OXFORD
UNIVERSITY PRESS

Archives of CLINICAL NEUROPSYCHOLOGY

Four-Subtest Index-based Short Form of WAIS-IV: Psychometric Properties and Clinical Utility

Hong-Zhen Fan[1], Jian-Jun Zhu[2], Jian Wang[3], Jie-Feng Cui[1], Nan Chen[1], Jing Yao[1], Shu-Ping Tan[1], Jing-Hui Duan[1], Hao-Tian Pang[4], Yi-Zhuang Zou[1,*]

[1]*Psychiatry Research Center, Beijing Huilongguan Hospital, Peking University, Beijing, China*
[2]*Pearson, San Antonio, TX, USA*
[3]*Psychology Clinic, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China*
[4]*Faculty of Psychology, University Complutense of Madrid, Madrid, Spain*

*Corresponding author at: Former Director of Psychiatry Research Center, Beijing Huilongguan Hospital, Peking University, Beijing 100096, China. Tel.: +86-10-62715511 ext. 6472; fax: +86-10-62718210; *E-mail address*: zouyizhuang@bjmu.edu.cn

Editorial Decision 23 January 2018; Accepted 8 February 2018

Abstract

Short form of the Wechsler Adult Intelligence Scale is often needed to quickly estimate intelligence for time-saving or screening in clinical practice. The present study aims to examine the psychometric properties of Chinese version of the four-subtest index-based short form (SF4) of WAIS-IV (FS) and to confirm its clinical application. 1,757 adults from the WAIS-IV Chinese version standardization sample and 239 mixed clinical samples including patients with schizophrenia or schizoaffective disorder (SCH), obsessive–compulsive disorder (OCD), and mild or moderate intellectual disability (ID) were used. Demographic data were collected and intelligence was assessed with WAIS-IV. The SF4 split-half reliability, test–retest stability coefficients and corrected SF4–FS correlations were good to excellent. The result of the Bland–Altman plot showed that the difference fell within 2*SD* was 95% and indicated a random error. The sensitivity, specificity positive predictive value (PPV), and negative predictive value (NPV) of the stepwise screening were good. There was an interaction ($p < .001$) between the IQ level (≥111) and gender on the accuracy of SF4, SF4 might get underestimated on females with the IQ level (≥111) than on males. In conclusion, SF4 is a valid and reliable instrument for use in the clinic, and its clinical application, stepwise screening and influencing factors in clinical use are discussed herein.

Keywords: Schizophrenia; Intelligence; Gender effects; Test construction

Introduction

The short form of the Wechsler Scales is considered to have a long history (Bersoff, 1971; Donders, 2001; Kaufman, Ishikuma, & Kaufmanpacker, 1991; Kaufman, Kaufman, Balgopal, & McLean, 1996; Levy, 1968; Rabin, 1943; Silverstein, 1967, 1990; Smith, McCarthy, & Anderson, 2000; Tellegen & Briggs, 1967; Watkins, 1986). It has been more than 70 years since the short form of Wechsler's original intelligence test, the Wechsler-Bellevue Intelligence Scale (W-B), published by Rabin (1943). Subsequently, a large number of Wechsler scale short forms have been developed (Allen et al., 1997; Axelrod & Naugle, 1998; Axelrod & Paolo, 1998; Bersoff, 1971; Blyler, Gold, Iannone, & Buchanan, 2000; Brooks & Weaver, 2005; Chen, Hua, & Zhu, 2007; Crawford, Allum, & Kinion, 2008; Denney, Ringe, & Lacritz, 2015; Donders, 2001; Donders, Elzinga, Kuipers, Helder, & Crawford, 2013; Donnell, Pliskin, Holdnack, Axelrod, & Randolph, 2007; Doppelt, 1956; Herring, 1952; Kaufman et al., 1991, 1996; Levy, 1968; Meyers, Zellinger, Kockler, Wagner, & Miller, 2013; Paolo & Ryan, 1993; Reynolds, 1978; Ryan, Georgemiller, & McKinney, 1984; Ryan, Kreiner, Gontkovsky, & Umfleet, 2015; Ryan, Weilage, & Spaulding, 1999; Silverstein, 1967, 1990; Smith et al., 2000; Thompson, 1987; Umfleet, Ryan, Gontkovsky, & Morris, 2012; van Duijvenbode, Didden, Hazel, & Engels, 2016; Watkins, 1986; Wymer, Rayls, & Wagner, 2003). Time-consuming (Christensen, Girard, & Bagby, 2007; Kaufman et al., 1991; Rabin, 1943; Silverstein, 1990; Smith et al., 2000;

