Routledge
Taylor & Francis Group

Journal of Clinical and Experimental Neuropsychology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncen20

The impact of race and other demographic factors on the false positive rates of five embedded Performance Validity Tests (PVTs) in a Veteran sample

John H. Denning & Michael David Horner

To cite this article: John H. Denning & Michael David Horner (2024) The impact of race and other demographic factors on the false positive rates of five embedded Performance Validity Tests (PVTs) in a Veteran sample, Journal of Clinical and Experimental Neuropsychology, 46:1, 25-35, DOI: 10.1080/13803395.2024.2314737

To link to this article: https://doi.org/10.1080/13803395.2024.2314737

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2024, VOL. 46, NO. 1, 25–35
https://doi.org/10.1080/13803395.2024.2314737

Routledge
Taylor & Francis Group

The impact of race and other demographic factors on the false positive rates of five embedded Performance Validity Tests (PVTs) in a Veteran sample

John H. Denning[a,b] and Michael David Horner[a,b]

[a]Mental Health Service, Ralph H. Johnson Veterans Affairs Health Care System, Charleston, SC, USA; [b]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, USA

ABSTRACT

Introduction: It is common to use normative adjustments based on race to maintain accuracy when interpreting cognitive test results during neuropsychological assessment. However, embedded performance validity tests (PVTs) do not adjust for these racial differences and may result in elevated rates of false positives in African American/Black (AA) samples compared to European American/White (EA) samples.

Methods: Veterans without Major Neurocognitive Disorder completed an outpatient neuropsychological assessment and were deemed to be performing in a valid manner (e.g., passing both the Test of Memory Malingering Trial 1 (TOMM1) and the Medical Symptom Validity Test (MSVT), (n = 531, EA = 473, AA = 58). Five embedded PVTs were administered to all patients: WAIS-III/IV Processing Speed Index (PSI), Brief Visuospatial Memory Test-Revised: Discrimination Index (BVMT-R), TMT-A (secs), California Verbal Learning Test-II (CVLT-II) Forced Choice, and WAIS-III/IV Digit Span Scaled Score. Individual PVT false positive rates, as well as the rate of failing two or more embedded PVTs, were calculated.

Results: Failure rates of two embedded PVTs (PSI, TMT-A), and the total number of PVTs failed, were higher in the AA sample. The PSI and TMT-A remained significantly impacted by race after accounting for age, education, sex, and presence of Mild Neurocognitive Disorder. There were PVT failure rates greater than 10% (and considered false positives) in both groups (AA: PSI, TMT-A, and BVMT-R, 12–24%; EA: BVMT-R, 17%). Failing 2 or more PVTs (AA = 9%, EA = 4%) was impacted by education and Mild Neurocognitive Disorder but not by race.

Conclusions: Individual (timed) PVTs showed higher false positive rates in the AA sample even after accounting for demographic factors and diagnosis of Mild Neurocognitive Disorder. Requiring failure on 2 or more embedded PVTs reduced false positive rates to acceptable levels across both groups (10% or less) and was not significantly influenced by race.

ARTICLE HISTORY
Received 19 September 2023
Accepted 11 January 2024

KEYWORDS
Veterans; demographics; neuropsychological tests; performance validity; race

During neuropsychological assessment, it is critical to determine if cognitive test results reflect the patient's true level of ability and the extent of any cognitive decline. Therefore, it has become standard and accepted practice to administer freestanding and/or embedded performance validity tests (PVTs) across all practice settings (clinical, forensic) as well as in research-only samples (Sherman et al., 2020; Shura et al., 2022; Sweet et al., 2021). Embedded PVTs provide some advantages over freestanding PVTs given they are derived from a cognitive test that was likely already planned to be administered, which increases testing time efficiency. However, despite the widespread utilization of numerous embedded PVTs and associated cutoff scores to determine invalid performance (Boone, 2021; Martin et al., 2015), there has been a relative lack of attention to the possible impact of race or ethnicity on the accuracy of these PVTs. It is imperative that one does not utilize PVT cutoffs that (inaccurately) over-identify a specific group (e.g., Dementia) as providing invalid testing when in fact they are likely providing valid performance (Bortnik & Dean, 2021; Denning, 2023; McGuire et al., 2019).

In contrast to the cognitive tests and normative scores from which many embedded PVTs are often based, these PVTs are not adjusted by race. Normative samples have often shown that European Americans/White (EA) individuals often perform significantly better, on average, than African American/Black (AA) samples across cognitive tests; therefore, several cognitive tests have adjustments to account for this difference by providing race-adjusted norms (Heaton et al., 2004; Norman et al., 2011). The differences between race-adjusted and non-race-adjusted scores in normative samples can be substantial, reaching 1+ standard deviations on certain tests (e.g., TMT-B, based on Heaton

CONTACT John H. Denning john.denning@va.gov Mental Health Service (116), Ralph H. Johnson Veterans Affairs Health Care System, 109 Bee St, Charleston, SC 29401, USA

et al., 2004; BVMT-R, based on; Norman et al., 2011), which can have far-reaching diagnostic implications. For example, Norman et al. described the importance of racially-based adjustments for a commonly used visual memory test (BVMT-R, Benedict, 1997). Before adjusting for race-related differences in raw score performance, over 30% percent of the normal functioning AA sample would be viewed as "impaired" (at least −1 SD below the mean). After adjustment, the proportion of the AA sample performing −1 SD were more in line with expectations of approximately 15% (and similar to the EA sample).

A variety of factors for which race serves as a proxy may be contributory, including quality of education, acculturation, literacy, socioeconomic status, effects of racism, stereotype threat, stressful early life events/trauma, test familiarity, access to medical care, toxic exposure, health-related disparities, nutrition, unrepresentative normative samples, and distrust of the examiner/healthcare system (Byrd et al., 2005; Glymour & Manly, 2008; Manly et al., 2004; Tan et al., 2021; Wong et al., 2000). The intended goal of these racial adjustments for cognitive tests is to more accurately determine the precise level of cognitive performance for any specific individual. Given the inherently broad, nonscientific, and evolving categories of socially-constructed racial categories used for cognitive score adjustments, others have encouraged alternative methods of normative adjustments that do not include race (Gasquoine, 2022).

