Performance on the Green Word Memory Test Following Operation Enduring Freedom/Operation Iraqi Freedom-Era Military Service: Test Failure is Related to Evaluation Context

Cortney L. McCormick1,2,*, Ruth E. Yoash-Gantz1,2,3, Scott D. McDonald1,4,5,6, Thomas C. Campbell1,4,5, Larry A. Tupler1,7,8

1Mid-Atlantic Mental Illness Research, Education, and Clinical Center, Durham, NC, USA
2Hefner Veterans Affairs Medical Center, Salisbury, NC, USA
3Department of Psychiatry and Behavioral Sciences, Wake Forest University, Winston-Salem, NC, USA
4McGuire Veterans Affairs Medical Center, Richmond, VA, USA
5Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA
6Department of Physical Medicine and Rehabilitation, Virginia Commonwealth University, Richmond, VA, USA
7Durham Veterans Affairs Medical Center, Durham, NC, USA
8Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

*Corresponding author at: James A. Haley VAMC, 13000 Bruce B. Downs Blvd., Tampa, FL 33612, USA. Tel.: +1-813-748-8100; fax: +1-813-631-3549.
E-mail address: cortney mccormick va.gov (C.L. McCormick).

Accepted 12 June 2013

Abstract
This study investigates prior reports of high neuropsychological symptom validity test (SVT) failure rates in post-deployed Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) active and veteran military personnel, using a large, multi-site sample (N = 214) drawn from three levels of the Department of Defense/Department of Veterans Affairs (VA) Polytrauma System of Care. The sample failure rate and its relationship to research versus dual research/clinical context of evaluation were examined, in addition to secondary variables explored in prior studies. Results yielded an overall failure rate of 25%, lower than prior reports describing OEF/OIF active-duty and veteran military personnel. Findings also supported the hypothesis that SVT failure rates would differ by context (dual > research). Participants with traumatic brain injury (TBI) failed more frequently than those without TBI in the dual context but not in the research context. Secondary analyses revealed that failure rates increased in the presence of depression, posttraumatic stress disorder, and male sex but were unrelated to active versus veteran military status, service connection (SC) or percentage of SC, age, education, or ethnicity. Further research is required to elucidate the underpinnings of these findings in light of the limited literature and variability between OEF/OIF-related SVT studies, as well as the substantial diagnostic and treatment implications for VA.

Keywords: Symptom validity tests; Effort; Neuropsychology; Military veterans; Traumatic brain injury; Posttraumatic stress disorder; Major depressive disorder; Service connection

Introduction
Given the elevated rates of traumatic brain injury (TBI) observed in the Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) cohort of over two million U.S. troops, the conclusion of military operations in Afghanistan (incipient) and Iraq (newly completed) will likely increase the demand for neuropsychological assessment (Jones, Ingram, & Ben-Porath, 2012). The rate of reported concussion/mild TBI (mTBI) among returning military personnel has been reported to be as high as 23%, depending on the diagnostic method and timing of inquiry (Hoge et al., 2008; Levin et al., 2010; MacGregor et al., 2010; Polusny et al., 2011; Schneiderman, Braver, & Kang, 2008; Terrio et al., 2009). These individuals may undergo neuropsychological evaluation for fitness-for-duty determination, assessment of cognitive sequelae of mental-health conditions, documentation of impairment...
for purposes of Veterans Benefits Administration (VBA) Compensation and Pension (C&P) decisions, rehabilitation guidance, and treatment planning (Lange, Pancholi, Bhagwat, Anderson-Barnes, & French, 2012). As such, neuropsychological evaluation findings and recommendations may have far-reaching implications for the vocational, financial, educational, psychosocial, and legal status of this cohort (Armstead-Jehle, 2010).

Due to the cognitive, behavioral, psychological, and functional impairments resulting from combat and other service-related injuries, many returning troops will have ongoing diagnostic and care needs (Golding, Bass, Percy, & Goldberg, 2009). In response, the Department of Defense (DoD) and Department of Veterans Affairs (VA) established four levels of a Polytrauma System of Care (i.e., multiple organ-system injuries; Belanger, Uomoto, & Vanderploeg, 2009b; Brahm et al., 2009; Cifu, Cohen, Lew, Jaffee, & Sigfورد, 2010; Cote, Syam, Vogel, & Cowper, 2007; Friedemann-Sanchez, Sayer, & Pickett, 2008; Hampton, 2011; Lew, 2005; Lew et al., 2007; Mernoff & Correia, 2010; Sayer et al., 2008; Sayer, Nelson, & Nugent, 2011; Sigfورد, 2008; Stelmach, Frith, Van Koevering, Rinne, & Stelmack, 2009). Four Level I Polytrauma Rehabilitation Centers (PRCs) provide intensive, multidisciplinary inpatient rehabilitation, each with a step-down Polytrauma Transitional Rehabilitation Program (PTRP). Level II consists of 22 Polytrauma Network Sites (PNSs) which offer post-acute rehabilitation and case management. Level III is composed of 83 Polytrauma Support Clinic Teams (PSCTs) providing follow-up outpatient rehabilitation services and case management. Level IV comprises 48 Polytrauma Points of Contact for referral and care coordination. Additionally, free medical care and monthly compensation are provided to many service members to address ongoing disability secondary to their military service via C&P evaluations for service connection (SC) and disability-percentage ratings, routinely informed by neuropsychological input (Boyko, Koepsell, Gaziano, Horner, & Feussner, 2000).