doi:10.1093/arclin/acy016

Watkins, 1986) and screening large samples for abbreviation were often mentioned in previous studies. The latest full version of the Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV) usually takes 65–90 min to complete (Wechsler, 2008a), which is similar to previous versions; and the amount of time was too long to nibble patient's sufferance or research effort. In addition, most studies focused on the global intelligence function or a group level rather than individuals' specific cognitive information (Axelrod, Dingell, Ryan, & Ward, 2000; Christensen et al., 2007; Donders & Axelrod, 2002). There were still other problems, such as the interference between two tests (Blyler et al., 2000; Christensen et al., 2007; Osato, Gorp, Kern, Satz, & Steinman, 1989; Silverstein, 1990; Smith et al., 2000). For instance, when intelligence was not the primary focus, the intelligence test could be shortened to reduce interferences among some analogous tests in a comprehensive neuropsychological evaluation, especially for cognitive change indicators, such as the MATRICS Consensus Cognitive Battery (MCCB) (Nuechterlein et al., 2008). In some cases, the whole cognitive test's accuracy was more essential than the intelligence information itself. The use of short form is not only about saving time but also more about the research purpose itself (Smith et al., 2000). Three questions cannot be avoided before using the short form: (1) Are the patients capable of completing the full version of the test? (2) Does the short form fulfill the patients' needs? (3) Does the short form meet the clinicians' purpose?

Based on the reasons above, the present study aims to develop a four-subtest short form (SF4) of the WAIS-IV Chinese version based on each index, and uses Equipercentile equating to estimate the full scale score. Equipercentile equating was improved for linear equating even if the relationship between the scores was nonlinear and had a lower random equating error (Kaufman et al., 1996; Kelecioğlu & Gübeş, 2013). The SF4 focuses on the following three aspects: (1) the reliability, including reliability and longitudinal stability; (2) the validity, including correlation with FS scores and consistency with FS scores using classification and Bland–Altman plots; and (3) clinical utility, including effective stepwise screening and influencing factors in clinical use.

Methods

Participants

The WAIS-IV Chinese version standardization sample ($N = 1{,}757$) with ages ranging from 16 to 86 years old was used for this study, which was collected from November 23, 2009 to August 31, 2011. A more detailed description of the sampling procedures is provided in Wang et al. (2013). Eighty-four adults from standardization sample participated in the retest during 13–70 days later (Mean = 32.2) and aged 16–65 years (male = 39, female = 45). The means for age and education were 28.09 years ($SD = 14.39$) and 11.70 years ($SD = 3.10$), respectively. A convenience mixed clinical sample ($N = 239$) with a DSM-IV-TR diagnosis (American Psychiatric Association, 2000) included patients with schizophrenia or schizoaffective disorder (SCH), $n = 119$ (49.8%); obsessive–compulsive disorder (OCD), $n = 30$ (12.6%); and mild or moderate intellectual disability (ID), $n = 90$ (37.7%). The patients with SCH diagnoses included paranoid, $n = 65$ (54.6%), and other, $n = 54$ (45.4%). Patients with SCH and OCD were from Beijing Huilongguan Hospital, while the others were from three special education schools in Beijing, and collected from November 10, 2010 to November 7, 2012. All patients with SCH were assessed during a clinically stable period (Positive and Negative Syndrome Scale positive scale was less 22 points) (Hirayasu, et al., 2016).

Short Form

The standardization sample was used for deriving the short form score. Block Design (BD), Information (IN), and Arithmetic (AR) were selected by high correlation coefficients with FS score for each index (see in Supplementary Table 1), and taking into account the shortest administration time of IN. Coding (CD) was considered on clinical sensitivity (Nuechterlein et al., 2008). The criteria of selecting scales were suggested from Kaufman (1972) and other researchers (De Andraca, Cobo, Rivera, & Pizarro, 1993; Donders, 1997; Silverstein, 1985, 1990). The mean of operation-completed time was 29.0 ($SD = 6.7$), 28.7 ($SD = 5.8$), 32.9 ($SD = 6.1$) and 23.1 ($SD = 7.2$) min in Norm, OCD, SCH, and ID group respectively. Deviation quotients were equated to FS scores by equipercentile equating (see in Supplementary Table 2).

Data Collection

The WAIS-IV full scale was administered to all participants according to the standardized procedures outlined in existing manuals (Wang et al., 2013; Wechsler, 2008a). A questionnaire, which included general information and demographic characteristics, was collected. All subjects were provided with written informed consent after demonstrating their understanding of consent documents. This study was approved by the Institutional Review Board of Beijing Huilongguan Hospital.