Two recent studies have explored the association between race or ethnicity and false positive rates associated with embedded PVT performance. Hood et al. (2022) compared the performance of 74 EA and 25 AA non-compensation-seeking patients across 13–25 PVT indices (derived from 5 freestanding and 9 embedded tests). The sample excluded those with Major Neurocognitive Disorder, those failing multiple PVTs (including at least 1 "extreme" failure), and ultimately anyone with Borderline IQ (WAIS-III Full Scale IQ less than 80). The latter exclusion criterion eliminated almost half the original AA group (47%), which suggests the sample may not be generalizable to other settings. The authors concluded that most PVTs do not differ by race; however, false positive rates for the b test, Dot Counting, and Stroop color-naming test were quite high in the AA group (false positives greater than 16%). In addition, when interpreting overall failure based on a ratio of failing at least 4 of 14 PVTs, both groups had failure rates of 8% (with the caveat that there is no consensus that this ratio procedure is accepted clinical practice; see Sweet et al., 2021). When utilizing a more conventional criterion of failing 2 or more embedded PVTs, there were slightly more failures in the AA group (56%) compared to the EA group (41%). Oddly, the EA group had twice as many PVTs with false positive rates greater than 10% (12 of 25 indices) compared to the AA group (6 of 25 indices), suggesting that the EA group might actually be more at risk of false positives than the AA group on some measures. Nevertheless, the authors concluded that AA patients may be more at risk of false positives on PVTs that have a processing speed component.

Hromas et al. (2022) analyzed a large sample of AA and EA patients (3 000, 50% AA) from the National Alzheimer's Coordinating Center, excluding those with Major/Mild Neurocognitive Disorder, those found to be cognitively impaired (but apparently not diagnosed with a Neurocognitive Disorder), and a MMSE less than 25. AA and EA groups were matched on age, education, sex, and MMSE score. Results suggested that it was relatively common for the combined sample to fail at least 1 of 6 embedded PVTs (25% to 31% failed 1). There was a higher failure rate in the AA group when relying on the criterion of failing 2 or more embedded PVTs (false positive rates ranged from 13% for raw scores to 21% for age-adjusted scores), whereas false positive rates for the EA group remained between 4% and 6%. Specific embedded PVTs showed low failure rates (less than 10%) across both racial groups (e.g., TMT-ratio score, digit span forward); however, only in the EA group were all embedded PVTs found to have false positive rates below 10%. The authors suggested a need for caution when interpreting embedded PVTs in underrepresented groups as there may need to be cutoff adjustments to improve accuracy.

Other studies have explored race and other demographic/patient factors that could impact performance on embedded and freestanding PVTs. Soble et al. (2023) found that age, education, sex, premorbid IQ, ethnoracial identification, and compensation-seeking accounted for minimal variance (7% to 26%) in the PVTs administered (5 freestanding, 1 embedded). Lower premorbid IQ, and in the Veteran sample, compensation-seeking status, showed the greatest impacts on PVT performance. The latter finding is consistent with other studies describing increased failure rates on the Test of Memory Malingering Trial 1 (TOMM1, Tombaugh, 1996) and/or the Medical Symptom Validity Test (MSVT, Green, 2004) in Veterans already receiving VA-related disability benefits, or simply being in the process of applying for disability (Braun et al., 2021; Horner et al., 2023). Soble et al. concluded that most PVTs were not significantly impacted by patient demographics except for low IQ. Although there was no overall impact of ethnoracial differences on any PVTs in the Soble et al.

study, in their much larger medical center subsample (specifically focusing on the AA and EA groups), failure on 2 or more PVTs was slightly more common in the AA group (25%) compared to the EA group (14%). Braun et al. also found significantly higher TOMM1 failure rates in AA (46%) compared to EA groups (34%); however, there were no such racial differences on the MSVT. Braun et al. concluded that it was possible that the TOMM was racially biased, and that a variety of factors (e.g., stereotype threat, distrust of healthcare system) and demographic factors should be given greater consideration when interpreting PVTs in minoritized groups.

The purpose of the current study is to determine if race is associated with false positive rates across embedded PVTs in a sample screened for valid test performance. In addition, we explored whether the impact of race on failing embedded PVTs remained a significant factor after taking into account multiple demographic and clinical factors known to impact cognitive test performance (e.g., age, education, sex, and cognitive impairment).

Materials and methods

Participants

Veterans referred for an outpatient neuropsychological evaluation and deemed to be providing valid cognitive test data (see procedures section for criteria) were included in the final sample ($n = 531$). The sample included only those identified in the medical record as European American/White (EA, $n = 473$) or African American/Black (AA, $n = 58$). Participants diagnosed with any type of Major Neurocognitive Disorder/Dementia were excluded. The overall sample demographics are as follows: age (mean = 52.2, sd = 14.9, range 21–84), education attainment (mean = 13.0 years, sd = 2.5, range 5–20), 89% EA, 11% AA, and 90% male.

The most common psychiatric diagnoses determined at the time of evaluation included Depressive Disorders (62%), Mild Neurocognitive Disorder (38%), PTSD (33%), history of Alcohol or other Substance Use Disorder (29%), Anxiety Disorders (28%), current Alcohol or other Substance Use Disorder (15%), and ADHD/developmental learning disorder (6%). The most common medical conditions/complaints included TBI (55%, 82% mild), hypertension (50%), elevated lipids (49%), chronic pain (48%), sleep apnea (28%), hearing loss (24%), diabetes (21%), vision problems (17%), coronary artery disease (16%), stroke (6%), and seizure disorder (5%).

Freestanding performance validity tests

The Medical Symptom Validity Test (MSVT, Green, 2004) is a computer administered, forced-choice PVT where 10 pairs of words are presented visually during 2 initial exposure trials. Immediately after presentation of the words the second time, the Immediate Recognition trial is presented where the individual selects the target word from a foil. After 10 min, the Delayed Recognition trial is presented in the same format, with the target words presented with different foils. A Consistency score is also calculated based on the consistency of responses across the 2 previous trials. Failure was based on the above 3 trials consistent with the test manual cutoffs.

The Test of Memory Malingering Trial 1 (TOMM1; Tombaugh, 1996), is a 50-item forced-choice picture recognition PVT. The cutoff for the current study to indicate failure was a score of 40 or less on Trial 1 (Cohen et al., 2022; Denning, 2021; Kraemer et al., 2020; Martin et al., 2020). In addition, given that 98% or more of those making zero errors across the first 10 items go on to pass TOMM1 (Denning, 2021; Grabyan et al., 2018), individuals were also considered to have passed TOMM1 if they were administered only the first 10 items and committed 0 errors (EA group = 12, AA group = 3). Based on clinical impressions at the time of the evaluation, those administered only the first 10 items tended to be younger, had no evidence to suggest possible Dementia, and performed quickly across the first 10 items (Denning, 2021).

