Archives of Assessment Psychology, Vol. 12, No. 1, (91-105)
Printed in U.S.A.

Preliminary Evidence Demonstrating the Utility of the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II) in the Neurocognitive Assessment of Alzheimer's Disease

Samuel T. Gontkovsky, Psy.D., Joseph J. Ryan, Ph.D., David S. Kreiner, Ph.D., and Gordon Teichner, Ph.D.

Abstract

Although memory deterioration is the primary symptom of Alzheimer's disease (AD), this may not be the first/only deficit of the disorder, necessitating more comprehensive assessment of neuropsychological functioning. This study examined the usefulness of the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II) in the neurocognitive assessment of AD. Participants were 68 individuals classified into AD, depression, and healthy comparison groups administered the WASI-II as part of an outpatient clinical neuropsychological evaluation. Patients with AD obtained significantly lower scores on the WASI-II relative to the standardization sample as well as healthy comparison and depressed participants. Classification accuracy statistics and receiver operating characteristics curve analyses revealed the WASI-II generally showed a moderate degree of diagnostic efficacy at best, with scores on the Full Scale IQ and Block Design subtest being the most effective discriminative measures. Results support the utility of the WASI-II in the neurocognitive evaluation of AD. Findings provide preliminary evidence to suggest the scale is sensitive to the neurocognitive sequelae of the disease and that scores may clarify whether significant intellectual deterioration is part of the presenting clinical picture.

Introduction

Deterioration of episodic memory is the hallmark symptom of Alzheimer's disease (AD) and is detectable in the preclinical stages of the disorder (Gallagher & Koh, 2011; Younan et al., 2020). It is therefore paramount that clinical neuropsychological evaluations of individuals suspected of having or already diagnosed with this disease include one or more formal tests of memory functioning. With the 2011 revision to the diagnostic guidelines for Alzheimer's disease (McKhann et al., 2011) came increased recognition, however, that memory dysfunction may not be the first or only major neurocognitive symptom of this disorder, thereby highlighting the need for greater emphasis on the broader clinical assessment of cognitive domains beyond that of memory.

Examination of intelligence has been a longstanding component of clinical neuropsychological evaluations, and surveys of test usage among clinical neuropsychologists in the United States and Canada have revealed that the various Wechsler intelligence scales have been the most commonly utilized measures in the field for assessing aspects of patient intellectual functioning (Rabin et al., 2005; Rabin et al., 2016). Empirical research also has demonstrated that

the current Fourth Edition of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008a) is sensitive to the neurocognitive sequelae of AD and therefore of value in assessing the manifestations of this disease (Hammers et al., 2020; Joung et al., 2020; Ruchinskas, 2019; Wechsler, 2008b). In cases requiring only an estimate of intellectual abilities, particularly when testing time is limited, as often may be the case among older individuals, secondary to concerns regarding patient fatigue or other factors, clinicians may elect to use either one of the various short forms developed from comprehensive intelligence scales, or a formal abbreviated measure of intelligence. The WASI-II (Wechsler, 2011) was developed to provide a short and reliable method for screening of intelligence. It is composed of four subtests and yields a Verbal Comprehension Index (VCI; Similarities + Vocabulary), Perceptual Reasoning Index (PRI; Block Design + Matrix Reasoning), and Full Scale IQs derived from two subtests (Vocabulary + Matrix Reasoning) or all four subtests. The WASI-II and WAIS-IV have four subtests in common. Correlations between WASI-II and WAIS-IV composite scores are high (Wechsler, 2011), including WAIS-IV Full Scale IQ with WASI-II Full Scale IQ – 4 subtest (corrected r = .92) and Full Scale IQ – 2 subtest (corrected r = .86), Verbal Comprehension indices (corrected r = .88), and Perceptual Reasoning indices (corrected r = .87). The psychometric properties of the WASI-II were shown to be adequate (Wechsler, 2011), and recent studies have provided additional information beyond the scale's manual to assist with interpretation of scores (McGeehan et al., 2017; Ryan & Gontkovsky, 2021).

Although there exist numerous studies investigating the clinical utility of the WASI-II in children and adolescents, samples have been composed primarily of individuals diagnosed with autism spectrum disorders, intellectual disabilities, and Attention-Deficit/Hyperactivity Disorder. Unfortunately, there is a paucity of WASI-II research utilizing adult samples with documented neurological injury or disease. The few reports that do exist have provided support for use of the WASI-II in detecting the deficits associated with traumatic brain injury (Wechsler, 2011) as well as frontotemporal lobar degeneration (Gontkovsky, 2017). The TBI study included in the scale's manual was based on a small sample (n = 21) with moderately severe injuries and examined only group statistics (e.g., means and standard deviations). As expected, the TBI patients obtained significantly lower scores than a healthy comparison group. The second report, although a significant contribution to the WASI-II literature at the time of publication, consisted of an in-depth analysis of a single case of progressive dementia. A more recent investigation focused on differentiating patients with neurological diagnoses from those with psychiatric conditions (Ryan et al., 2021). Findings of this study revealed that both the subtest and composite WASI-II scores of neurological cases were significantly below those of the psychiatric controls, although the pattern of performance across groups was relatively similar. The need for additional investigations to support the use of the WASI-II in the assessment of individuals with neurological disorders is obvious.

