Evaluation of seven-subtest short forms of the Wechsler Adult Intelligence Scale-III in a referred sample

Bradley N. Axelrod a,*, Joseph J. Ryan b, L. Charles Ward c

a John D. Dingell Department of Veterans Affairs Medical Center, Detroit, MI, USA
b Dwight D. Eisenhower Department of Veterans Affairs Medical Center, Detroit, MI, USA
c Tuscaloosa Department of Veterans Affairs Medical Center, Detroit, MI, USA

Abstract

The application of seven-subtest short forms of the Wechsler Adult Intelligence Scale-III (WAIS-3) was evaluated in a sample of 281 mixed clinical patients from three Veterans Affairs Medical Centers. Short-form summary scores were derived from deviation quotient tables and from prorating. They included either Block Design or Matrix Reasoning. Short-form summary scores for Full-Scale IQ (FSIQ) and Verbal IQ (VIQ) demonstrated good alternate-forms reliability with the full WAIS-3 scores, whereas Performance IQ (PIQ) summary scores were less accurate. Short forms derived from deviation quotients and prorating did not differ from each other. However, the inclusion of Matrix Reasoning resulted in somewhat better accuracy with WAIS-3 PIQ than did Block Design. The results of this study support the use of the seven-subtest short form of the WAIS-3 in estimating full WAIS-3 summary scores, especially for FSIQ and VIQ. © 2000 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Many shortened versions of the Wechsler Intelligence Scales reduce evaluation time without significantly impacting the reliability or validity of the resulting summary scores. The use of shortened versions of tests is most beneficial for obtaining a quick estimation of intellectual functioning in an individual, for the comparison of group data in a research setting, or for a fast estimation of intelligence (Sattler & Ryan, 1998). The publishers of the Wechsler scales, planning to release the Wechsler Abbreviated Scale of Intelligence in 1999, noted that short forms can also be used for screening evaluations or for vocational rehabilitation purposes.

* Corresponding author. Psychology Service (116B), 4646 John R. Street, Department of Veterans Affairs Medical Center, Detroit, MI 48201-1916.
E-mail address: bradley.axelrod@med.va.gov (B.N. Axelrod).
The seven-subtest short form proposed by Ward (1990) for the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) includes the verbal subtests of Information, Digit Span, Arithmetic, and Similarities. The performance subtests of Picture Completion, Block Design, and Digit Symbol were also included. The time needed to administer this short form ranges from 37 to 45 minutes (Ryan & Rosenberg, 1984; Ward, Selby, & Clark, 1987), in contrast to the almost 75 to 90 minutes for the same individuals to take the entire WAIS-R. In comparison to other short-form estimates, prior research with the WAIS-R consistently found this seven-subtest short form to generate the most reliable estimate of Full-Scale IQ (FSIQ) across a spectrum of clinical samples (e.g., Abraham, Axelrod, & Paolo, 1997; Allen et al., 1997; Benedict, Schretlen, & Bobholz, 1992; Callahan, Schopp, & Johnstone, 1997; Ehrenreich, 1996; Iverson, Myers, Bengston, & Adams, 1996; Satterfield, Martin, & Leiker, 1994; Woodard, Godsall, & Henry, 1996; Zubicaray, Smith, & Anderson, 1996). Similar findings have also been observed in nonclinical individuals (Axelrod & Paolo, 1998; Paolo & Ryan, 1991, 1993; Paolo, Ryan, Ward, & Hilmer, 1996; Schretlen & Ivnik, 1996). Summarizing the findings for the WAIS-R, Pearson correlations of the seven-subtest FSIQ with full WAIS-R FSIQ averaged 0.97, with about 93% of the scores on the short form falling within 5 points of FSIQ obtained on the full version. These findings were the same for clinical and control participants.