Analysis

The split-half and test–retest reliability coefficients were used and recommended by Wechsler (2008b). The SF4–FS correlations were corrected by using Levy's formula (Levy, 1967). A paired *t*-test or Related-Samples Wilcoxon Signed Rank Test was used to compare SF4 and FS or test and retest scores. The intraclass correlation coefficient (ICC) and the Bland–Altman plot (Bland & Altman, 1986) were calculated to evaluate consistency between SF4 and FS. Effect size was computed using Cohen's *d* Formula (Cohen, 1988; Fritz, Morris, & Richler, 2012). The Bland–Altman plot and following methods used the total sample. The relationship between the SF4–FS difference and FS scores used locally weighted regression analysis (LOWESS) and 1,000 times bootstrap resampling. The interaction analysis between IQ group and gender was assessed with a two-way ANOVA. The receiver operating characteristic (ROC) analysis was used to estimate the diagnosis of an IQ less than 90. A p value of less than .05 was considered statistically significant. Statistical analyses were yielded by using the IBM SPSS Statistics 20.0 program (SPSS, 2011), and the plot used the R 3.3.2 package (https://www.r-project.org/).

Results

Demographic Characteristics

The detailed demographic variables are shown in Table 1.

Reliability and Longitudinal Stability

The SF4 split-half coefficients were high, ranging from 0.93 to 0.95 in nine age groups. The average reliabilities for FS and SF4 were 0.98 and 0.94, respectively. The test–retest reliabilities corrected by normative sample for FS and SF4 were 0.91 and 0.90, respectively.

Validity

Relation to FS score. SF4 correlated highly with FS ($r = 0.95$) in the standardization sample, and other groups ranged from 0.93 to 0.96 after correction. The intraclass correlation coefficient (*ICC*) of each group ranged from 0.92 to 0.96 and was similar to the simple correlation coefficients. The mean value and the differences between FS and SF4 in different groups are provided in Table 2. The difference between FS and SF4 in the ID group was significant ($z = 3.248$, $p = .001$, effect size = 0.13), and the effect size could be considered as small (Cohen, 1988). No other differences approached significance.

Consistency with FS scores. In a combined standardization plus mixed clinical sample, the Bland–Altman plot showed that the mean of the SF4–FS difference was –0.04, near zero, and the SF4–FS difference fell within ±4.24 points (2*SEM*) was 65.6%, and 2*SD* was 95% (±9.48 points) (Fig. 1a). The plot also showed a random error between SF4 and FS as in an acceptable range. The SF classification was 73.2% (Norm: 72.3%, OCD: 63.3%, SCH: 86.5%, ID: 91.1%) in the combined sample, which was consistent with FS seven level classifications, and 99.8% (Norm: 99.8%, OCD: 96.7%, SCH: 100.0%, ID: 100.0%) fell within the ±1 level. The consistencies of IQ categories included less than 70 or less than 90 were 98.4% and 92.0% in the combined sample, respectively.

Table 1. Demographic variables in standardization and clinical samples

Demographic variables	Norm ($N = 1757$)	OCD ($N = 30$)	SCH ($N = 119$)	ID ($N = 90$)
Age	37.81 ± 17.95	25.70 ± 5.97	35.89 ± 12.07	26.68 ± 9.62
Gender (Male, %)	834 (47.5)	17 (56.7)	56 (47.1)	56 (62.2)
Education (years)	10.42 ± 3.30	13.40 ± 2.62	12.99 ± 2.97	—

Note: Norm = standardization sample; OCD = obsessive–compulsive disorder; SCH = schizophrenia or schizoaffective disorder; ID = mild or moderate intellectual disability.

Table 2. The correlation coefficients of FS and SF4 in standardization and clinical samples

Group	*N*	FS		SF4		*r*	*r*[a]	*ICC*	*t*/*z*	*p*	Mean difference	Effect size
		Mean	*SD*	Mean	*SD*							
Norm	1757	100.0	15.0	100.0	15.0	.95	0.95	.95	0.152	.879	0.02	0.00
OCD	30	101.8	13.0	101.2	15.7	.93	0.93	.92	0.592	.559	0.60	0.05
SCH	119	90.8	17.0	91.7	17.3	.96	0.96	.96	1.840	.068	−0.90	−0.05
ID[b]	90	51.2	11.8	49.7	13.0	.94	0.95	.94	3.248	.001	1.50	0.13

[a]Corrected by Levy's formula (1967); [b]Wilcoxon test was used. FS = Wechsler Adult Intelligence Scale Fourth Edition Full Scale; SF4 = four-subtest index-based short form of the Wechsler Adult Intelligence Scale Fourth Edition; Norm = standardization sample; OCD = obsessive–compulsive disorder; SCH = schizophrenia or schizoaffective disorder; ID = mild or moderate intellectual disability.