The present study therefore was undertaken to examine the power of the WASI-II to detect the neurocognitive sequelae of AD. In order to accomplish this goal, the performance of healthy control participants and patients with AD or depression were compared to that of the WASI-II normative sample. It was hypothesized that participants with AD would obtain scores significantly below those of the normative group, whereas healthy controls and patients with depression would perform at similar levels to each other and to the standardization sample. We simultaneously examined the level and pattern of WASI-II composite and subtest scores across the three study groups. For this analysis, it was hypothesized that patients with AD would score significantly below both comparison groups, and also demonstrate a unique pattern of performance on the WASI-II subtests, as well as the Verbal Comprehension-Perceptual Reasoning Index comparison.

Because significance testing of group means does not address the diagnostic validity of a scale, we supplemented the ANOVA results by providing diagnostic validity statistics (i.e., sensitivity, specificity, hit rate, positive predictive value, and negative predictive value) for the WASI-II scores. These statistics reflect the clinical value of each WASI-II component in the context of specific cutoff scores for group classification. They also allow one to determine the probability that a given individual does or does not have AD (Smith et al., 2003).

Methods

The total sample consisted of 68 individuals referred for outpatient clinical neuropsychological evaluation. The 28 patients comprising the AD group were consecutive adult referrals who met diagnostic criteria for probable major (n = 18) or mild (n = 10) neurocognitive disorder due to Alzheimer's disease as specified by the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013). Comparison groups consisted of 23 consecutive adult referrals who met DSM-5 criteria for diagnosis of a depressive disorder and 17 consecutive adult referrals who did not meet full DSM-5 diagnostic criteria for a specific clinical disorder or condition, who were considered to have a neuropsychological profile consistent with normal aging, and who had no documented or reported history of neurological injury/disease. Nevertheless, it is acknowledged that participants without a diagnosable disorder either presented clinically with subjective cognitive decline or demonstrated one or more signs prompting referral for evaluation of suspected neurocognitive concerns.

Participants were 69% female and 31% male. The racial composition of the sample was 94% White, 2% Latinx/Hispanic, and 4% Black. The sample was 96% right-handed and 4% left-handed. Analyses revealed no significant differences among the groups across any of these aforementioned demographic variables. For the AD, depression, and healthy comparison participants, mean ages were 70.5 (SD = 8.1), 66.3 (7.3), and 62.1 (12.0) years, respectively. There was a significant age difference across groups, $F(2, 65) = 4.68$, $p = .013$, partial $\eta^2 = .13$, with the AD group being significantly older than the healthy comparison group. This difference should be of minimal concern, however, as the WASI-II utilizes age-corrected scores. Years of education were highly similar (i.e., AD M = 14.2 years, SD = 2.6; Depression M = 14.8 years, SD = 2.5; Healthy Controls M = 14.7 years, SD = 3.0), with an absence of significant differences across the groups.

Participants underwent neuropsychological evaluation using a relatively brief, fixed-flexible battery developed for older adult referrals, with the core intellectual measure being the WASI-II. Examinations were conducted by qualified psychometrists under close supervision of a senior neuropsychologist licensed at the independent practice level. All assessments adhered to standard administration procedures, as detailed in the WASI-II manual. None of the patient referrals were for forensic purposes. Nevertheless, participants were administered the Test of Memory Malingering (Tombaugh, 1996) and/or Reliable Digit Span from the WAIS-IV (Wechsler, 2008a) as part of the neuropsychological battery to assess performance validity; two potential participants with depressive disorders were excluded from this study, one secondary to failure on both the second trial and the retention trial of the Test of Memory Malingering and the other secondary to failure on Reliable Digit Span. This retrospective investigation using archival clinical data was compliant with institutional standards for human research and was conducted in accordance with the 1964 Declaration of Helsinki (World Medical Association, 2013) and its subsequent amendments as well as with the ethical principles of the American Psychological Association (2002).

In addition to standard parametric tests (e.g., *t*-tests and ANOVA), data analyses included receiver operating characteristics (ROC) curve analyses and the application of classification accuracy/diagnostic validity statistics (Glaros & Kline, 1988; Smith et al., 2003; Streiner, 2003), including sensitivity (SN), specificity (SP), hit rate (HR), positive predictive value (PPV), and negative predictive value (NPV). For the present study, SN represents the probability that each test score correctly identifies the presence of AD; SP is the probability that each test score correctly identifies the absence of AD; HR is the percentage of individuals who are accurately classified by each test score; PPV is the probability that each test score is accurate when it predicts the presence of AD for the individual case; and NPV is the probability that each test score is accurate when it predicts the absence of AD for the individual case.