In addition to taking one half of the time for administration and generating good estimates of the full FSIQ, the seven-subtest short form for the WAIS-R (Ward, 1990) also allowed the clinician to estimate VIQ and PIQ separately. The comparability for estimated VIQ with full WAIS-R VIQ has consistently been found to have higher Pearson correlations, ranging from .91 to .97, in contrast to the correlations for PIQ, which averaged .83 to .93. Furthermore, the percentage of short-form cases that fell within 5 points of full WAIS-R IQ summary scores were 93% and 78% for VIQ and PIQ, respectively. In a sample of patients with lateralized or diffuse lesions, the VIQ minus PIQ difference scores with the short form fell within 5 points of that of the full WAIS-R for 75% of the sample (Ryan, Abraham, Axelrod, & Paolo, 1996). Interestingly, the seven subtests included in this estimate of full WAIS-R performance were as equally effective when scaled scores were weighted differently as when they were given equal weights in prorating the sum of scaled scores (Iverson, Myers, & Adams, 1997).

The Wechsler Adult Intelligence Scale-III (WAIS-3; Wechsler, 1997a) is the latest revision of the Wechsler Intelligence Scale, which began with the Wechsler-Bellevue. The WAIS-3 generates three summary scores (FSIQ, VIQ, and PIQ) and four factor analytically derived index scores (Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed). The overall test is composed of a total of 14 subtests, one of which (Object Assembly) is considered optional, as it is not included in any of the summary or index scores. Two additional subtests, Letter Number Sequencing and Symbol Search, are not included in the IQ scores, but are used in calculating index scores for the WAIS-3.

Sattler and Ryan (1998) developed deviation IQs using the technique described in Tellegen and Briggs (1967) and information presented in the WAIS-3 manual (Wechsler, 1997b). In addition to presenting data on the original seven-subtest short form, they also included information on an alternate seven-subtest short form that replaces Block Design with Matrix
Reasoning. They found correlations between full FSIQ and short forms with both Block Design and Matrix Reasoning to be 0.98 for all ages in the standardization sample. Correlation among these forms was 0.96 following the correction for redundancy (Levy, 1967). VIQ summary scores averaged a correlation of 0.98 and a corrected correlation of 0.95. For PIQ scores, the short forms with Block Design and Matrix Reasoning correlated 0.94 and 0.95, respectively, with full PIQ. When corrected for redundancy, the correlations were 0.89 when Block Design was used and 0.90 when Matrix Reasoning was included. Although the results with Block Design and Matrix Reasoning were comparable, the practical advantages of limited test equipment, ability to administer to patient in the supine position, and the shorter administration time were given as reasons for preference of Matrix Reasoning (Ryan & Ward, 1999).

The purpose of the present study was to evaluate the utility of seven-subtest short forms in a clinical sample. Not only were we interested in the alternate-forms reliability of FSIQ scores, but also of VIQ and PIQ summary scores. Furthermore, the relative utility of summary scores derived from Block Design versus Matrix Reasoning was assessed. Finally, the use of deviation IQs versus prorated scores to achieve more accurate estimates was evaluated.

1. Method

1.1. Participants

The sample was composed of 281 patients who had been referred for neuropsychological evaluation at one of three Department of Veterans Affairs Medical Centers. The centers each contributed 128, 108, and 45 cases. The sample averaged 52.7 (SD = 14.7) years of age and 12.1 (SD = 2.5) years of reported education. Almost 96% of the sample was male and 90% was right-handed. The ethnic origin of the sample was 65% White, 33% Black, and 2% Hispanic.

1.2. Procedure

In this archival study, participants had completed the WAIS-3 as part of a larger neuropsychological evaluation. The WAIS-3 was administered and scored according to the standardized procedures outlined in the manual (Wechsler, 1997a). As is protocol for the WAIS-3, raw scores were converted to age-corrected scaled score equivalents (Wechsler, 1997a; Table A.1). Sums of scaled scores for verbal, performance, and all subtests were converted to IQ score equivalents (Wechsler, 1997a; Table A.3).