Embedded performance validity tests

All patients were administered the same 5 embedded PVTs in the order listed below, within a comprehensive neuropsychological test battery. Many of the embedded PVTs are frequently used by neuropsychologists (LaDuke et al., 2018; Schroeder et al., 2016; Young et al., 2016), were derived from some of the most commonly administered cognitive tests, and assessed multiple cognitive domains (e.g., visual/verbal memory, visual scanning/processing speed, and auditory attention). Even though there are often a variety of cutoffs found in prior studies for these embedded PVTs, cutoffs for the current study were specifically chosen to emphasize low false positives rates in clinical samples (10% or less).

Trailmaking Part A (TMT-A, Heaton et al., 2004) is a timed paper and pencil measure where individuals connect a series of numbered circles, in order, as fast as possible. The cutoff time indicating invalid

performance was determined to be 63 s or greater (Iverson et al., 2002).

The WAIS-III/IV Processing Speed Index (PSI, Wechsler, 1997, 2008) is a timed measure based on the combined performance of Digit Symbol Coding and Symbol Search. The cutoff score indicating invalid performance was determined to be a standard score of 71 or less (Curtis et al., 2009; Etherton et al., 2006; Ovsiew et al., 2020).

The Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997), is a visual-spatial memory test where 6 designs are presented on a single page for 10 s. The recognition discrimination score is the number of correct hits (maximum = 6) minus the number of false positives (maximum = 6). This score ranges from 0–6, with invalid performance reflecting a score of 4 or less (Bailey et al., 2018; Jennette et al., 2022; Olsen et al., 2019; Pliskin et al., 2021; Resch et al., 2023).

WAIS-III/IV Digit Span (DS, Wechsler, 1997, 2008) cutoff used to indicate invalid performance was determined to be an age-corrected scaled score of 5 or less which was consistent across both versions of the WAIS (Greve et al., 2007; Jasinski et al., 2011; Resch et al., 2022; Shura et al., 2020; Spencer et al., 2013; Webber & Soble, 2018; Whitney et al., 2009). Digit Span scaled score was used in the current study because the database of the current study spanned several years before reliable digit span (RDS) was more routinely collected in our clinic. In addition, there have been several studies and reviews of Digit Span scaled score (as cited above) that show very similar (or slightly better) accuracy characteristics compared to RDS.

The California Verbal Learning Test-Second Edition (CVLT-II, Delis et al., 2000) is a list learning verbal memory test. The forced-choice trial was used as the embedded PVT with a cut off of 14 or less indicating invalid performance (Denning, 2012; Resch et al., 2020; Schwartz et al., 2016).

Procedures

Valid cognitive testing was indicated by passing both freestanding PVTs (TOMM1 and MSVT). Research has shown that in clinical samples without Major Neurocognitive Disorder (similar to the current study), failure on even 1 freestanding, forced-choice, memory-based PVT significantly (negatively) impacts cognitive test performance (e.g., 1+ standard deviation decline) and strongly suggests invalid test results (Axelrod & Schutte, 2010; Bar-On Kalfon et al., 2016; Denning, 2021; Donders et al., 2021; Erdodi et al., 2021; Grills & Armistead Jehle, 2016; Keary et al., 2013; Locke et al., 2008; Nauta et al., 2022; Salinsky et al., 2020; Suchy et al., 2012). Both TOMM1 and MSVT have low false positive rates even in patients with mild cognitive decline (Cohen et al., 2022; Resch et al., 2022). As all patients in the present sample passed both TOMM1 and MSVT, failure on any of the 5 embedded PVTs was considered a false positive.

Data analysis

A series of chi-square analysis initially explored race-based differences in failure rates on each of the 5 embedded PVTs, the total number of embedded PVTs failed, and the percentage of patients failing 2 or more embedded PVTs. To determine if race remained associated with PVT failures after taking demographic and other factors into account, we carried out a series of hierarchical logistic regressions, with failure on each embedded PVT as the dependent variable in separate analyses, and age, sex, education, presence of Mild Neurocognitive Disorder, and race as the predictor variables. In each regression, age, sex, and education were entered in Step 1, presence of Mild Neurocognitive Disorder was entered at Step 2, and race was entered at Step 3.

Results

The first set of analyses compared AA ($n = 58$) and EA ($n = 473$) patients on demographic variables (Table 1). There were no significant group differences in age ($t = 1.17$, n.s.) or education ($t = 1.39$, n.s.). Chi-square analyses indicated no significant group differences in the proportion of male patients ($\chi^2 = 1.05$, n.s.) or the proportion of patients who were diagnosed with Mild Neurocognitive Disorder at the time of

Table 1. Characteristics of the sample, stratified by race.

	Black/African-American ($n = 58$)		White/European-American ($n = 473$)	
	Mean	(SD)	Mean	(SD)
Age	54.4	(11.1)	52.0	(15.2)
Years of education	13.4	(2.6)	12.9	(2.4)
Male (%)	86.2		90.5	
Diagnosis of Mild Neurocognitive Disorder (%)	46.6		36.6	

evaluation ($\chi^2 = 2.19$, n.s.).The mean number of embedded PVTs failed in this sample was 0.34 (SD = 0.61). As all patients in the sample had passed both TOMM1 and MSVT, all such failures on embedded PVTs were considered to be false-positives. As a group, AA patients failed slightly more embedded PVTs (mean = 0.53, SD = 0.80) than EA patients (mean = 0.32, SD = 0.58; $t = 2.01$, $p < .05$, Cohen's $d = 0.30$). The next set of analyses examined performance on each of the five embedded PVTs as a function of race. Specifically, a chi-square analysis was performed for each embedded PVT separately, comparing AA and EA patients passing vs. failing the PVT. These results are summarized in Table 2. AA patients scored below the cutoff more frequently than EA patients on the PSI ($\chi^2 = 10.30$, $p = .001$) and TMT-A ($\chi^2 = 4.60$, $p < .05$). There were no group differences in failure rates on the BVMT-R: Discrimination ($\chi^2 = 1.61$, n.s.), CVLT-II Forced Choice ($\chi^2 = 0.28$, n.s.), or Digit Span Scaled Score ($\chi^2 = 0.01$, n.s.). Only 4.5% of the overall sample failed two or more embedded PVTs; although the proportion of patients failing two or more PVTs was not statistically significant between AA and EA groups ($\chi^2 = 2.54$, n.s.), the AA group did show failure rates twice as high as the EA group (AA, $n = 5$, 9% vs. EA, $n = 19$, 4%).