Results

Descriptive statistics for the three groups on the WASI-II are presented in Table 1. One-sample *t*-tests were calculated to compare group scores with those of the scale's standardization sample. Mean WASI-II composites and subtest scores of patients with AD were significantly below the standardization sample, all p's < .01, with estimated effect sizes ranging from d = .62 on the Vocabulary subtest to d = .86 on Full Scale IQ. Mean composite and subtest scores of the depression group were uniformly within the average range and did not differ significantly from those of the standardization sample. Likewise, all mean scores for the healthy control group were within the average range and, with the exception of the Block Design subtest, significantly above those of the standardization sample. Estimated effect sizes ranged from d = .60 on the Vocabulary subtest to d = .92 on Full Scale IQ.

Table 1

WASI-II Scores of Patients with Alzheimer's Disease or Depression and Healthy Controls

	Group								
	AD (n = 28)			Depression (n = 23)			Healthy Controls (n = 17)		
	M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range
VCI	88.9	13.9	67-123	104.7	12.6	82-123	107.9	10.0	85-123
PRI	85.2	20.2	54-125	101.7	11.8	79-128	108.9	13.3	76-122
FSIQ	86.0	16.2	59-111	103.7	12.1	82-122	109.7	10.5	78-124
SI	42.8	9.0	28-63	52.5	7.4	33-62	56.0	9.5	39-80
VC	43.6	10.3	27-67	53.7	11.0	32-78	53.9	6.6	43-67
BD	40.4	11.8	25-69	51.8	6.4	37-62	53.3	8.9	34-65
MR	41.0	14.0	22-67	50.3	10.1	28-70	57.4	8.6	37-68

Note. AD = Alzheimer's disease, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ, SI = Similarities, VC = Vocabulary, BD = Block Design, MR = Matrix Reasoning.

Initial between-group analyses examined test scores of participants with AD relative to the comparison groups. A one-way ANOVA on the Full Scale IQs was significant, $F(2, 68) = 19.05$, $p < .001$, $\eta^2 = .370$. Bonferroni post hoc tests indicated that the scores of patients with AD were significantly below, $p <. 001$, those of both participants with depression and healthy controls, but the latter groups performed similarly from a statistical perspective. A two-way mixed ANOVA conducted on the VCI and PRI scores showed a significant main effect of group, $F(2, 65) = 18.84$, $p < .001$, $\eta^2 = .367$. Post hoc Bonferroni tests indicated that the scores of patients with AD were significantly (all $p < .001$) below those of the two comparison groups, the latter of which did not differ from each other. Pattern differences across groups in terms of the VCI-PRI relationship were absent, as the interaction effect was not significant, $F(2,65) = 0.59$, $p = .556$, $\eta^2 = .367$.

The four WASI-II subtest scores were assessed by a two-way mixed ANOVA that yielded a significant main effect for groups, $F(2,65) = 19.73$, $p < .001$, $\eta^2 = .378$. Bonferroni post hoc comparisons indicated that the AD group scored significantly, $p < .001$, lower than the other groups across all WASI-II subtests. Significant differences were not found between the depression and healthy control groups. The interaction effect was not significant, $F(5.23, 170.10) = 0.92$, $p = .470$, $\eta^2 = .028$, indicating the absence of subtest pattern differences across groups. Note that in cases in which Mauchly's test indicated a violation of the sphericity assumption, Greenhouse-Geisser corrected values for *F*-tests are reported.

Table 2

WASI-II Scores of Patients with Alzheimer's Disease and Combined Control Group

	Group						
	AD (n = 28)			Controls (n = 40)			Cohen's *d*
	M	*SD*	Range	*M*	*SD*	Range	
VCI	88.9	13.9	67-123	106.1	11.6	82-123	1.35
PRI	85.2	20.2	54-125	104.8	12.9	76-128	1.15
FSIQ	86.0	16.2	59-111	106.3	11.7	78-124	1.45
SI	42.8	9.0	28-63	54.0	8.2	33-80	1.30
VC	43.6	10.3	27-67	53.8	9.3	32-78	1.04
BD	40.4	11.8	25-69	52.4	7.5	34-65	1.24
MR	41.0	14.0	22-67	53.3	10.0	28-70	1.02

Note. AD = Alzheimer's disease, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ, SI = Similarities, VC = Vocabulary, BD = Block Design, MR = Matrix Reasoning.

