient is .91 for IQ_{sf1} and the lowest is .83 for IQ_{sf4}. The mean difference between the IQ_{sf} and the FSIQ is almost zero for the four short forms. These results are not surprising, since the individual differences are distributed almost normally around a mean close to zero; the positive and negative differences therefore tend to cancel out. A Student t test for paired data was performed to compare the means of the IQs obtained with the full form of the WAIS-IV and

Table 3
Mean and standard deviation of the four IQ_{sf}, correlation and mean difference of the IQ_{sf} with the FSIQ, Student's t test and Cohen's d comparing the means of the IQ_{sf} and FSIQ.

Short form	Mean	SD	Correlation with FSIQ[a]	FSIQ- IQ_{sf}	t	dl	p	d[b]
IQ_{sf1}	99.97	14.99	0.91	0.03	0.185	875	0.853	0.013
IQ_{sf2}	100.05	14.99	0.89	−0.04	−.0196	875	0.844	−0.013
IQ_{sf3}	100.03	15.03	0.87	−0.02	−.0100	875	0.921	−0.007
IQ_{sf4}	100.08	15.16	0.83	−0.08	−0.280	875	0.779	−0.019
FSIQ	100,00	15,03						

[a] Correlation corrected with Levy's formula (1967) modified by Girard and Christensen (2008).
[b] Cohen's d calculated with Rosenthal's formula 2.14 (1991).

Table 4
Mean and standard deviation of the absolute differences between the IQ_{sf} and the FSIQ.

Short form	Mean	SD
IQ_{sf1}	4.24	3.16
IQ_{sf2}	4.92	3.76
IQ_{sf3}	5.30	4.23
IQ_{sf4}	6.28	4.88

with each of the short forms. All the *t* values are very far from the threshold of statistical significance. Cohen's *d*, which measures the effect size with a standard deviation unit, confirms that the four differences are especially low. All are of the order of one hundredth of a standard deviation, very far from the conventional thresholds proposed by Cohen (1988) for a difference to be worth taking into consideration.When we look at the absolute differences between the FSIQ and the IQ_{sf}, that is, without taking into account their plus or minus sign, the assessment is very different, as shown by the results in Table 4. The mean absolute differences range from 4.24 for form 1 to 6.28 for form 4, with standard deviations that go from 3.16 to 4.88.

Table 5 gives a more practical picture of the size of the absolute differences between the FSIQ and each of the IQ_{sf}. For each value of the absolute difference ranging from 0 to 10 and more, this table shows the percentage of subjects of the standardization sample. For IQ_{sf4}, we note that almost a quarter of the subjects show a difference greater than or equal to 10 points, which is very substantial and disqualifies this form as a valid estimate of FSIQ.

Table 6 provides information on identification of individuals with mental retardation or a high intellectual potential on the basis of each of the short forms. The standard thresholds of ≤ 70 and ≥ 130 have been used to identify the individuals in the standardization sample with mental retardation or a high intellectual potential, respectively. The identification made with the short forms has been compared with that made with the full form, taken as the reference, using ROC curves. Table 6 shows the area under the ROC curve, the percentage of true positives and the percentage of false positives obtained with each IQ_{sf} during identification of those with mental retardation and high intellectual potentials.

Figs. 1–4 show the Bland-Altman plots for each short form. Fig. 1 represents the extent of agreement between the FSIQ and IQ_{sf1}. In this graph, the mean score on the FSIQ and on IQ_{sf1} is plotted along the abscissa. The difference between the FSIQ and IQ_{sf1} is plotted along the ordinate. Each point corresponds to the coordinates of an individual on the two axes. The size of the points varies as a function of the number of individuals with the same coordinates. The solid straight line that crosses the point cloud represents the mean difference between the FSIQ and IQ_{sf1} for the entire standardization sample, which is equal to 0.03. The two dotted lines located at 1.96 standard deviations on either side of the mean define an interval which includes 95% of the differences. Since the standard deviation of the distribution of differences between the FSIQ and IQ_{sf1} is equal to 5.29, the upper and lower bounds of this interval are 10.40 and −10.34 respectively.