Next, binomial logistic regressions further examined race (AA or EA) as a predictor of failing the embedded PVTs. Performance (pass vs. fail) on each PVT was used as the dependent variable in a separate regression. For each regression, age, sex, and education were entered as predictor variables in Step 1; diagnosis of Mild Neurocognitive Disorder (coded dichotomously as yes/no, based on diagnosis assigned at the time of clinical neuropsychological evaluation) was entered in Step 2; and race (AA or EA) was entered in Step 3. The predictor variables were specified in this sequence in order to delineate the effect of race on PVT performance, after accounting for other demographic and clinical factors that could affect performance on these embedded PVTs (and on cognitive tests more generally).

Table 3 shows logistic regression analysis of predictors of failing the PSI. Lower education was marginally significantly associated with failing the PSI. Diagnosis of Mild Neurocognitive Disorder was strongly associated with failing the PSI. Even after accounting for these effects, race was significantly associated with failing the PSI, with AA patients 4 times as likely as EA patients to fail. Similarly, as shown in Table 4, higher age, lower education, and diagnosis of Mild Neurocognitive Disorder were all significantly associated with failing TMT-A; but even with these factors in the model, being AA was marginally significantly associated with failing. Race was not a significant predictor of performance on the BVMT-R: Discrimination, CVLT-II: Forced Choice, or Digit Span Scaled Score embedded

Table 2. Chi-square analyses of percentage of patients failing each PVT.

Embedded PVTs	Cut Offs	Percentage of patients failing Black/African-American (n = 58)	White/European-American (n = 473)
WAIS-III/IV: PSI **	≤ 71	12.1	3.2
BVMT-R: Discrimination (raw)	≤ 4	24.1	17.3
Trail Making, Part A (sec) *	≥ 63	12.1	5.3
CVLT-II: Forced Choice (raw)	≤ 14	1.7	3.0
WAIS-III/IV: Digit Span (ss)	≤ 5	3.5	3.2
Patients Failing 2+ PVTs		9.0	4.0

*$p < .05$.
**$p = .001$.
Note. BVMT-R = Benton Visual Retention Test, Revised. CVLT-II = California Verbal Learning Test, Second Edition. WAIS-III/IV: PSI = Wechsler Adult Intelligence Scale, Third or Fourth Edition, Processing Speed Index.

Table 3. Binary logistic regression analysis of predictors of failing processing speed index embedded PVT (n = 531).

Predictor Variables	β	S.E.	Wald X^2	df	p	Exp (β)
Age	−.03	.02	3.44	1	.06	.97
Sex	.03	1.09	.00	1	.98	1.03
Years of education	**−.21**	**.11**	**3.86**	**1**	**.05**	**.81**
Diagnosis of mild NCD	**3.81**	**1.05**	**13.13**	**1**	**.00**	**44.92**
Race	**−1.47**	**.54**	**7.54**	**1**	**.01**	**.23**
(Constant)	−.534	2.28	.06	1	.81	.59

Note. PSI embedded PVT was coded as 0 = pass and 1= fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

Table 4. Binary logistic regression analysis of predictors of failing TMT-A embedded PVT ($n = 531$).

Predictor Variables	β	S.E.	Wald X^2	df	*p*	Exp (β)
Age	**.05**	**.02**	**6.85**	**1**	**.01**	**1.06**
Sex	−.20	.80	.06	1	.80	.82
Years of education	**−.25**	**.08**	**8.89**	**1**	**.00**	**.78**
Diagnosis of mild NCD	**1.75**	**.52**	**11.36**	**1**	**.00**	**5.77**
Race	**−.98**	**.50**	**3.80**	**1**	**.05**	**.38**
(Constant)	−2.89	1.79	2.60	1	.11	.06

Note. TMT-A embedded PVT was coded as 0 = pass and 1 = fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

Table 5. Binary logistic regression analysis of predictors of failing BVMT-R: discrimination embedded PVT ($n = 531$).

Predictor Variables	β	S.E.	Wald X^2	df	*p*	Exp (β)
Age	**.03**	**.01**	**9.73**	**1**	**.00**	**1.03**
Sex	.32	.46	.47	1	.50	1.37
Years of education	−.05	.05	1.13	1	.29	.95
Diagnosis of mild NCD	**.90**	**.25**	**13.15**	**1**	**.00**	**2.46**
Race	−.35	.35	1.00	1	.32	.71
(Constant)	−2.88	.95	9.21	1	.00	.06

Note. BVMT-R embedded PVT was coded as 0 = pass and 1 = fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

Table 6. Binary logistic regression analysis of predictors of failing CVLT-2 forced choice embedded PVT ($n = 531$).

Predictor Variables	β	S.E.	Wald X^2	df	*p*	Exp (β)
Age	−.02	.02	1.33	1	.25	.98
Sex	.27	1.06	.07	1	.80	1.31
Years of education	**−.25**	**.13**	**3.83**	**1**	**.05**	**.78**
Diagnosis of mild NCD	−.14	.60	.06	1	.81	.87
Race	.36	1.05	.11	1	.74	1.43
(Constant)	−.02	2.38	.00	1	.99	.98

Note. CVLT-2 Forced Choice was coded as 0 = pass and 1 = fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

Table 7. Binary logistic regression analysis of predictors of failing digit span scaled score embedded PVT ($n = 531$).

Predictor Variables	β	S.E.	Wald X^2	df	*p*	Exp (β)
Age	−.01	.02	.34	1	.56	.99
Sex	.24	1.06	.05	1	.82	1.27
Years of education	−.08	.11	.54	1	.46	.93
Diagnosis of mild NCD	**1.47**	**.58**	**6.47**	**1**	**.01**	**4.34**
Race	.01	.78	.00	1	.99	1.01
(Constant)	−2.90	2.10	1.90	1	.17	.06

Note. Digit Span embedded PVT was coded as 0 = pass and 1 = fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

PVTs after accounting for demographic factors and Mild Neurocognitive Disorder (Tables 5–7).