In response to these findings, participants designated either as depressed or healthy were combined into a single control group (n = 40), which was used in all subsequent analyses. The combined group had means for age and education of 64.5 years (SD = 9.6) and 14.7 years (SD = 2.7), respectively. While the AD and control groups had similar levels of education (AD M = 14.2 years, SD = 2.6; controls M = 14.7 years, SD = 2.7), the groups differed significantly, (AD: M =

70.5 years, *SD* = 8.1; controls: *M* = 64.5 years, *SD* = 9.6) in age, $t(66) = 2.66$, $p = .01$, $d = .66$. As noted previously, the WASI-II scores are all age-adjusted, a fact that should minimize the effects of this average six-year discrepancy.

Table 2 reports the means, standard deviations, and ranges for AD and control participants on the WASI-II. Patients with AD scored significantly worse than controls on the Full Scale IQ, $t(66) = 5.99$, $p < .001$, $d = 1.45$. A mixed ANOVA on the VCI and PRI scores indicated a significant level of performance difference across groups, $F(1,66) = 35.57$, $p < .001$, $\eta^2 = .350$, indicating that patients with AD scored significantly ($p < .001$) below control participants. There was no VCI-PRI pattern difference across groups, as the interaction effect was nonsignificant, $F(1,66) = 0.44$, $p = .512$, $\eta^2 = .007$. A mixed ANOVA conducted on the four-subtest score means demonstrated that patients with AD were significantly impaired relative to the controls, $F(1,66) = 37.39$, $p < .001$, $\eta^2 = .362$, whereas pattern differences were absent, as the groups by subtests interaction effect was not significant, $F(2.64, 174.36) = 0.28$, $p = .818$, $\eta^2 = .004$.

Table 3

Diagnostic Validity Statistics for WASI-II Composites Using Three Cutoff Values for Identification of Alzheimer's

	SN	SP	HR	PPV	NPV
Cutoff ≤ 85					
VCI	.50	.92	.75	.82	.72
PRI	.61	.92	.79	.85	.77
FSIQ	.57	.90	.76	.80	.75
Cutoff ≤ 77					
VCI	.21	1.0	.68	1.0	.64
PRI	.43	.97	.75	.92	.71
FSIQ	.39	1.0	.75	1.0	.70
Cutoff ≤ 70					
VCI	.04	1.0	.60	1.0	.60
PRI	.25	1.0	.69	1.0	.65
FSIQ	.25	1.0	.69	1.0	.66

Note. SN = Sensitivity, SP = Specificity, HR = Hit Rate, PPV = Positive Predictive Value, NPV = Negative Predictive Value, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ.

Null-hypothesis testing addresses the impact of AD on WASI-II performance, and provides an estimation of how patient scores differ from those of controls. Examination of means, standard deviations, and ranges are of limited assistance to clinicians, however, when faced with diagnostic decision-making for individual cases. Therefore, classification accuracy statistics, as described previously, were calculated to address the diagnostic validity of each WASI-II score (Glaros & Kline, 1988; Smith et al., 2003). Cutoff scores of 1, 1.5, and 2 standard deviations below the normative mean of 100, defined as ≤ 85 for AD and ≥ 86 for controls, ≤ 77 for AD and ≥ 78 for controls, and ≤ 70 for AD and ≥ 71 for control, respectively, were used to analyze the VCIs, PRIs, and Full Scale IQs. For the WASI-II subtests, the three cutoffs were based on the normative subtest

mean T-value of 50 and were defined as ≤ 40 for AD and ≥ 41 for controls, ≤ 34 for AD and ≥ 35 for controls, and ≤ 30 for AD and ≥ 31 for controls. Each of the validity statistics presented in Tables 3 and 4 can range from 0 to 1 and may be interpreted as follows: poor (≤ .69), moderate (≥ .70), strong (≥ .80), and excellent (≥ .90).

Table 4

Diagnostic Validity Statistics for WASI-II Subtests Using Three Cutoff Values for Identification of Alzheimer's

	SN	SP	HR	PPV	NPV
Cutoff ≤ 40					
SI	.39	.95	.72	.85	.69
VC	.39	.92	.70	.79	.68
BD	.61	.95	.81	.89	.78
MR	.57	.90	.76	.80	.75
Cutoff ≤ 34					
SI	.18	.97	.65	.83	.63
VC	.21	.95	.65	.75	.63
BD	.43	.97	.75	.92	.71
MR	.39	.95	.72	.85	.69
Cutoff ≤ 30					
SI	.11	1.0	.63	1.0	.61
VC	.04	1.0	.60	1.0	.60
BD	.25	1.0	.69	1.0	.66
MR	.32	.95	.69	.82	.67

Note. SN = Sensitivity, SP = Specificity, HR = Hit Rate, PPV = Positive Predictive Value, NPV = Negative Predictive Value, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ.