Fig. 2 represents the extent of agreement between the FSIQ and IQ_{sf2}. The mean of the differences between the FSIQ and IQ_{sf2} is −0.04. Since the standard deviation of the distribution of differences between the FSIQ and IQ_{sf2} is 6.19, the upper bound of the 95% interval on either side of the mean is 12.10, and the lower bound is −12.18. Fig. 3 represents the extent of agreement between the FSIQ and IQ_{sf3}. The mean of the differences between the FSIQ and IQ_{sf2} is −0.23. Since the standard deviation of the distribution of differences between the FSIQ and IQ_{sf2} is 6.78, the upper bound of the 95% interval on either side of the mean is 13.08, and the lower bound is −13.52. Finally, Fig. 4 represents the extent of agreement

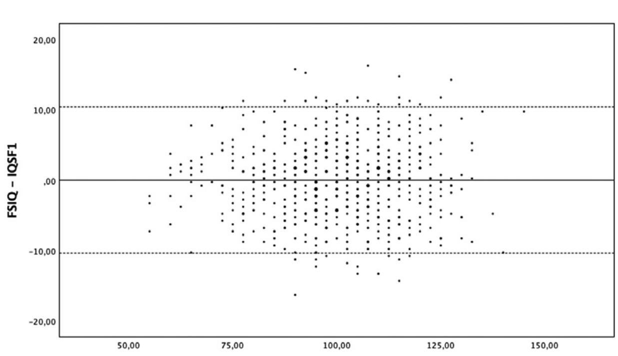

Fig. 1. Bland-Altman plot representing the extend of agreement between the FSIQ and IQ_{sf1}.

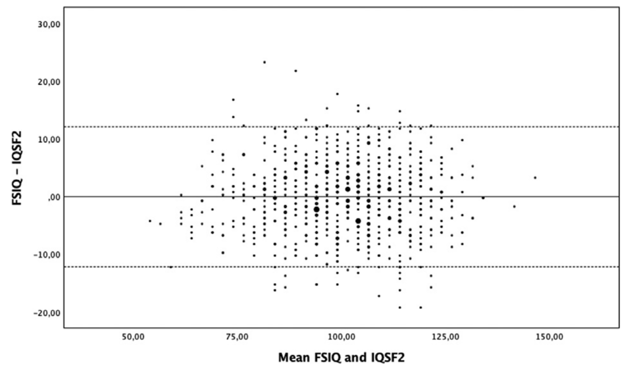

Fig. 2. Bland-Altman plot representing the extend of agreement between the FSIQ and IQ_{sf2}.

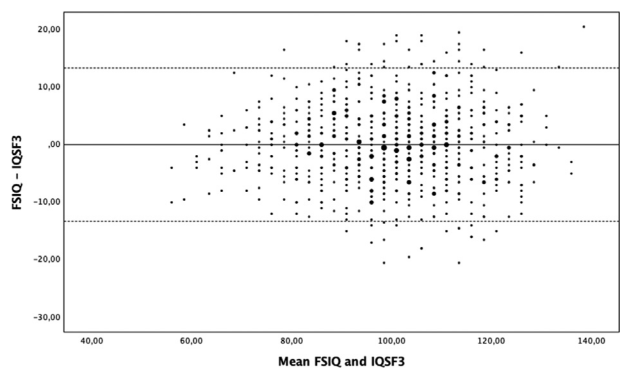

Fig. 3. Bland-Altman plot representing the extend of agreement between the FSIQ and IQ_{sf3}.