Finally, a separate logistic regression analysis used as the dependent variable a dichotomous classification of whether the patient failed two or more embedded PVTs (Table 8). Predictor variables were identical to those in the previous logistic regressions and were entered in the same fashion. Lower education and diagnosis of Mild Neurocognitive Disorder were associated with failing two or more PVTs, but race was not a significant predictor after accounting for these other factors.

Discussion

The purpose of the current study was to assess the impact of race (AA or EA) on the false positive rates of several embedded PVTs in a clinical sample of Veterans providing valid testing. Additionally, after

Table 8. Binary logistic regression analysis of predictors of failing two or more embedded PVTs ($n = 531$).

Predictor Variables	β	S.E.	Wald X^2	df	*p*	Exp (β)
Age	−.00	.02	.02	1	.88	1.00
Sex	.25	1.07	.06	1	.82	1.28
Years of education	**−.20**	**.10**	**4.34**	**1**	**.04**	**.82**
Diagnosis of mild NCD	**2.88**	**.77**	**14.11**	**1**	**.00**	**17.81**
Race	−.71	.56	1.63	1	.20	.49
(Constant)	−2.06	2.02	1.04	1	.31	.13

Note. Failing two or more PVTs was coded as 0 = pass and 1 = fail. Sex was coded as 0 = female and 1 = male. Diagnosis of mild neurocognitive disorder was coded as 0 = no and 1 = yes. Race was coded as 0 = Black/African-American and 1 = White/European-American. NCD = Neurocognitive Disorder.

accounting for the effect of several demographic and patient factors, we assessed the association of race with failure on these PVTs. We also took steps to reduce factors associated with false positive rates on validity measures by excluding patients diagnosed with Major Neurocognitive Disorder.Across the combined sample, only 4.5% failed 2 or more of the embedded PVTs. However, the AA sample failed slightly more PVTs, and on 2 of the 5 PVTs (PSI, TMT-A), AA patients failed at significantly higher rates. Although not reaching statistical significance, the AA group also failed 2 or more PVTs at twice the rate as the EA group (9% vs 4%); however, this was still below the conventionally acceptable level of 10%. These differences in failure rates were followed up to determine the extent that race remained a significant factor in PVT failures after accounting for other demographics. Results showed that failing 2 or more embedded PVTs (indicative of false positives in our sample) was most impacted by a diagnosis of Mild Neurocognitive Disorder (and to a much smaller extent lower education), with race not being a significant factor. However, failing the PSI or TMT-A was associated with race even after accounting for age, education, and the presence of Mild Neurocognitive Disorder. Specifically, the AA group was 4 times as likely to fail the PSI, and twice as likely to fail the TMT-A, as the EA group. Failure on the other 3 embedded PVTs (Digit Span, CVLT-II: Forced Choice, BVMT-R: discrimination) was mostly explained by the presence of Mild Neurocognitive Disorder (BVMT-R, digit span), higher age (BVMT-R), and lower education (CVLT-II).

Our findings are consistent with the recommendations of Hromas et al. (2022) that some caution should be used when interpreting failure on individual, embedded PVTs. Our findings are also in line with some of the concerns noted by Hood et al. (2022), where AA groups may be more at risk for false positives on PVTs that have a processing speed component. We found that failure on the PVTs which rely heavily on motor/processing speed (PSI, TMT-A) were most impacted by race. However, unlike Hromas et al. (2022), we found acceptable PVT false positive rates across the entire sample (less than 5%) when relying on the criterion of 2 or more failures. Instead of suggesting alternative cutoffs for AA samples as mentioned by Hromas et al. (2022), it might be more straightforward to simply rely on the failure of 2 or more embedded PVTs to reduce false positives to acceptable levels (10% or less). The higher false positive rates in their AA group may have been at least partially due to their lack of screening to ensure valid test data compared to our rigorous/performance-based screening requiring the passing of 2 freestanding, memory based, forced-choice PVTs. It should be emphasized that failure on 3 of the 5 embedded PVTs in the current study were not significantly impacted by race. In addition, the CVLT-II: Forced Choice and Digit Span showed low false positive rates across both EA and AA groups (2%-4%). Failing 2 or more embedded PVTs was heavily impacted by the presence of mild cognitive decline. This finding highlights the fact that despite individuals experiencing mild cognitive decline, they can still pass 2 freestanding PVTs (TOMM1 and MSVT) compared to some of the embedded PVTs. For example, the BVMT-R: Discrimination was failed by 24% of the AA group and 17% of the EA group. Given this high rate of failure across both groups, and the significant impact of both mild cognitive decline and advanced age, use of this embedded PVT may need to be reexamined in these specific populations regardless of race. Even after taking into account several demographic and patient-related factors, race remained a significant predictor of failure on 2 of the 5 embedded PVTs. Higher failure rates on processing speed-based, embedded PVTs might suggest that situational factors may have a larger effect on these types of tests. For example, stereotype threat, perceived discrimination, and discordance between patient and examiner race have been suggested as possible mechanisms that could decrease performance on cognitive tests and PVTs (Braun et al., 2021; Shewatch et al., 2019; Thames et al., 2013; VanLandingham et al., 2022; Walton & Spencer, 2009). However, despite stereotype threat being viewed as one of the top 8 reasons for PVT failure in clinical settings (Martin et al., 2015), there is no research supporting this speculation. Evidence

showing declines on much more challenging cognitive tests have often relied on studies with small samples (Thames et al., 2013), and demographic factors have been shown to be stronger predictors of test performance than stereotype threat (Whaley, 2021). In addition, several meta-analyses found that the impact of stereotype threat on test performance may be statistically significant, but not highly impactful at the individual level (e.g., $d = -0.14$ to -0.18, Shewatch et al., 2019; Walton & Spencer, 2009). Nevertheless, one should remain aware of various contextual factors that could impact optimal neuropsychological test performance of minority groups and work to mitigate these factors when possible (VanLandingham et al., 2022). In sum, stereotype threat's effects are likely not strong enough to result in failure on embedded PVTs (or freestanding PVTs), but the factors responsible for the race-based differences on embedded PVT failures in our sample require additional study.

Limitations

The current study has certain limitations which may limit the generalizability of the findings. The sample included Veterans who were predominantly male with many co-morbid psychiatric and medical disorders. Although the AA sample was substantially smaller than the EA sample (11% of the sample), this was highly consistent with the most recent US Census data for non-Hispanic AA in the United States (12%, United States Census Bureau, 2021).