As can be seen from Tables 3 and 4, when considering group-relevant statistics, SN was poor for the four subtests and three composites, while excellent SP was determined for all seven WASI-II components. These findings held regardless of which cutoff scores were applied. It was also clear that SN declined as the cutoff scores became more conservative, whereas SP improved with more conservative scores. In terms of overall HRs, the PRI was the only composite that produced moderate levels of correct classifications for each cutoff value. Block Design HRs at 1 SD and 1.5 SDs below the scale mean score were strong and moderate, respectively. All of the subtests yielded poor HRs when the cutoff for AD was 2 SDs below the T-value average of 50. It should be noted that the percentages of correct classifications (HRs) were all substantially higher than the total sample AD base rate of 41%.

Table 5

Results of ROC Curve Analyses of the WASI-II for Differentiating Patients with Alzheimer's Disease from Controls

	Area Under the Curve	Lower Bound	Upper Bound
VCI	.831	.727	.935
PRI	.795	.671	.920
FSIQ	.850	.761	.939
SI	.827	.772	.933
VC	.778	.661	.894
BD	.806	.681	.930
MR	.747	.618	.876

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ, SI = Similarities, VC = Vocabulary, BD = Block Design, MR = Matrix Reasoning.

The PPVs, which pertain to the diagnostic meaning of a single score, indicated that each of the WASI-II composites were strong to excellent when identifying the individual patient with AD. The PPVs for the subtests suggested that Block Design was the best at recognizing AD. When healthy individuals were the focus (NPV), the composite and subtest scores were poorly to moderately accurate in ruling out AD. Finally, a series of ROC curve analyses was conducted to determine how well each of the WASI-II composite and subtest scores discriminated between patients with AD and control participants. Area under the curve (AUC) statistics are presented in Table 5 and range from .747 for the Matrix Reasoning subtest to .850 for the Full Scale IQ. Recognizing the fact that an AUC of .500 indicates the absence of clinical utility, and a value of 1.000 indicates perfect group discrimination, the values in Table 5 reflect moderate diagnostic effectiveness (Streiner & Cairney, 2007), with Full Scale IQ being the most effective group discriminator. Tables 6 and 7 provide false-positive and false-negative rates for the seven WASI-II scores at three cutoff values. Clearly, in the present study the scale produced high rates of false-negative diagnoses and should not be relied upon to rule in AD in the individual case.

Discussion

The first hypothesis of the study was confirmed. Patients with AD obtained significantly lower WASI-II composite and subtest scores than did the scale's standardization sample. Also, healthy control participants and patients with depression scored at a level similar to that of the normative sample. Our second hypothesis was partially supported as patients with AD performed significantly below healthy elderly controls and patients with depression on the subtests and composites. The intellectual functioning of the two comparison groups was at similar levels and did not differ significantly from one another.

Conversely, the second part of this hypothesis was not supported, as pattern differences were absent across the subtest scores and the VCI-PRI contrasts. Relative to patients with AD, healthy control participants and individuals with depression produced similar score variability in terms of both direction and magnitude. As mentioned previously, the observation that healthy control participants and patients with depression performed similarly on the WASI-II justified combining the samples into a single control group (n = 40) for all subsequent analyses. This was substantiated by the results of a mixed ANOVA on the AD and combined control group. Results

were the same as those of the original three-group analysis in that the AD sample scored significantly below controls on the subtests and composites, with all effect sizes > 1.02, but there were no pattern differences across groups.

Table 6

False Positive and False Negative Error Rates for WASI-II Composites at Three Cutoff Values.

	False Positive	False Negative
Cutoff ≤ 85		
VCI	4.4% (3/68)	20.6% (14/68)
PRI	4.4% (3/68)	16.2% (11/68)
FSIQ	5.9% (4/68)	17.6% (12/68)
Cutoff ≤ 77		
VCI	0.0% (0/68)	32.3% (22/68)
PRI	1.5% (1/68)	23.5% (16/68)
FSIQ	0.0% (0/68)	25.0% (17/68)
Cutoff ≤ 70		
VCI	0.0% (0/68)	38.2% (26/68)
PRI	1.5% (1/68)	23.5% (16/39)
FSIQ	0.0% (0/68)	30.9% (21/68)

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, FSIQ = Full Scale IQ.

To supplement the traditional null hypothesis testing approach, classification accuracy statistics were calculated to address the diagnostic validity of each WASI-II score for the AD versus the combined control groups differential (Smith et al., 2003). Inspection of Tables 3 and 4 indicates that each of the WASI-II composites and subtests has poor sensitivity, regardless of which cutoff value for AD is used. Conversely, excellent specificity was recorded for the composites and subtests across the three cutoffs under study.

As an example, when the T-value cutoff for identifying AD is ≤ 40, the Block Design subtest hit rate of .81 (SN = .61; SP = .95) far exceeds the AD base rate of 41% (28/68). Because Block Design is 61% sensitive, a total of 17 cases were classified as AD (≤ 40 score; true positives) and 11 were incorrectly designated as normal (≥ 41 score; false negatives). The Block Design subtest was 95% specific, indicating that of the 40 control participants, 38 were correctly identified as normal (≥ 41 score; true negatives), and two were misclassified as AD (≤ 40 score; false positive).