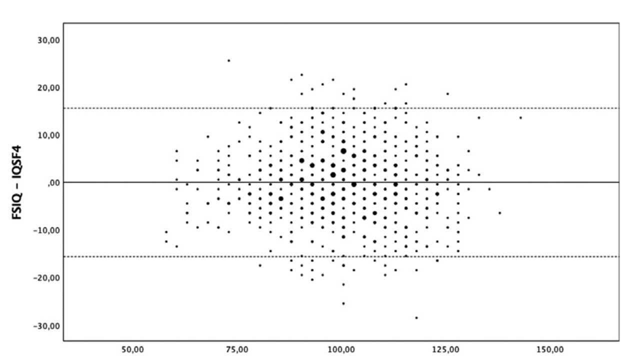

Fig. 4. Bland-Altman plot representing the extend of agreement between the FSIQ and IQ_{sf4}.

Table 5
Percentage of subjects of standardization sample for each absolute difference between the IQ_{sf} and the FSIQ.

FSIQ - IQ_{sf}	IQ_{sf1}	IQ_{sf2}	IQ_{sf3}	IQ_{sf4}
0	6.7	5.6	6.8	6.6
1	15.6	13.4	13.6	7.2
2	14.4	13.8	11.0	10.7
3	10.6	9.9	9.0	11.3
4	13.2	11.1	10.0	8.9
5	9.2	8.0	9.1	8.0
6	7.2	8.8	8.2	7.6
7	5.5	7.8	6.4	6.3
8	6.4	5.3	5.8	5.3
9	4.0	3.9	4.6	5.0
10 and more	7.2	12.4	15.5	23.1

Table 6
Identification of subjects with mental retardation or a high intellectual potential on the basis of each of the IQ_{sf}: areas under the ROC curve and percentages of true and false positives.

	IQ_{sf1}	IQ_{sf2}	IQ_{sf3}	IQ_{sf4}
IQ ≤ 70				
Area under the ROC curve	0.94***	0.83***	0.87***	0.87***
True positive rate	89.3%	67.9%	75.0%	75.0%
False positive rate	0.5%	1.2%	< 0.1%	< 0.1%
IQ ≥ 130				
Area under the ROC curve	0.79***	0.83***	0.66	0.74*
True positive rate	58.3%	66.7%	33.3%	50.0%
False positive rate	0.7%	< 0.1%	< 0.1%	< 0,1%

* $p < .05$.
*** $p < .001$.

Table 7
Reliability coefficient and standard error of the four IQ_{sf} and the FSIQ.

Short form	Reliability	Standard error	90% Bounds	95% Bounds
IQ_{sf1}	0.944	3.556	±5.867	±6.969
IQ_{sf2}	0.949	3.381	±5.579	±6.627
IQ_{sf3}	0.945	3.505	±5.784	±6.870
IQ_{sf4}	0.938	3.727	±6.149	±7.304
FSIQ	0.970	2.598		

between the FSIQ and IQ_{sf4}. The mean of the differences between the FSIQ and IQ_{sf4} is −0.08. Since the standard deviation of the distribution of differences between the FSIQ and IQ_{sf4} is 7.95, the upper bound of the 95% interval on either side of the mean is 15.51, and the lower bound is −15.66. Referring to the four plots, we note that the dispersion of differences around the mean increases from one form to the next, with the 95% interval being narrowest for IQ_{sf1} and broadest for IQ_{sf4}.

Table 7 presents the reliability coefficients of the four short forms, their standard measurement errors, and the values determining the 90 and 95% confidence intervals. We see that the reliability coefficients of the four short forms are all high, but remain lower than that of the FSIQ. Logically, the more subtests the short forms use, the higher their reliability. The least reliable short form is that using only two subtests. Nevertheless, the reliability of this latter form remains very satisfactory.