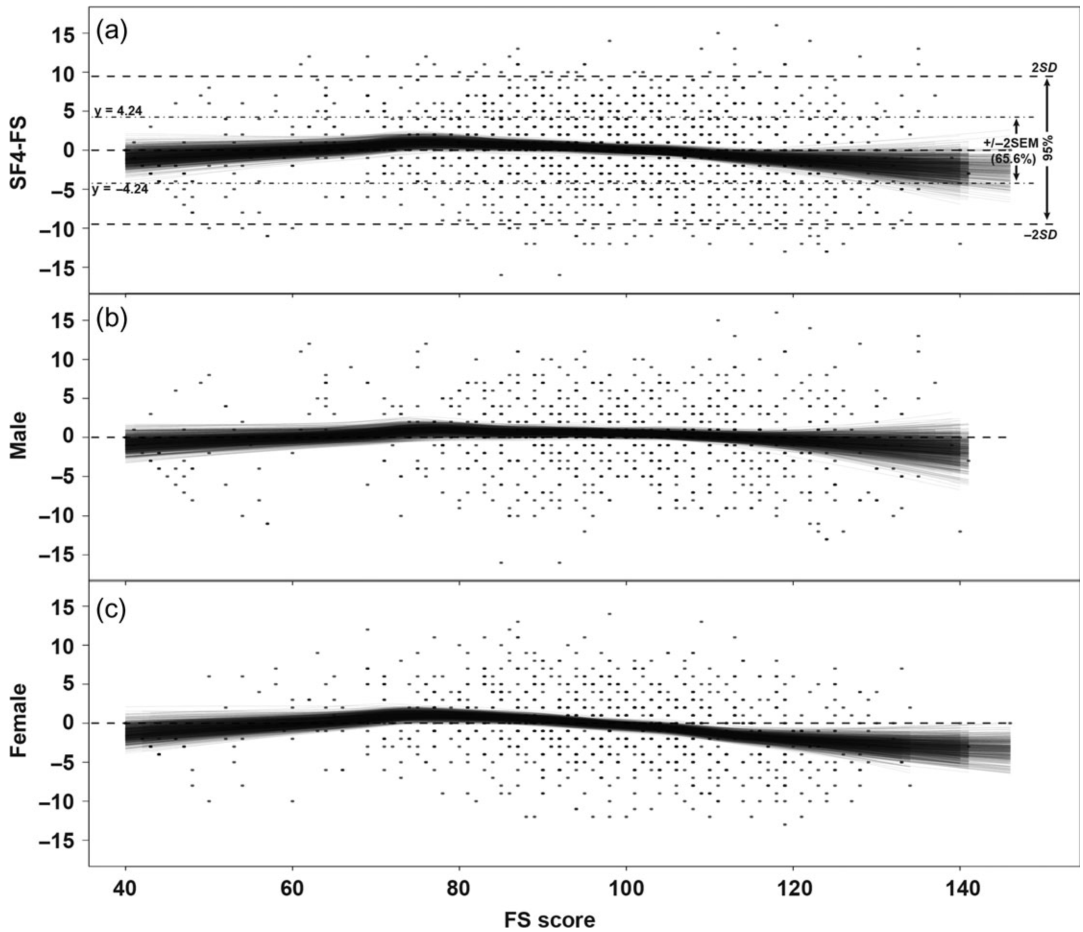

Fig. 1. The association between SF4–FS differences and FS scores in the combined sample.

Clinical Utility

Stepwise screening. The results of the receiver operating characteristic (ROC) analysis of the diagnostic value (IQ < 90) of SF4 in different groups are presented in Table 3. The sensitivity (specificity) were 93.1% (88.0%), 100.0% (84.0) and 94.3% (86.4) in Norm, OCD and SCH group respectively, and 100% classification was consistence with IQ level (IQ < 90) in ID group. To improve the effectiveness of classification, the first two subscales were added and the criterion of scale score (SS) <7 was used, which was one standard deviation below the mean. The AUC of the stepwise screening in the combined sample was 0.900 (95% CI: 0.885–0.912), with 95.6% sensitivity and 84.3% specificity. The positive predictive value (PPV) was 70.9% and negative predictive value (NPV) was 98.0%. The mean (*SD*, *Min*, and *Max*) of the normal IQ (IQ greater or equal to 90) misclassified as suspected ID or ID (IQ less than 90) was 94.60 (4.84, 90, 116), and the converse was 87.52 (1.56, 83, 89).