Despite concerted efforts by the DoD and VA to address TBI, the diagnostic process established to assess deployment-related injury and to inform resource allocation has prompted some to express reservations concerning the publicizing of TBI as a “signature” injury of OEF/OIF (Golding et al., 2009; Hoge et al., 2008; Hoge, Goldberg, & Castro, 2009; Ivins, 2010; Wilk et al., 2010). Accurate diagnosis and treatment can be challenging given that many combat-incurred TBIs, particularly mTBIs, go undocumented in the field due to mission priorities and are often only later diagnosed from self-report. False-positive diagnosis in particular may be encouraged due to time elapsed since injury: for example, studies of OEF/OIF veterans have observed an increase over time in rates of self-reported mTBI history (Polusny et al., 2011) as well as an increase in complaints of memory impairment rather than the expected decrease (Lange et al., 2012). Other problems include lack of specificity of diagnostic criteria (such as the subjective “alteration of consciousness” after the event) and/or potential misattribution of psychological factors or post-deployment maladjustment to TBI events (Bryant, 2008; Hoge et al., 2008; Otis, McGlinchey, Vasterling, & Kerns, 2011; Polusny et al., 2011; Russo, 2012; Terrio, Nelson, Betthauser, Harwood, & Brenner, 2011). VA’s emphasis on addressing “occult” (Lew, Poole, Alvarez, & Moore, 2005, p. 393) or “invisible” (Tanielian & Jaycox, 2008, p. xix) wounds in this population, often based on the non-specific Neurobehavioral Symptom Inventory (Benge, Pastorek, & Thornton, 2009; Cicerone & Kalmar, 1995; Hill, Mobo, & Cullen, 2009; Lippa, Pastorek, Benge, & Thornton, 2010), can promote over- attribution of neurologic contributions to patient presentation, which, combined with negative expectations for recovery, can foster iatrogenic illness (Hoge et al., 2008; Polusny et al., 2011; Russo, 2012; Sayer et al., 2011). Compounding the issue, if OEF/OIF veterans with putative TBI demonstrate inconsistent responses or inaccuracies of self-reported injury, additional problems of reliability and validity arise (Van Dyke, Axelrod, & Schutte, 2010). False-positive diagnosis of TBI can be especially promulgated when subjective cognitive complaints are accompanied by poor neuropsychological performance, which may be secondary to poor effort applied to testing. Improved discrimination is therefore necessary of signs indicative of veridical self-presentation reflecting neurological contributions of TBI rather than fictitious, psychogenic, somatoform, or malingered findings in this population.

With regard to the specific isolation of neurocognitive findings secondary to TBI/mTBI, clinicians and researchers agree that symptom validity tests (SVTs) represent a cornerstone of a comprehensive neuropsychological evaluation, particularly for populations in which secondary gain (monetary, legal, psychological) may influence test performance (Heilbronner et al., 2010; Jones et al., 2012; Lange et al., 2012). Currently, over 300 publications address symptom validity in civilian forensic populations (Green, 2007), with a wide range of failure rates depending on population characteristics, evaluation context, and influence of secondary gain. Examples include SVT failure between 8% and 39% for medical and mTBI cases, respectively (Larrabee, 2003; Mittenberg et al., 2002), 18% failure rate among memory-disorder clinic patients with financial incentive to appear impaired (Howe, Anderson, Kaufman, Sachs, & Loring, 2007), and observed rates as high as 54% in criminal forensic settings (Ardolf, Denney, & Houston, 2007). Due to unique military-related characteristics of the OEF/OIF cohort such as deployment into combat, the equivocal neuropsychological implications of blast injury, and repeated exposure to other extreme stressors, the civilian body of research may not generalize to this population.