3. Discussion

On the basis of all the analyses, IQ_{sf1} appears to be the best short form, as it provides the closest estimate to the FSIQ obtained with the full version of the WAIS-IV. The correlation of IQ_{sf1} with the FSIQ is greater than that observed with the other short forms. The mean of the absolute differences compared to the FSIQ is the smallest (4.24), with a smaller standard deviation (3.16). The Brand-Altman dispersion plot (Fig. 1) allows this smaller dispersion of the differences, reflected in a narrower 95% dispersion interval, to be visualized. Although IQ_{sf2} seems at first sight to also be a satisfactory short form, its correlation with the FSIQ is a bit lower. In addition, the mean of the absolute differences compared to the FSIQ is greater than that observed with IQ_{sf1} (4.92), and the 95% dispersion interval represented in the Brand-Altman dispersion plot (Fig. 2) is distinctly broader. IQ_{sf2} appears to perform better than IQ_{sf1} in only a single case, that of identification of persons with high intellectual potential. Its percentage of correct identification is 66.7%, versus 58.3% with IQ_{sf1}. By contrast, IQ_{sf1} is distinctly better than IQ_{sf2} for identifying persons with mental retardation (89.3% versus 67.9%). In terms of their reliability, IQ_{sf1} and IQ_{sf2} can be considered equivalent.

We note that many errors in identification are committed by using the different short forms. However, IQ_{sf1} allows satisfactory identification of persons with mental retardation. On the other hand, the identification of persons with a high intellectual potential is subject to a significant percentage of errors when this short form is used. These errors are undoubtedly due to the inclusion of the *Coding* subtest, which is the least successful test for many individuals with high intellectual potential, in its calculation (Sparrow, Pfeiffer & Newman, 2005).

The other two short forms (IQ_{sf3} and IQ_{sf4}) are distinctly less satisfactory. In particular, the short form including only two subtests can produce very large differences from the FSIQ. With the short form IQ_{sf4}, almost 25% of individuals obtain an IQ that deviates by 10 points or more from the FSIQ obtained with the full form. The size of this difference corresponds to 2/3 of a standard deviation, which is significant. Consequently, this short form should not be used.

Despite the attributes of the short form IQ_{sf1}, it is important to stress its limits and not to systematically substitute it for administration of the full form. IQ_{sf1} provides only an approximation of the FSIQ and can lead to errors in identification. In clinical practice, these errors can have adverse consequences for the persons evaluated. Furthermore, use of this short form eliminates the possibility of calculating the four indices and collecting a great deal of other information useful for diagnosis and comprehension of the cognitive functioning of the persons examined. The short form IQ_{sf1} should therefore only be used if the individual implications of the evaluation are limited. This is the case in research in which the collective results take priority. In this context, the time savings allowed by a short form can be valuable.

Disclosure of interest

The authors declare that they have no competing interest.

References

Axelrod, B. N., Ryan, J. J., & Ward, L. C. (2001). Evaluation of seven-subtest short forms of the Wechsler Adult Intelligence Scale-III in a referred sample. *Archives of Clinical Neuropsychology*, *16*(1), 1–8. https://doi.org/10.1016/S0887-6177(99)00054-2

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods. *Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*, *346*(8982), 1085–1087. https://doi.org/10.1016/S0140-6736(95)91748-9

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Bulzacka, E., Meyers, J. E., Boyer, L., Le Gloahec, T., Fond, G., Szöke, A., Leboyer, M., & Schürhoff, F. (2016). WAIS-IV seven-subtest short form: validity and clinical use in schizophrenia. *Archives of Clincal Neuropsychology*, *31*(8), 915–925. https://doi.org/10.1093/arclin/acw063

Chen, H., & Hua, M. S. (2020). Selecting Tetradic Short Forms of the Taiwan Wechsler Adult Intelligence Scale IV. *Assessment*, *27*(7), 1633–1644. https://doi.org/10.1177/1073191119831787

Cohen, J. (1988). *Statistical power analysis for behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fan, H. Z., Zhu, J. J., Wang, J., Cui, J. F., Chen, N., Yao, J., & Zou, Y. Z. (2019). Four-Subtest Index-based Short Form of WAIS-IV: Psychometric Properties and Clinical Utility. *Archives of Clinical Neuropsychology*, *34*(1), 81–88. https://doi.org/10.1093/arclin/acy016