For the short-form IQ scores, data from eight subtests were included. Specifically, the verbal subtests of Similarities, Arithmetic, Digit Span, and Information were included. For both performance estimates, Picture Completion and Digit Symbol-Coding were included. Either Block Design or Matrix Reasoning was added as the third performance subtest. Deviation IQ scores were derived from Sattler and Ryan (1998; Tables O-8, O-9, and O-11) using verbal, performance, and full-scale sum of scaled scores obtained from the WAIS-3.
Prorated sums of scaled scores were obtained by multiplying verbal sum of scaled scores by 6/4 and performance sum of scaled scores by 5/3. Full-scale sum of scaled scores was the total of prorated verbal and performance sum of scaled scores.

2. Results

Mean performance for FSIQ, VIQ, and PIQ as well as for each of the estimated versions of these summary scores appear in Table 1. A repeated measures analysis of variance was performed across each of the full-scale, verbal, and performance summary scores relative to performance with the complete version of the WAIS-3. For FSIQ, the least difference was FSIQ with the estimate derived with the deviation quotient using Block Design, which arithmetically fell 0.45 points lower than FSIQ, $F(1, 280) > 6.43$, $P = .01$. The summary score derived from the full version of the test was statistically higher than each of the other short form estimates, $F(1, 280) > 14.65$, $P < .001$. More interesting, from a practical standpoint, the mean difference scores were at most 1 point higher for the estimates relative to the full version of FSIQ. Whereas FSIQ was higher than estimates of FSIQ, full WAIS-3 VIQ scores averaged 0.7 to 0.8 points lower than estimated VIQ scores, $F(1, 280) > 14.90$, $P < .001$. Short-form estimates of PIQ averaged 1.9 to 2.7 points higher than PIQ obtained from the full WAIS-3, $F(1, 280) > 53.24$, $P < .001$.

The Pearson correlation coefficients between summary scores for the full version and each of the short forms of the WAIS-3 are presented in the fourth column of Table 1. As can clearly be seen, correlations with FSIQ, VIQ, and PIQ were all greater than 0.975, 0.973, and 0.945, respectively. However, correlating short versions of a test with the full version of the same test inflates Pearson correlations by including the mathematical redundancy. Consequently, a correction can be made that incorporates the reliability of the short form as well as the standard deviations of the estimated and full versions of the measure (Levy, 1967). To achieve this end, the reliabilities for each of the short forms were calculated (Moiser, 1943) using the intersubtest correlations for “all ages” (Wechsler, 1997b; Table 4.12) and average reliability coefficients from from the WAIS-3 manual (Wechsler, 1997b; Table 3.1). The results are presented in the third column of Table 1. Reliability scores for all estimates of FSIQ and VIQ fall above the internal consistency cut-off ($r > 0.95$) deemed as “preferred” (Nunnally, 1978, p. 246). The reliability for the PIQ estimates fell above the tolerated cut-off ($r > 0.90$). Correlation coefficients corrected for redundancy were calculated and are in the final column of the table. None of the corrected correlations differed from each other for full-scale ($V < 0.4$, $ns$), verbal ($V = 0.0$, $ns$), or performance ($V < 0.9$, $ns$) scores.

Estimated IQ scores for full scale, verbal, and performance were compared to each other to determine the clinical utility of each of these scores. These comparisons were made for estimations based on deviation quotients as well as those based on prorating. For estimations derived from deviation quotients with either Block Design or Matrix Reasoning, correlations of FSIQ and VIQ estimates did not differ from each other as computed by $t$ test following the Fisher $r$ to $Z$ transformation ($V < 1.40$, $ns$). In contrast, correlations of PIQ with estimates of PIQ were consistently lower than those obtained among the correlations of FSIQ ($V > 5.15$, $P < .001$) and VIQ ($V > 3.87$, $P < .001$) scores. The findings that the short-form to full-
version correlations between FSIQ and VIQ were comparable and that PIQ correlations were significantly lower than the other two sets of scores were repeated when correlations among prorated scores were analyzed.