When we focus on individual cases, the PPVs and NPVs are of primary interest. In the present study, using a cutoff score of ≤ 40, the PPV is strong (.89), while the NPV is at a moderate level (.78). With an AD base rate of 41%, the PPV indicates that a score of ≤ 40 on Block Design has an 89% chance of correctly identifying AD and the NPV reflects a 78% chance that a cut off ≥ 41 classifies the individual as normal. It appears that a low score on the WASI-II Block Design subtest, when interpreted in conjunction with historical data and results of memory and other neuropsychological measures, could be used to support the likelihood of an AD diagnosis. Overall,

the power to discriminate AD from controls as assessed by ROC curve analysis was at a moderate level for both the subtest (AUC = .747 to .827) and composite (AUC = .795 to .850) scores (Streiner & Cairney, 2007).

Table 7

False Positive and False Negative Error Rates for WASI-II Subtests at Three Cutoff Values.

	False Positive	False Negative
Cutoff ≤ 40		
SI	2.9% (2/68)	25.0% (17/68)
VC	4.4% (3/68)	25.0% (17/68)
BD	1.5% (1/68)	16.2% (11/68)
MR	5.9% (4/68)	17.6% (12/68)
Cutoff ≤ 34		
SI	1.5% (1/68)	33.8% (23/68)
VC	4.4% (2/68)	32.3% (22/68)
BD	1.5% (1/68)	23.5% (1/16)
MR	2.9% (2/68)	25.0% (17/68)
Cutoff ≤ 30		
SI	0.0% (0/68)	36.8% (25/68)
VC	2.9% (2/68)	39.7% (27/68)
BD	0.0% (0/68)	30.9% (21/68)
MR	4.4% (2/68)	27.9% (19/68)

Note. SI = Similarities, VC = Vocabulary, BD = Block Design, MR = Matrix Reasoning.

The PPV and NPV, unlike SN and SP, are influenced by the base rate or prevalence of the condition of interest (i.e., AD). Moreover, experts agree that tests perform best when the base rate is 50% (Streiner, 2003), a value close to that of 41% AD in the current investigation. Had the WASI-II been used to screen the general population of persons ≥ 65 years of age with a 10% base rate for AD (Alzheimer's Association, 2022), the PPVs would have been substantially smaller and the NPVs substantially larger (Glaros, 1988; Smith et al., 2003; Streiner, 2003). In this situation, the score on the Block Design subtest discussed previously would likely be employed most effectively, along with the medical history and results of other neuropsychological measures, to rule out the possibility of AD.

The present results are in line with past research (Gontkovsky, 2017; Ryan et al., 2005; Ryan et al., 2021, Wechsler, 2011), as the WASI-II performance of individuals with brain damage or disease was significantly compromised. This conclusion was verified via contrasting the scores of patients with AD to those of the standardization sample as well as a control group. Unfortunately, a WASI-II score pattern unique to AD was not identified, indicating that the effect of AD was purely a level of performance phenomenon. WASI-II scores were associated with an

unacceptably high false-negative rate. Under the current scenario, these errors are unacceptable because they ignore the presence of a progressive dementia, the sequelae of which will not remit spontaneously but will increase in magnitude and variety. A false-negative error might deprive an individual of the latest medical interventions and delay the development of viable strategies to handle impending placement and financial issues.

Conversely, if the less likely false-positive error is made, healthy individuals might be referred for neuroradiological tests (e.g., computed tomography, magnetic resonance imaging, etc.) and/or scheduled for a series of future assessments to monitor AD progression. When the identification of AD is the focus of assessment, false-positive errors result in both inconvenience and unnecessary expense for the patient.

A competent clinician would not rely exclusively on a brief intelligence test to identify the presence of AD. As part of a battery of measures that assess multiple adaptive capacities (e.g., orientation, attention, memory, visual-spatial analysis, language, executive capabilities, and emotional status), however, the WASI-II can be highly useful. The scale can provide an estimate of the premorbid intellectual level (Holdnack et al., 2013; Pearson, 2009; Ryan et al., 2021) and an approximation concerning the extent of cognitive deterioration. Moreover, observations during scale administration provide an opportunity to evaluate a patient's approach to tasks of verbal fluency, visual-spatial problem solving, and verbal abstraction under standardized conditions.

The detection of problems such as word-finding difficulties, anomia, failure to maintain the square block design configuration, concretization of verbal reasoning, and perseverative tendencies may contribute to the final diagnosis and assist with the development of the appropriate care plan for the individual with AD. WASI-II scores can distinguish between healthy controls and individuals with AD at a moderate level, but this estimate should be considered minimal. In clinical situations, some patients with AD are likely to show evidence of impairment that is not reflected by scores on the WASI-II.