Table 3. The receiver operating characteristic (ROC) analysis of the SF4 for stepwise screening (IQ < 90) in standardization and clinical samples

Index	SF4					BD/AR/SF4 stepwise screening (Combined sample)
	Norm	OCD	SCH	ID	Combined sample	
AUC	0.906 (0.891–0.919)	0.920 (0.761–0.987)	0.904 (0.836–0.950)	—	0.975 (0.967–0.981)	0.900 (0.885–0.912)
Sensitivity (%)	93.1 (90.3–95.3)	100.0 (47.8–100.0)	94.3 (84.3–98.8)	—	94.4 (92.2–96.1)	95.6 (93.6–97.1)
Specificity (%)	88.0 (86.2–89.7)	84.0 (63.9–95.5)	86.4 (75.7–93.6)	—	87.9 (86.1–89.5)	84.3 (82.3–86.1)
PPV (%)	71.1 (67.1–74.8)	55.6 (21.2–86.3)	84.7 (73.0–92.8)	—	75.7 (72.3–78.8)	70.9 (67.5–74.1)
NPV (%)	97.6 (96.6–98.4)	100.0 (83.9–100.0)	95.0 (86.1–99.0)	—	97.5 (96.5–98.3)	98.0 (97.0–98.7)
Youden index	0.811	0.840	0.807	—	0.823	0.799
Optimal criterion			≤92			<7/<7/≤92

Note: Values in parentheses were 95% confidence intervals. Norm = standardization sample; OCD = obsessive–compulsive disorder; SCH = schizophrenia or schizoaffective disorder; ID = mild or moderate intellectual disability; SF4 = four-subtest index-based short form of the Wechsler Adult Intelligence Scale Fourth Edition; BD = block design; AR = arithmetic; *AUC* = area under the ROC curve. *PPV* = positive predictive value. *NPV* = negative predictive value.

Influencing factors. The correlations of the SF4–FS difference with gender and IQ scores in the combined sample were significant ($r = 0.063$ and –0.081, respectively, $p < .01$) and were still significant even after adjusting for age and education (partial correlation coefficient: 0.066 and –0.065, respectively, $p < .01$). The plot of LOWESS between the SF4–FS difference and IQ scores showed that the SF4-SF difference could be above zero, while IQ scores varied from 60 to 100, especially for scores near 75, and others might be below zero. There was a slight difference between males and females, as shown in Fig. 1b and c. The interaction between the IQ group (less than 110 and more than 111) and gender was significant ($F = 10.705$, $p = .001$), and a further simple effect analysis showed that, the SF4–FS difference in females (Mean =−1.99) was lower ($t = 3.967$, $p < .001$) than males (Mean =−0.02) in the high IQ group (more than 111) and was not significant in the low IQ group (less than 110) ($t = 1.313$, $p = .189$). Females were therefore more likely to be underestimated about 2 scores than males by SF4 when IQs were greater than 111.

Discussion

Reliability and Validity

The reliability and validity of SF4 were acceptable in clinical use. SF4 had a consistence reliability of 0.94 and corrected test–retest reliability of 0.90, which was not as good as the full version but still an excellent value with regard to clinical significance (Cicchetti, 1994; Donders & Axelrod, 2002; Nunnally & Bernstein, 1994; Resnick & Entin, 1971; Smith et al., 2000). The correlation coefficient between SF4 and FS ranged from 0.93 to 0.96 after the correction for redundancy (Levy, 1967) in different samples, similar to most previous studies ranging from 0.92 to 0.97 (Benedict, Schretlen, & Bobholz, 1992; Eisenstein & Engelhart, 1997; Hoffman & Nelson, 1988; Ryan et al., 1984; Ryan, 1983; Thompson, Howard, & Anderson, 1986), and satisfied the criteria suggested by Donders and Axelrod (2002). SF4 demonstrated a reasonable agreement with FS but was still insufficient to use in clinics for diagnosis (Lange, Iverson, Viljoen, & Brink, 2007). The result of classification of individuals was similar to that above and is not supported as an accurate classification as of yet (Alley, Allen, & Leverett, 2007; Devinney, Kamnetz, Chan, & Hattori, 1998; Hoffman & Nelson, 1988; Mccusker, 1994; Ryan et al., 1984; Thompson et al., 1986). Even so, there was still a very high accuracy of 99.8% for classification intervals (plus or minus one level for a rough estimate), and good generalizability in different samples was observed. As a result, SF4 was not valid as an accurate diagnostic tool, but it was useful for a rough estimate or screening.