The percentage of cases of estimated summary scores that fell within 5 and 10 points of actual summary scores was calculated for each of the IQ estimates (see Table 2). For the FSIQ estimates, all versions demonstrated comparable classifications of short-form performance relative to FSIQ performance using all 11 subtests. The proportions were compared using the method described by Bruning and Kintz (1977, pp. 222–225). The same finding was observed with VIQ estimates relative to VIQ derived from the full version of the WAIS-3.

Table 1
Comparison of full and seven-subtest short-form versions of the WAIS-III

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
<th>Reliability</th>
<th>Pearson r</th>
<th>Corrected r</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIQ</td>
<td>87.34 (14.52)</td>
<td>.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSIQ-BD w/DQ</td>
<td>86.53 (14.84)</td>
<td>.965</td>
<td>.977</td>
<td>.946</td>
</tr>
<tr>
<td>FSIQ-MR w/DQ</td>
<td>86.89 (14.96)</td>
<td>.965</td>
<td>.980</td>
<td>.944</td>
</tr>
<tr>
<td>FSIQ-BD Prorated</td>
<td>86.44 (14.32)</td>
<td>.965</td>
<td>.976</td>
<td>.942</td>
</tr>
<tr>
<td>FSIQ-MR Prorated</td>
<td>86.69 (14.43)</td>
<td>.965</td>
<td>.980</td>
<td>.945</td>
</tr>
<tr>
<td>VIQ</td>
<td>89.25 (14.04)</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIQ w/DQ</td>
<td>89.97 (14.34)</td>
<td>.957</td>
<td>.975</td>
<td>.931</td>
</tr>
<tr>
<td>VIQ Prorated</td>
<td>90.05 (13.69)</td>
<td>.957</td>
<td>.973</td>
<td>.931</td>
</tr>
<tr>
<td>PIQ</td>
<td>86.22 (14.73)</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIQ-BD w/DQ</td>
<td>83.84 (15.49)</td>
<td>.917</td>
<td>.945</td>
<td>.858</td>
</tr>
<tr>
<td>PIQ-MR w/DQ</td>
<td>84.33 (15.56)</td>
<td>.915</td>
<td>.960</td>
<td>.871</td>
</tr>
<tr>
<td>PIQ-BD Prorated</td>
<td>83.49 (15.16)</td>
<td>.917</td>
<td>.946</td>
<td>.861</td>
</tr>
<tr>
<td>PIQ-MR Prorated</td>
<td>84.10 (15.09)</td>
<td>.915</td>
<td>.960</td>
<td>.874</td>
</tr>
</tbody>
</table>

Note: WAIS-III = Wechsler Adult Intelligence Scale-III; FSIQ = Full-Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; -BD = estimated with Block Design; -MR = estimated with Matrix Reasoning; w/DQ = estimated derived via deviation quotients.

Table 2
Percentage of short-form WAIS-III protocols that fell within 5 and 10 points of full-version performance on the WAIS-III

<table>
<thead>
<tr>
<th>Variable</th>
<th>± 5 Points</th>
<th>± 10 Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIQ-BD w/DQ</td>
<td>92.6</td>
<td>99.6</td>
</tr>
<tr>
<td>FSIQ-MR w/DQ</td>
<td>94.7</td>
<td>100.0</td>
</tr>
<tr>
<td>FSIQ-BD Prorated</td>
<td>90.8</td>
<td>99.2</td>
</tr>
<tr>
<td>FSIQ-MR Prorated</td>
<td>93.9</td>
<td>99.6</td>
</tr>
<tr>
<td>VIQ w/DQ</td>
<td>91.5</td>
<td>100.0</td>
</tr>
<tr>
<td>VIQ Prorated</td>
<td>93.1</td>
<td>99.7</td>
</tr>
<tr>
<td>PIQ-BD w/DQ</td>
<td>66.6</td>
<td>95.0</td>
</tr>
<tr>
<td>PIQ-MR w/DQ</td>
<td>74.8</td>
<td>97.9</td>
</tr>
<tr>
<td>PIQ-BD Prorated</td>
<td>68.4</td>
<td>94.3</td>
</tr>
<tr>
<td>PIQ-MR Prorated</td>
<td>78.3</td>
<td>96.1</td>
</tr>
</tbody>
</table>