The WASI-II cannot be employed independently to identify cases of AD. Utilized in conjunction with memory measures, such as the Wechsler Memory Scale-Forth Edition (WMS-IV; Wechsler, 2009), however, it can provide important information to aid in the diagnostic process.

Consider the example of a 77-year-old woman with a high school education who earned a WASI-II Full Scale IQ of 109, with a nonsignificant VCI-PRI discrepancy and minimal scatter across the subtests. She was oriented for time and place with grossly intact interpersonal, attention, and language skills and obtained a WMS-IV Immediate Memory Index of 93. When considered in isolation, these results represent a false-negative error, a major limitation of the WASI-II in the present study. Inspection of the five WMS-IV indices alongside those of the WASI-II, however, rendered this false-negative error harmless, as the remaining WMS-IV indices revealed impairments in auditory, visual, and delayed memory. Inspection of the WASI-II scores and the WMS-IV indices should immediately alert one to a problem since the Full Scale IQ of 109 is markedly larger than the Auditory Memory, Visual Memory, Immediate Memory, and Delayed Memory indices by 25, 33, 16, and 44 points, respectively.

This situation underscores the importance of including both intellectual and memory tests in a battery when AD is being considered in the differential diagnoses. This combination of scores, if considered in isolation, likely describes an individual who is capable of presenting a "facade of adequacy." With grossly intact cognitive functions, language, and immediate memory (i.e., Immediate Memory Index = 93), this elderly woman with 12 years of education could appear to be functioning adequately in most superficial everyday encounters, especially upon first meeting

her. Using scores obtained from the WASI-II and WMS-IV in combination, however, allowed the true extent of her memory disability to be detected.

Although memory deficits are the prominent feature of AD, assessing other domains of neurocognition is important not only for diagnostic purposes but also for formulating recommendations and treatment plans. The current findings provide evidence to suggest that the WASI-II not only is moderately sensitive to the cognitive sequelae of AD, but also can be used to address the possibility that intellectual deterioration is part of the presenting clinical picture.

Overall, the generalizability of these findings is limited because of several methodological issues, including the non-probability sampling method, the small sample size, and the fact that patients with AD were, on average, 6 years younger than the control group. Scores also were not adjusted for potential influence of the Flynn effect. Nevertheless, this preliminary data provides evidence to support the utility of the WASI-II in the neurocognitive evaluation of individuals with AD. It also should be noted that the most liberal cutoff scores (i.e., 1 SD below the WASI-II normative sample) provided the best SN and SP estimates, whereas relying on more extreme cutoffs (e.g., 2 SDs below the WASI-II standardization sample) rendered the composites and subtest scores too insensitive for practical use.

On the other hand, if the differential diagnosis involved "normal aging versus nonpsychotic depression versus AD," a Full Scale IQ $\leq$ 70 was a virtual certainty concerning the presence of AD. Also, it is the case that SN and SP figures will differ in samples of patients and controls that are not identical to those in the current investigation. Thus, diagnostic validity statistics will need to be recalculated in future studies since variables such as early versus late onset, duration of illness, and disease severity were not addressed in this study.

Finally, it is important to reiterate the fact that SN and SP are not influenced by the base rate of the condition of interest in a given setting. The PPV and NPV statistics change, however, as the base rate of AD in the population under study changes (Smith et al., 2003). If a clinician uses the WASI-II and WMS-IV to screen referrals at a dementia clinic, the base rate will certainly exceed the 41% figure for AD that characterized the present study. Likewise, if the tests are to be used to screen for AD in a general medical practice, the base rate of the disease will be markedly lower.

To deal with this situation, clinicians can calculate the base rates of interest and derive new cutoff values specific to their own settings. The formulae in Smith et al. can be used to derive diagnostic validity statistics for the various tests used with their patient populations. Additional investigations are needed to either support or refute the clinical utility of the WASI-II for the screening and assessment of patients with suspected or confirmed AD. It would be informative if future investigations examined patients with mild cognitive impairment as well as those with early and late onset of AD. In order to further our knowledge concerning the differential diagnostic utility of the WASI-II, it also would be helpful to contrast the scores of patients with AD to those of individuals diagnosed with other relevant neurological disorders.

About the Authors

Samuel T. Gontkovsky, Psy.D., is affiliated with Empower Behavioral Health and Intervention in Columbus, OH, and The Ohio State University Wexner Medical Center College of Medicine Department of Psychiatry and Behavioral Health in Columbus, OH.

Joseph J. Ryan, Ph.D., ABPP-CN, is former Chair of the Department of Psychological Science and currently an emeritus professor at the University of Central Missouri in Warrensburg, MO.

David S. Kreiner, Ph.D., is a Professor and Department Chair in the Department of Psychological Science at the University of Central Missouri in Warrensburg, MO.

Gordon Teichner, Ph.D., ABPP, is a Neuropsychologist and Licensed Psychologist at Charleston Neuropsychology in Charleston, SC.