Clinical Utility

SF4 could quickly screen whether IQs reached a normal level of 90 and above and had a sensitivity of approximately 95% while using the SF4 criteria of not less than 92, whether for a standardization sample or a clinical sample. The stepwise screening had a higher sensitivity and negative predictive value (NPV) and a lower specificity and positive predictive value (PPV), but the use of these methods is more efficient for quickly screening than only using the SF4 total score alone. The other valuable utility of SF4 was to estimate group intelligence. The SF4 mean was close to the FS mean, which was similar to the results of a previous study (Benedict et al., 1992). There were some factors that had influence on the utility of SF4, such as IQ levels. The closer the group intelligence was to 100, the more accurate the SF4 was in estimating the FS. SF4 was overestimated when group intelligence was less than 100, as it was the case in many previous studies (Bawden & Byrne, 1991; Margolis, Taylor, & Greenlief, 1986; Mattis, Hannay, & Meyers, 1992; Ryan et al., 1984; Ryan, Larsen, & Prifitera, 1983; Thompson et al., 1986; van Duijvenbode et al., 2016). The test underestimated individuals with IQs more than 111 especially for females and might be unsuitable for high IQ subjects prone to significant intersubtest scatter (Benedict et al., 1992; Matarazzo, 1972; Matarazzo, Daniel, Prifitera, & Herman, 1988). It is possible that more information was needed to obtain from the individuals with lower or higher scores, and the former's weakness and the latter's strength would be gradually exposed as more information was collected. Moreover, the SF4 contained more operating tests than verbal tests and could therefore underestimate females with high IQ scores more than that of the males. Altogether, the SF4 could be a choice of FS under most clinical situations, and considered as a supplement of the full version in which it is difficult to conduct.

Limitations

First, there are several methodological considerations in abbreviating scales (Levy, 1968), such as atypical samples and dependent norms (Zhu, Cayton, & Chen, 2016). Second, the findings discussed might not be fully extended to other clinical patients with specific neurocognitive deficits (Axelrod & Paolo, 1998; Chen et al., 2007). Moreover, other variables might also influence the use of the short form, such as geographic region, race, disease type and state, drugs and drug dosages that need to be established.

Supplementary Material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

This work was supported by the Capital Foundation of Medical Developments of Beijing (No. 2007-1019). This project was also supported in part with the Capital Health Research and Development of Special of Beijing (No. 2011-2013-2) and the Natural Science Foundation of Beijing Municipality (No. 710208).

Conflict of Interest

None.

Acknowledgements

The authors would like to thank Dong Zhang for his suggestion for stepwise screening and Yi-Hui Xie for advice on the LOWESS method. We would also like to thank all of the participants and evaluators who took part in the study.

References

Allen, D. N., Huegel, S. G., Gurklis, J. A., Jr., Kelley, M. E., Barry, E. J., & van Kammen, D. P. (1997). Utility of WAIS-R short forms in schizophrenia. *Schizophrenia Research*, *26*, 163–172.

Alley, P. J., Allen, R. A., & Leverett, J. P. (2007). Validity of two selected-item short forms of the WAIS-III in an intellectually deficient sample. *Journal of Clinical Psychology*, *63*, 1145–1152.

American Psychiatric Association. (2000). *The diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Axelrod, B. N., Dingell, J. D., Ryan, J. J., & Ward, L. C. (2000). Estimation of Wechsler Adult Intelligence Scale-III index scores with the 7-subtest short form in a clinical sample. *Assessment*, *7*, 157–161.

Axelrod, B. N., & Naugle, R. I. (1998). Evaluation of two brief and reliable estimates of the WAIS-R. *International Journal of Neuroscience*, *94*, 85–91.

Axelrod, B. N., & Paolo, A. M. (1998). Utility of WAIS–R seven-subtest short form as applied to the standardization sample. *Psychological Assessment*, *10*, 33–37.

Bawden, H. N., & Byrne, J. M. (1991). Use of the Hobby WISC—R Short Form with patients referred for neuropsychological assessment. *Psychological Assessment*, *3*, 660–666.

Benedict, R. H., Schretlen, D., & Bobholz, J. H. (1992). Concurrent validity of three WAIS-R Short forms in psychiatric inpatients. *Psychological Assessment*, *4*, 322–328.

Bersoff, D. N. (1971). Short forms of individual intelligence tests for children: Review and critique. *Journal of School Psychology*, *9*, 310–320.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*, 307–310.

Blyler, C. R., Gold, J. M., Iannone, V. N., & Buchanan, R. W. (2000). Short form of the WAIS-III for use with patients with schizophrenia. *Schizophrenia Research*, *46*, 209–215.

Brooks, B. L., & Weaver, L. E. (2005). Concurrent validity of WAIS-III short forms in a geriatric sample with suspected dementia: verbal, performance and full scale IQ scores. *Archives of Clinical Neuropsychology*, *20*, 1043–1051.

Chen, H., Hua, M., & Zhu, J. (2007). Development of the Taiwan WAIS-III Two-Subtest Short Forms. *Psychological Testing*, *54*, 305–330.

Christensen, B. K., Girard, T. A., & Bagby, R. M. (2007). Wechsler Adult Intelligence Scale-Third Edition short form for index and IQ scores in a psychiatric population. *Psychological Assessment*, *19*, 236–240.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crawford, J. R., Allum, S., & Kinion, J. E. (2008). An index-based short form of the WAIS-III with accompanying analysis of reliability and abnormality of differences. *The British Journal of Clinical Psychology*, *47*, 215–237.