Although we rigorously screened for valid test performance, we unfortunately did not have information regarding disability-seeking status, which has been found to significantly increase freestanding PVT failure rates (Horner et al., 2023) and could have impacted the findings to some unknown extent. Despite utilizing some of the most commonly administered freestanding and embedded PVTs, we realize that there are dozens of possible combinations of validity measures which may be unique to any particular practice setting and patient population. Therefore, results may differ if the embedded PVTs used in the current study were administered in a different order. In the current study, freestanding and embedded PVTs were administered to every individual in the same order, which likely reduced variability due to order effects (Zuccato et al., 2018) or missing test data. Finally, we had no available information about the race of the examiner and the possible impact of the patient/examiner race match/mismatch on the rate of embedded PVT failures. Future research may benefit from assessing the possible impact of examiner/patient race mismatch on PVT performance.

Conclusions

The association between race and cognitive test performance has been known for several decades, with multiple factors likely contributing to the lower average performance of AA compared to EA samples. In contrast, the impact of race on embedded PVTs has been given far less scrutiny. Our findings provide evidence that race can impact false positive rates on select embedded PVTs, even after accounting for demographic factors and mild levels of cognitive decline. However, the impact of race was mitigated when relying on the failure of at least 2 out of 5 embedded PVTs, as false positive rates were reduced across both groups to acceptable levels (less than 10%).

Acknowledgments

The views and opinions contained in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Veteran Affairs. This work was supported by the Office of Research and Development, Department of Veterans Affairs.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Axelrod, B. N., & Schutte, C. (2010). Analysis of the dementia profile on the medical symptom validity test. *The Clinical Neuropsychologist*, *24*(5), 873–881. https://doi.org/10.1080/13854040903527295

Bailey, K. C., Soble, J. R., Bain, K. M., & Fullen, C. (2018). Embedded performance validity tests in the hopkins verbal learning test-revised and the brief visuospatial memory test-revised: A replication study. *Archives of Clinical Neuropsychology*, *33*(7), 895–900. https://doi.org/10.1093/arclin/acx111

Bar-On Kalfon, T., Gal, G., Shorer, R., & Ablin, J. N. (2016). Cognitive functioning in fibromyalgia: The central role of effort. *Journal of Psychosomatic Research*, *87*, 30–36. https://doi.org/10.1016/j.jpsychores.2016.06.004

Benedict, R. H. B. (1997). *Brief visuospatial memory test-revised*. Psychological Assessment Resources.

Boone, K. B. (2021). *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford Press.

Bortnik, K. E., & Dean, A. C. (2021). Performance validity testing in patients with dementia. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed., pp. 481–503). Guilford Press.

Braun, S. E., Fountain-Zaragoza, S., Halliday, C. A., & Horner, M. D. (2021). Demographic differences in performance validity test failure. *Applied Neuropsychology: Adult*, *30*(5), 483–491. https://doi.org/10.1080/23279095.2021.1958814

Byrd, D. A., Sanchez, D., & Manly, J. J. (2005). Neuropsychological test performance among Caribbean-born and U.S.-born African American elderly: The role of age, education and reading level. *Journal of Clinical and Experimental Neuropsychology*, *27*(8), 1056–1069. https://doi.org/10.1080/13803390490919353

Cohen, C. D., Rhoads, T., Keezer, R. D., Jennette, K. J., Williams, C. P., Hansen, N. D., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2022). All of the accuracy in half of the time: Assessing abbreviated versions of the Test of Memory Malingering in the context of verbal and visual memory impairment. *The Clinical Neuropsychologist*, *36*(7), 1933–1949. https://doi.org/10.1080/13854046.2021.1908596

Curtis, K. L., Greve, K. W., & Bianchini, K. J. (2009). The Wechsler Adult Intelligence scale-III and malingering in traumatic brain injury: Classification accuracy in known groups. *Assessment*, *16*(4), 401–414. https://doi.org/10.1177/1073191109338161

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test* (2nd ed.). Psychological Corporation.

Denning, J. H. (2012). The efficiency and accuracy of the Test of Memory Malingering trial 1, errors on the first 10 items of the test of memory malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, *27*(4), 417–432. https://doi.org/10.1093/arclin/acs044

Denning, J. H. (2021). When 10 is enough: Errors on the first 10 items of the test of memory malingering (TOMMe10) and administration time predict freestanding performance validity tests (PVTs) and underperformance on memory measures. *Applied Neuropsychology: Adult*, *28*(1), 35–47. https://doi.org/10.1080/23279095.2019.1588122

Denning, J. H. (2023). The TOMM1 discrepancy index (TDI): A new performance validity test (PVT) that differentiates between invalid cognitive testing and those diagnosed with dementia. *Applied Neuropsychology: Adult*, *30*(1), 83–90. https://doi.org/10.1080/23279095.2021.1910951

Donders, J., Lefebre, N., & Goldsworthy, R. (2021). Patterns of performance and symptom validity test findings after mild traumatic brain injury. *Archives of Clinical Neuropsychology*, *36*(3), 394–402. https://doi.org/10.1093/arclin/acz057

Erdodi, L. A., Klyce, D., Mills, A., & Dukarm, P. (2021). Five shades of gray: Conceptual and methodological issues around multivariate models of performance validity. *NeuroRehabilitation*, *49*(2), 179–213. https://doi.org/10.3233/NRE-218020

Etherton, J. L., Bianchini, K. J., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and performance on the WAIS-III processing speed index. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1218–1237. https://doi.org/10.1080/13803390500346595

Gasquoine, P. G. (2022). Performance-based alternatives to race-norms in neuropsychological assessment. *Cortex*, *148*, 231–238. https://doi.org/10.1016/j.cortex.2021.12.003

Glymour, M. M., & Manly, J. J. (2008). Lifecourse social conditions and racial and ethnic patterns of cognitive aging. *Neuropsychology Review*, *18*(3), 223–254. https://doi.org/10.1007/s11065-008-9064-z

Grabyan, J. M., Collins, R. L., Alverson, W. A., & Chen, D. K. (2018). Performance on the test of memory malingering is predicted by the number of errors on its first 10 items on an inpatient epilepsy monitoring unit. *The Clinical Neuropsychologist*, *32*(3), 468–478. https://doi.org/10.1080/13854046.2017.1368715

Green, P. (2004). *Manual for green's medical symptom validity test (MSVT)*. Green's Publishing Inc.

Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., & Ciota, M. A. (2007). Malingering in toxic exposure: Classification accuracy of reliable digit span and WAIS-III Digit Span scaled scores. *Assessment*, *14*(1), 12–21. https://doi.org/10.1177/1073191106295095

Grills, C. E., & Armistead Jehle, P. (2016). Performance validity test and neuropsychological assessment battery screening module performances in an active-duty sample with a history of concussion. *Applied Neuropsychology: Adult*, *23*(4), 295–301. https://doi.org/10.1080/23279095.2015.1079713

Heaton, R. K., Miller, S. W., Taylor, M. J., & Iverson, I. (2004). *Revised comprehensive norms for an expanded halstead-reitan battery: Demographically-adjusted norms for African American and caucasian adults*. Psychological Assessment Resources.

Hood, E. D., Boone, K. B., Miora, D. S., Cottingham, M. E., Victor, T. L., Zeigler, E. A., Zeller, M. A., & Wright, M. J. (2022). Are there differences in performance validity test scores between African American and White American neuropsychology clinic patients? *Journal of Clinical and Experimental Neuropsychology*, *44*(1), 31–41. https://doi.org/10.1080/13803395.2022.2069230

Horner, M. D., Denning, J. H., & Cool, D. L. (2023). Self-reported disability-seeking predicts PVT failure in veterans undergoing clinical neuropsychological evaluation. *The Clinical Neuropsychologist*, *37*(2), 387–401. https://doi.org/10.1080/13854046.2022.2056923

Hromas, G., Rolin, S., & Davis, J. J. (2022). Racial differences in positive findings on embedded performance validity tests. *Applied Neuropsychology: Adult* Advance online publication, 1–9. https://doi.org/10.1080/23279095.2022.2146504

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. D. (2002). Detecting exaggeration and malingering with the trail making test. *The Clinical Neuropsychologist*, *16*(3), 398–406. https://doi.org/10.1076/clin.16.3.398.13861

Jasinski, L. J., Berry, D. T., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult intelligence scale digit span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 300–314. https://doi.org/10.1080/13803395.2010.516743

Jennette, K. J., Rhoads, T., Resch, Z. J., Cerny, B. M., Leib, S. I., Sharp, D. W., Ovsiew, G. P., & Soble, J. R. (2022). Multivariable analysis of the relative utility and additive value of eight embedded performance validity tests for classifying invalid neuropsychological test performance. *Journal of Clinical and Experimental Neuropsychology*, *44*(7), 451–460. https://doi.org/10.1080/13803395.2022.2128067

Keary, T. A., Frazier, T. W., Belzile, C. J., Chapin, J. S., Naugle, R. I., Najm, I. M., & Busch, R. M. (2013). Working memory and intelligence are associated with victoria symptom validity test hard item performance in patients with intractable epilepsy. *Journal of the International Neuropsychological Society*, *19*(3), 314–323. https://doi.org/10.1017/S1355617712001397

Kraemer, L. D., Soble, J. R., Phillips, J. I., Webber, T. A., Fullen, C. T., Highsmith, J. M., Alex, A. W., & Critchfield, E. A.(2020). Minimizing evaluation time while maintaining accuracy: Cross-validation of the test of memory malingering (TOMM) trial 1 and first 10-item errors as briefer performance validity tests. *Psychological Assessment*, *32* (5), 442–450. https://doi.org/10.1037/pas0000802

LaDuke, C., Barr, W., Brodale, D. L., & Rabin, L. A. (2018). Toward generally accepted forensic assessment praactices among clinical neuropsychologists: A survey of professional practice and common test use. *The Clinical Neuropsychologist*, *32*(1), 145–164. https://doi.org/10.1080/13854046.2017.1346711

Locke, D. E., Smigielski, J. S., Powell, M. R., & Stevens, S. R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation*, *23*(3), 273–281. https://doi.org/10.3233/NRE-2008-23310

Manly, J. J., Byrd, D. A., Touradji, P., & Stern, Y. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders. *Applied Neuropsychology*, *11*(1), 37–46. https://doi.org/10.1207/s15324826an1101_5

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of north american professionals. *The Clinical Neuropsychologist*, *29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114–131. https://doi.org/10.1093/arclin/acy012

Nauta, I. M., Bertens, D., van Dam, M., Huiskamp, M., Driessen, S., Geurts, J., Uitdehaag, B., Fasotti, L., Hulst, H. E., de Jong, B. A., & Klein, M. (2022). Performance validity in outpatients with multiple sclerosis and cognitive complaints. *Multiple Sclerosis Journal*, *28*(4), 642–653. https://doi.org/10.1177/13524585211025780

Norman, M. A., Moore, D. J., Taylor, M., Franklin, D., Jr., Cysique, L., Ake, C., Lazarretto, D., Vaida, F., Heaton, R. K., & Group, H. (2011). Demographically corrected norms for African Americans and Caucasians on the hopkins verbal learning test-revised, brief visuospatial memory test-revised, stroop color and Word test, and Wisconsin card sorting test 64-card version. *Journal of Clinical and Experimental Neuropsychology*, *33*(7), 793–804. https://doi.org/10.1080/13803395.2011.559157

Olsen, D. H., Schroeder, R. W., Heinrichs, R. J., & Martin, P. K. (2019). Examination of optimal embedded PVTs within the BVMT-R in an outpatient clinical sample. *The Clinical Neuropsychologist*, *33*(4), 732–742. https://doi.org/10.1080/13854046.2018.1501096

Ovsiew, G. P., Resch, Z. J., Nayar, K., Williams, C. P., & Soble, J. R. (2020). Not so fast! Limitations of processing speed and working memory indices as embedded performance validity tests in a mixed neuropsychiatric sample. *Journal of Clinical and Experimental Neuropsychology*, *42*(5), 473–484. https://doi.org/10.1080/13803395.2020.1758635

Pliskin, J. I., DeDios Stern, S., Resch, Z. J., Saladino, K. F., Ovsiew, G. P., Carter, D. A., & Soble, J. R. (2021). Comparing the psychometric properties of eight embedded performance validity tests in the Rey Yuditory Verbal Learning Test, Wechsler Memory Scale Logical Memory, and Brief Visuospatial memory test-Revised Recognition Trials for detecting invalid neuropsychological test performance. *Assessment*, *28*(8), 1871–1881. https://doi.org/10.1177/1073191120929093