References

Alzheimer's Association. (2022). 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 18*, 700–789. https://doi.org/10.1002/alz.12638

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).https://doi.org/10.1176/appi.books.9780890425596

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073. https://www.apa.org/ethics/code

Gallagher, M., & Koh, M. T. (2011). Episodic memory on the path to Alzheimer's disease. *Current Opinion in Neurobiology, 21*, 929–934. https://doi.org/10.1016/j.conb.2011.10.021

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology, 44*(6), 1013–1023. https://doi.org/10.1002/1097-4679(198811)44:6<1013::AID-JCLP2270440627>3.0.CO;2-Z

Gontkovsky, S. T. (2017). Sensitivity of the Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II) to the neurocognitive deficits associated with the semantic dementia variant of frontotemporal lobar degeneration: A case study. *Applied Neuropsychology: Adult, 24*, 288–293. https://doi.org/10.1080/23279095.2016.1154857

Hammers, D. B., Kucera, A., Spencer, R. J., Abildskov, T. J., Archibald, Z. G., Hoffman, J. M., & Wilde, E. A. (2020). Examining the relationship between a verbal incidental learning measure from the WAIS-IV and neuroimaging biomarkers for Alzheimer's pathology. *Developmental Neuropsychology, 45*, 95–109. https://doi.org/10.1080/87565641.2020.1762602

Holdnack, J. A., Schoenberg, M. R., Lange, R. T., & Iverson, G. L. (2013). Predicting premorbid ability for WAIS-IV, WMS-IV and WASI-II. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation* (pp. 217–278). Elsevier.

Joung, H., Yi, D., Byun, M. S., Lee, J. H., Lee, Y., Ahn, H., Lee, D. Y. (2020). Functional neural correlates of the WAIS-IV Block Design test in older adult with mild cognitive impairment and Alzheimer's disease dementia. *Alzheimer's & Dementia, 16*(S6) e044262. https://doi.org/10.1002/alz.044262

McGeehan, B., Ndip, N., & McGill, R. J. (2017). Exploring the multidimensional structure of the WASI-II: Further insights from Schmid–Leiman higher order and exploratory bifactor

solutions. *Archives of Assessment Psychology, 7*, 7–27. https://www.assessmentpsychologyboard.org/journal/index.php/AAP/article/view/86

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Jr., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S. and Phelps, C.H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*, 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Pearson. (2009). *Advanced clinical solutions for use with the WAIS-IV and WMS-IV.* Pearson.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33–65. https://doi.org/10.1016/j.acn.2004.02.005.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability of test usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: Follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*, 206–230. https://doi.org/10.1093/arclin/acw007

Ruchinskas, R. (2019). Wechsler Adult Intelligence Scale-4th Edition Digit Span performance in subjective cognitive complaints, amnestic mild cognitive impairment, and probable dementia of the Alzheimer type. *The Clinical Neuropsychologist, 33*, 1436–1444. https://doi.org/10.1080/13854046.2019.1585574

Ryan, J. J., Carruthers, C. A., Miller, L. J., Souheaver, G. T., Gontkovsky, S. T., & Zehr, M. D. (2005). The WASI Matrix Reasoning subtest: Performance in traumatic brain injury, stroke, and dementia. *International Journal of Neuroscience, 115*, 129–136. https://doi.org/10.1080/00207450490512704

Ryan, J. J., & Gontkovsky, S. T. (2021). Reliabilities of discrepancy scores and supplemental tables for the WASI-II. *Journal of Psychoeducational Assessment, 39*, 930–937. https://doi.org/10.1177/07342829211040595

Ryan, J. J., Kreiner, D. S., Teichner, G., & Gontkovsky, S. T. (2021). Validity of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II) as an indicator of neurological disease/injury: A pilot study. *Brain Injury, 35*, 1624–1629. https://doi.org/doi: 10.1080/02699052.2021.1978547

Smith, G. E., Cerhan, J. H., & Ivnik, R. J. (2003). Diagnostic validity. In D. S., Tulsky, D. H. Saklofske, G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. Bornstein . . . M. F. Ledbetter (Eds.), *Clinical interpretation of the WAIS-Ill and WMS-III* (pp.273–301). Elsevier.

Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic tests. *Journal of Personality Assessment, 81*, 209–219. https://doi.org/10.1207/S15327752JPA8103_03

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *Canadian Journal of Psychiatry, 52*, 121–128. https://doi.org/10.1177/070674370705200210

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM).* Multi-Health Systems.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale–Fourth Edition.* Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition technical and interpretive manual.* Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale-Fourth Edition.* Pearson.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence--Second Edition*. Pearson.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Younan, D., Petkus, A. J., Widaman, K. F., Wang, X., Casanova, R., Espeland, M. A., . . . Chen, J-C. (2020). Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain, 143*, 289–302. https://doi.org/10.1093/brain/awz348