De Andraca, I., Cobo, C., Rivera, F., & Pizarro, F. (1993). Intelligence evaluation with short forms of WAIS for populations of low socioeconomic status. *Revista De Saude Publica*, *27*, 334–339.

Denney, D. A., Ringe, W. K., & Lacritz, L. H. (2015). Dyadic Short Forms of the Wechsler Adult Intelligence Scale-IV. *Archives of Clinical Neuropsychology*, *30*, 404–412.

Devinney, D. J., Kamnetz, B., Chan, F., & Hattori, K. (1998). Weschler scale short forms in vocational assessment and evaluation settings. *Vocational Evaluation & Work Adjustment Journal*, *31*, 4–10.

Donders, J. (1997). A short form of the WISC-III for clinical use. *Psychological Assessment*, *9*, 15–20.

Donders, J. (2001). Using a short form of the WISC-III: sinful or smart? *Child Neuropsychology*, *7*, 99–103.

Donders, J., & Axelrod, B. N. (2002). Two-subtest estimations of WAIS-III factor index scores. *Psychological Assessment*, *14*, 360–364.

Donders, J., Elzinga, B., Kuipers, D., Helder, E., & Crawford, J. R. (2013). Development of an eight-subtest short form of the WISC-IV and evaluation of its clinical utility in children with traumatic brain injury. *Child Neuropsychology*, *19*, 662–670.

Donnell, A. J., Pliskin, N., Holdnack, J., Axelrod, B., & Randolph, C. (2007). Rapidly-administered short forms of the Wechsler Adult Intelligence Scale-3rd edition. *Archives of Clinical Neuropsychology*, *22*, 917–924.

Doppelt, J. E. (1956). Estimating the full scale score on the Wechsler adult intelligence scale from scores on four subtests. *Journal of Consulting Psychology*, *20*, 63–66.

Eisenstein, N., & Engelhart, C. I. (1997). Comparison of the K-BIT with Short Forms of the WAIS-R in a neuropsychological population. *Psychological Assessment*, *9*, 57–62.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18.

Herring, F. H. (1952). An evaluation of published short forms of the Wechsler-Bellevue Scale. *Journal of Consulting Psychology*, *16*, 119–123.

Hirayasu, Y., Sato, S. I., Takahashi, H., Iida, S., Shuto, N., Yoshida, S., et al. (2016). A double-blind randomized study assessing safety and efficacy following one-year adjunctive treatment with bitopertin, a glycine reuptake inhibitor, in Japanese patients with schizophrenia. *BMC Psychiatry*, *16*, 66.

Hoffman, R. G., & Nelson, K. S. (1988). Cross-validation of six short forms of the WAIS-R in a healthy geriatric sample. *Journal of Clinical Psychology*, *44*, 952–957.

Kaufman, A. S. (1972). A short form of the Wechsler Preschool and Primary Scale of Intelligence. *Journal of Consulting & Clinical Psychology*, *39*, 361–369.

Kaufman, A. S., Ishikuma, T., & Kaufmanpacker, J. L. (1991). Amazingly short forms of the WAIS—R. *Journal of Psychoeducational Assessment*, *9*, 4–15.

Kaufman, A. S., Kaufman, J. C., Balgopal, R., & McLean, J. E. (1996). Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology*, *25*, 97–105.

Kelecioğlu, H., & Gübeş, N. Ö. (2013). Comparing linear equating and equipercentile equating methods using random groups design. *International Online Journal of Educational Sciences*, *32*, 217–222.

Lange, R. T., Iverson, G. L., Viljoen, H., & Brink, J. (2007). Clinical validation of Canadian WAIS-III Index short forms in inpatient neuropsychiatry and forensic psychiatry. *Clinical Neuropsychologist*, *21*, 434–441.

Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, *23*, 84–86.

Levy, P. (1968). Short-form tests. A methodological review. *Psychological Bulletin*, *69*, 410–416.

Margolis, R. B., Taylor, J. M., & Greenlief, C. L. (1986). A cross-validation of two short forms of the WAIS-R in a geriatric sample suspected of dementia. *Journal of Clinical Psychology*, *42*, 145–146.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence*. New York: Oxford University Press.

Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *Journal of Clinical Psychology*, *44*, 940–950.

Mattis, P. J., Hannay, H. J., & Meyers, C. A. (1992). Efficacy of the Satz-Mogel short form WAIS—R for tumor patients with lateralized lesions. *Psychological Assessment*, *4*, 357–362.