Note: WAIS-III = Wechsler Adult Intelligence Scale-III; FSIQ = Full-Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; -BD = estimated with Block Design; -MR = estimated with Matrix Reasoning; w/DQ = estimated derived via deviation quotients.
In contrast to the lack of differences among the versions of FSIQ and VIQ estimates, short-form estimations of PIQ did differ depending on the subtest used. The use of Block Design in estimating PIQ generated significantly fewer protocols falling within 5 points of full WAIS-3 PIQ than did short form estimates using Matrix Reasoning. This finding was the same whether the scores were derived via deviation quotients ($z = 2.67, P < .01$) or by prorating ($z = 2.14, P < .05$).

3. Discussion

The first finding of import for the present study is that FSIQ and VIQ short-form summary scores are more accurate estimations of full versions than are PIQ short-form scores. This finding replicated the results obtained previously in studies with the WAIS-R (e.g., Axelrod & Paolo, 1998) as well as with the standardization sample of the WAIS-3 (Sattler & Ryan, 1998). The internal consistency of the short forms also demonstrated higher reliability for FSIQ and VIQ relative to PIQ estimations. The lower internal consistency of performance subtests clearly impacted the resulting corrected correlation coefficients of the alternate-forms reliability.

Similar to previous findings (Iverson et al., 1997), there was no significant difference between summary scores derived from deviation quotients versus prorating. Internal consistency, alternate-forms reliability, comparisons of mean performance, and percentage of cases that fell within 5 or 10 points of the full WAIS-3 score all pointed to the same finding. From a practical point of view, clinicians might find simple prorating easier than using a lookup table, but there appears to be no difference in the accuracy of the scores.

The WAIS-3 allows for the option of using Matrix Reasoning in place of Block Design in computing short-form summary scores. In direct comparisons, there was no clinical advantage to using either subtest in calculating FSIQ. However, for estimated PIQ, the inclusion of Matrix Reasoning provided a slight advantage over Block Design. This statistical finding is consistent with the more practical benefits of Matrix Reasoning as well as the shorter length of mean administration time in a sample (Ryan & Ward, 1999). Be that as it may, clinicians using a seven-subtest short form of the WAIS-3 will need to assess the abilities of each patient in determining the practical utility of each subtest. Some clinicians have anecdotally reported that having unlimited time on Matrix Reasoning may unnecessarily lengthen the evaluation session, when the use of Block Design might instead result in discontinuing earlier.

As can be seen in the data presented here, short forms may provide a good estimation of FSIQ and VIQ with group data. However, the implementation for individuals may be less stable and less accurate. In addition, the estimations of PIQ with any of the four methods outlined here are inferior to the short form scores for FSIQ and VIQ. This finding is entirely consistent with the results of studies with the WAIS-R (e.g., Axelrod & Paolo, 1998). As noted before, the reduced internal consistency of the short forms of the WAIS-3 reduces both the alternate-forms reliability (Nunnally, 1978, pp. 243–245) as well as increases the standard error of measurement (Axelrod, Woodard, Schretlen, & Benedict, 1996). However, as presented previously (Ryan & Ward, 1999), the standard errors of measurement are not
vastly different for the seven-subtest FSIQ in comparison to the FSIQ ($SEM = 2.72$) derived from the full WAIS-3 ($SEM = 2.30$).

References