Girard, T. A., & Christensen, B. K. (2008). Clarifying problems and offering solutions for correlated error when assessing the validity of selected-subtest short forms. *Psychological Assessment*, *20*(1), 76–80. https://doi.org/10.1037/1040-3590.20.1.76

Girard, T. A., Axelrod, B. N., Patel, R., & Crawford, J. R. (2015). Wechsler Adult Intelligence Scale–IV Dyads for Estimating Global Intelligence. *Assessment*, *22*(4), 441–448. https://doi.org/10.1177/1073191114551551

Grégoire, J. (2000). *L'évaluation clinique de l'intelligence de l'enfant*. Liège: Mardaga.

Grégoire, J. (2006). *L'examen clinique de l'intelligence de l'enfant*. Liège: Mardaga.

Grégoire, J. (2013). Measuring components of Intelligence: Mission impossible? *Journal of Psychoeducational Assessment*, *31*(2), 138–147. https://doi.org/10.1177/0734282913478034jpa.sagepub.com

Grégoire, J., & Wierzbicki, C. (2009). Comparaison de quatre short forms de l'échelle d'intelligence de Wechsler pour adultes - troisième édition (WAIS-III). *Revue Européenne de Psychologie Appliquee*, *59*(1), 17–24. https://doi.org/10.1016/j.erap.2007.08.003

Kaufman, J. C., & Kaufman, A. S. (2001). Time for changing the guard: a farewell to short forms of intelligence tests. *Journal of Psychoeducational Assessment*, *19*(3), 245–267. https://doi.org/10.1177/073428290101900305

Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, *23*(1), 84–86. https://doi.org/10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Merz, Z. C., Lace, J. W., Eisenstein, A. M., & Grant, A. F. (2019). Examination of Wechsler adult Intelligence Scale-Fourth Edition short-form IQ estimations in an out-patient psychoeducational sample. *Applied Neuropsychology:Adult*, *0*(0), 1–10. https://doi.org/10.1080/23279095.2019.1687480

Meyers, J. E., Zellinger, M. M., Kockler, T., Wagner, M., & Miller, R. M. (2013). A validated seven-subtest short form for the WAIS-IV. *Applied Neuropsychology: Adult*, *20*(4), 249–256. https://doi.org/10.1080/09084282.2012.710180

Pilgrim, B. M., Meyers, J. E., Bayless, J., & Whetstone, M. M. (1999). Validity of the Ward seven-subtest WAIS-III short form in a neuropsychological population. *Applied Neuropsychology*, *6*(4), 243–246. https://doi.org/10.1207/s15324826an0604_7

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publishing.

Sparrow, S. S., Pfeiffer, S. I., & Newman, T. N. (2005). Assessment of the children who are gifted with the WISC-IV. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV. Clinical use and interpretation*. San Diego, CA: Elsevier Academic Press.

Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*, *31*(5), 499–506. https://doi.org/10.1037/h0024963

van Ool, J. S., Hurks, P. P. M., Snoeijen-Schouwenaars, F. M., Tan, I. Y., Schelhaas, H. J., Klinkenberg, S., & Hendriksen, J. G. M. (2018). Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. *Developmental Neurorehabilitation*, *21*(2), 101–107. https://doi.org/10.1080/17518423.2016.1277799

Ward, L. C. (1990). Prediction of verbal, performance, and full scale IQs from seven subtests of the wais-r. *Journal of Clinical Psychology*, *46*(4), 436–440. https://doi.org/10.1002/1097-4679(199007)46:4<436::AID-JCLP2270460411>3.0.CO;2-M

Wechsler, D. (2008). *WAIS–IV administration and scoring manual*. San Antonio, TX: Pearson.

Wechsler, D. (2009). *WAIS–IV–Manuel d'administration*. Paris: ECPA.