Resch, Z. J., Cerny, B. M., Ovsiew, G. P., Jennette, K. J., Bing-Canar, H., Rhoads, T., & Soble, J. R. (2023). A direct comparison of 10 WAIS-IV digit span embedded validity indi cators among a mixed neuropsychiatric sample with varying degrees of cognitive impairment. *Archives of Clinical Neuropsychology*, *38*(4), 619–632. https://doi.org/10.1093/arclin/acac082

Resch, Z. J., Rhoads, T., Ovsiew, G. P., & Soble, J. R. (2022). A known-groups validation of the medical symptom validity test and analysis of the genuine memory impairment profile. *Assessment*, *29*(3), 455–466. https://doi.org/10.1177/1073191120983919

Salinsky, M., Binder, L., Storzbach, D., Parko, K., Rutecki, P., Goy, E., & Tadrous-Furnanz, S. (2020). Validity testing in veterans with epileptic seizures and psychogenic nonepileptic seizures. *Epilepsy and Behavior*, *111*, 107246. https://doi.org/10.1016/j.yebeh.2020.107246

Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. *The Clinical Neuropsychologist*, *30*(4), 515–535. https://doi.org/10.1080/13854046.2016.1177118

Schwartz, E. S., Erdodi, L., Rodriguez, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A., & Roth, R. M. (2016). CVLT-II forced choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society*, *22* (8), 851–858. https://doi.org/10.1017/S1355617716000746

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shewach, O. R., Sackett, P. R., & Quint, S.(2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, *104*(12), 1514–1534. https://doi.org/10.1037/apl0000420

Shura, R. D., Denning, J. H., Armistead Jehle, P., & Axelrod, B. A.(2022). Validity assessment within VA and active duty. In R. W. Schoeder, & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing non-credible performance* (pp. 235–253). Guilford.

Shura, R. D., Martindale, S. L., Taber, K. H., Higgins, A. M., & Rowland, J. A. (2020). Digit Span embedded validity indicators in neurologically-intact veterans. *The Clinical Neuropsychologist*, *34*(5), 1025–1037. https://doi.org/10.1080/13854046.2019.1635209

Soble, J. R., Cerny, B. M., Rhoads, T., DeBoer, A. B., Sharp, D. W., Ovsiew, G. P., Phillips, M. S., Pesanti, S. D., Jennette, K. J., & Resch, Z. J. (2023). Do demographic factors influence detection of invalid neuropsychological test performance using common performance validity tests? A multisite investigation. *Neuropsychology*, *37*(2), 218–232. https://doi.org/10.1037/neu0000880

Spencer, R. J., Axelrod, B. N., Drag, L. L., Waldron-Perrine, B., Pangilinan, P. H., & Bieliauskas, L. A. (2013). WAIS-IV reliable digit span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteran sample undergoing evaluation for mTBI. *The Clinical Neuropsychologist*, *27*(8), 1362–1372. https://doi.org/10.1080/13854046.2013.845248

Suchy, Y., Chelune, G., Franchow, E. I., & Thorgusen, S. R. (2012). Confronting patients about insufficient effort: The impact on subsequent symptom validity and memory performance. *The Clinical Neuropsychologist*, *26*(8), 1296–1311. https://doi.org/10.1080/13854046.2012.722230

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference, P. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Tan, Y. W., Burgess, G. H., & Green, R. J. (2021). The effects of acculturation on neuropsychological test performance: A systematic literature review. *The Clinical Neuropsychologist*, *35*(3), 541–571. https://doi.org/10.1080/13854046.2020.1714740

Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Mindt, M. R., Arentoft, A., & Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and Examiner race on neuropsychological performance: Simple as black and White?. *Journal of the International Neuropsychological Society: JINS*, *19*(5), 583–593. https://doi.org/10.1017/S1355617713000076

Tombaugh, T. N. (1996). *Test of memory malingering*. Multi-Health Systems.

U.S. Census Bureau. (2021). Racial and ethnic diversity in the united states: 2010 and 2020 census. https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html

VanLandingham, H., Ellison, R. L., Laique, A., Cladek, A., Khan, H., Gonzalez, C., & Dunn, M. R. (2022). A scoping review of stereotype threat for BIPOC: Cognitive effects and intervention strategies for the field of neuropsychology. *The Clinical Neuropsychologist*, *36*(2), 503–522. https://doi.org/10.1080/13854046.2021.1947388

Walton, G. M., & Spencer, S. J. (2009). Latent Ability. *Psychological Science*, *20*(9), 1132–1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV digit span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *The Clinical Neuropsychologist*, *32*(4), 657–670. https://doi.org/10.1080/13854046.2017.1415374

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). The Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). Pearson Clinical Assessment.

Whaley, A. L. (2021). Stereotype threat and neuropsychological test performance in the U.S. African American Population. *Archives of Clinical Neuropsychology*, *36*(7), 1361–1366. https://doi.org/10.1093/arclin/acab020

Whitney, K. A., Davis, J. J., Shepard, P. H., Bertram, D. M., & Adams, K. M. (2009). Digit span age scaled score in middle-aged military veterans: Is it more closely associated with TOMM failure than reliable digit span? *Archives of Clinical Neuropsychology*, *24*(3), 263–272. https://doi.org/10.1093/arclin/acp034

Wong, T. M., Strickland, T. L., Fletcher-Jansen, E., Ardila, A., & Reynolds, C. R. (2000). Theoretical and practical issues in the neuropsychological assessment and treatment of culturally dissimilar patients. In E. Fletcher-Janzen, T. L. Strickland, & C. R. Reynolds (Eds.), *Critical issues in neuropsychology. Handbook of cross-cultural neuropsychology* (pp. 3–18). Kluwer Academic Publishers.

Young, J. C., Roper, B. L., & Arentsen, T. J. (2016). Validity testing and neuropsychology practice in the VA healthcare system: Results from recent practitioner survey. *The Clinical Neuropsychologist*, *30*(4), 497–514. https://doi.org/10.1080/13854046.2016.1159730

Zuccato, B. G., Tyson, B. T., & Erdodi, L. A. (2018). Early bird fails the PVT? The effects of timing artifacts on performance validity tests. *Psychological Assessment*, *30*(11), 1491–1498. https://doi.org/10.1037/pas0000596