Mccusker, P. J. (1994). Validation of Kaufman, Ishikuma, and Kaufman-Packer's Wechsler Adult Intelligence Scale-Revised short forms on a clinical sample. *Psychological Assessment*, *6*, 246–248.

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*, 249–256.

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., et al. (2008). The MATRICS Consensus Cognitive Battery, part 1: Test selection, reliability, and validity. *American Journal of Psychiatry*, *165*, 203–213.

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

Osato, S. S., Gorp, W. G. V., Kern, R. S., Satz, P., & Steinman, L. (1989). The Satz-Mogel Short Form of the in an elderly, demented population. *Psychological Assessment*, *1*, 339–341.

Paolo, A. M., & Ryan, J. J. (1993). Test-retest stability of the Satz-Mogel WAIS-R short form in a sample of normal persons 75 to 87 years of age. *Archives of Clinical Neuropsychology*, *8*, 397–404.

Rabin, A. I. (1943). A short form of the Wechsler-Bellevue test. *Journal of Applied Psychology*, *27*, 320–324.

Resnick, R. J., & Entin, A. D. (1971). Is an abbreviated form of the Wisc valid for Afro American children? *Journal of Consulting & Clinical Psychology*, *36*, 97–99.

Reynolds, W. M. (1978). A question as to the validity of the verbal scale IQ as a WAIS short form. *Journal of Consulting & Clinical Psychology*, *46*, 1535–1536.

Ryan, J. J. (1983). Clinical utility of a WAIS-R short form. *Journal of Clinical Psychology*, *39*, 261–262.

Ryan, J. J., Georgemiller, R. J., & McKinney, B. E. (1984). Application of the four-subtest WAIS-R short form with an older clinical sample. *Journal of Clinical Psychology*, *40*, 1033–1036.

Ryan, J. J., Kreiner, D. S., Gontkovsky, S. T., & Glass Umfleet, L. (2015). Classification accuracy of sequentially administered WAIS-IV short forms. *Applied Neuropsychology: Adult*, *22*, 409–414.

Ryan, J. J., Larsen, J., & Prifitera, A. (1983). Validity of two- and four-subtest short forms of the WAIS-R in psychiatric sample. *Journal of Consulting & Clinical Psychology*, *51*, 460.

Ryan, J. J., Weilage, M. E., & Spaulding, W. D. (1999). Accuracy of the seven subtest WAIS-R short form in chronic schizophrenia. *Schizophrenia Research*, *39*, 79–83.

Silverstein, A. B. (1967). A short short form of the WISC and WAIS for screening purposes. *Psychological Reports*, *20*, 682.

Silverstein, A. B. (1985). An appraisal of three criteria for evaluating the usefulness of WAIS-R short forms. *Journal of Clinical Psychology*, *41*, 676–680.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 3–11.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102–111.

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

Thompson, A. P. (1987). Methodological issues in the clinical evaluation of two- and four-subtest short forms of the WAIS-R. *Journal of Clinical Psychology*, *43*, 142–144.

Thompson, A. P., Howard, D., & Anderson, J. (1986). Two- and four-subtest short forms of the WAIS-R: Validity in psychiatric sample. *Canadian Journal of Behavioural Science*, *18*, 287–293.

Umfleet, L. G., Ryan, J. J., Gontkovsky, S. T., & Morris, J. (2012). Estimating WAIS-IV Indexes: Proration versus linear scaling in a clinical sample. *Journal of Clinical Psychology*, *68*, 390–396.

van Duijvenbode, N., Didden, R., van den Hazel, T., & Engels, R. C. (2016). Psychometric qualities of a tetrad WAIS-III short form for use in individuals with mild to borderline intellectual disability. *Developmental Neurorehabilitation*, *19*, 26–30.

Wang, J., Zou, Y. Z., Cui, J. F., Fan, H. Z., Chen, N., Yao, J., et al. (2013). Reliability and construct validity of Chinese version of Wechsler Adult Intelligence Scale-Fourth Edition. *Chinese Mental Health Journal*, *27*, 62–67.

Watkins, C. E., Jr. (1986). Validity and usefulness of WAIS-R, WISC-R, and WPPSI short forms. *Professional Psychology: Research and Practice*, *17*, 36–43.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale* (4th ed.) Administration and Scoring Manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale* (4th ed.) Technical and Interpretive Manual. San Antonio, TX: Psychological Corporation.

Wymer, J. H., Rayls, K., & Wagner, M. T. (2003). Utility of a clinically derived abbreviated form of the WAIS-III. *Archives of Clinical Neuropsychology*, *18*, 917–927.

Zhu, J., Cayton, T. G., & Chen, H. (2016). Substitution, proration, or a retest? The optimal strategy when standard administration of the WPPSI-IV Is infeasible. *Psychological Assessment*, *28*, 5800–5814.