Although neurocognitive performance following TBI has been investigated in the OEF/OIF cohort, few studies report observations on symptom validity per se, potentially limiting inferences regarding its role in neuropsychological presentations: SVTs are either not employed (Luethcke, Bryan, Morrow, & Isler, 2011) or used to exclude cases with inadequate performance to avoid potential contamination of the data (Belanger, Kretzmer, Yoash-Gantz, Pickett, & Tupler, 2009a; Brenner et al., 2010; Ivins,
Kane, & Schwah, 2009; Marx et al., 2009; Nelson et al., 2012; Ruff, Huff, & Wang, 2009; Vasterling et al., 2006, 2012). A limited number of recent studies of OEF/OIF military personnel with TBI history, however, have focused specifically on SVT performance characteristics (Armistead-Jehle, 2010; Nelson et al., 2010; Russo, 2012; Whitney, Shepard, Williams, Davis, & Adams, 2009) with varying methods, results, and conclusions.

Whitney and colleagues (2009) retrospectively examined consecutive records of 23 OEF/OIF post-deployed clinical-neuropsychological referrals (all but one self-reporting exposure to explosive blasts) including both veterans (n = 19, 61%) and active-duty military (n = 9, 39%) at a VA Polytrauma System of Care PNS (outpatient clinic). A 17% failure rate (n = 4) was observed on the Medical Symptom Validity Test (MSVT; Green, 2004), a forced-choice measure designed to identify individuals with cognitive performance incommensurate with clinical circumstances. Of the four individuals failing, all were still active military, and all four failed at least one other SVT. The findings led the authors to conclude that these individuals satisfied diagnostic criteria for probable malingering neurocognitive dysfunction (Slick, Sherman, & Iverson, 1999). Without a more comprehensive understanding of the origin of 100% MSVT specificity among the active-duty personnel in the Whitney and colleagues (2009) study, this inference based on only four participants may be premature.

In a similar study also employing the MSVT, Armistead-Jehle (2010) examined a retrospective clinical sample of 45 consecutive OEF/OIF veterans with a self-reported history of TBI (over 90% from blasts) and observed that 58% (n = 26) failed. Further analysis yielded no differences between pass and fail rates on the MSVT in relation to age, sex, education, mean percentage of SC, or frequency of posttraumatic stress disorder (PTSD) or substance-use disorder. However, of the 26 patients failing the MSVT, 69% had a self-reported history of depression and 96% had SC. The author concluded that depressed individuals with SVT failure may be “prone to symptom exaggeration” (Armistead-Jehle, 2010, p. 57) and thus exert a broader influence on SVT performance. Armistead-Jehle (2010) also noted that the “potential for financial gain in the VHA [Veterans Health Administration] system is omnipresent” (p. 57) and that “external incentive to appear more compromised than one might objectively be is potentially ubiquitous in this system” (p. 58). These observations and conclusions speak to the importance of further investigation, particularly the association between SVT performance and both self-reported psychopathology and external influences on behavioral presentation, such as SC.

Russo (2012) examined a retrospective VA sample of 38 consecutive OEF/OIF combat veterans receiving PSCT Polytrauma System of Care consult requests following diagnoses of definite TBI, using the Green Word Memory Test (WMT; Green, 2005). The WMT is a forced-choice cognitive measure upon which the MSVT is based. A failure rate of 68.4% was observed. Examination of demographic and SC variables yielded no significant differences between pass/fail rates for sex, education, ethnicity, percent with SC, or mean percentage of SC. However, participants who failed the WMT were significantly older than those who passed.

Nelson and colleagues (2010) improved upon these studies by using a larger sample (N = 119); non-TBI, research-only, and non-OEF/OIF-combat-era controls; and a fixed battery of SVT measures, including the Victoria Symptom Validity Test (VSST; Slick, Hopp, Strauss, & Thompson, 1997). Results showed that in veterans largely (n = 82) with concussion (71% from blasts), SVT failure was less frequent in the research OEF/OIF group (10.7%, n = 75) than the forensic (C&P) group (59.1%, n = 44), irrespective of service era. Of the total sample restricted to OEF/OIF (research and forensic), 22 of 99 participants (22%) had invalid SVT results, of which 16 of 24 participants (67%) were in the forensic group and 6 of 69 participants (8%) were in the research group. These findings suggest that invalid SVT performance occurred in both forensic and research contexts, although multiple times more frequently in the former context involving VBA disability claimants.

These four studies of SVT performance in participants deployed to OEF/OIF thus yield whole-sample failure rates from 17% to 68%, the upper limit of which is cautionary, if not alarming. In fact, the SVT failure rates of 58% (Armistead-Jehle, 2010), 67% (Nelson et al., 2010), and 68% (Russo, 2012) in clinical samples of OEF/OIF veterans with TBI are markedly higher than the 30%–40% typically reported among civilian settings (Larrabee, 2003; Mittenberg et al., 2002) and more consistent with the 54% found in the criminal setting (Ardolf et al., 2007). The findings, however, may be influenced by small sample sizes, differing methods of data procurement, or retrospective analyses in clinical samples. In view of these limitations and the paucity of relevant research on this cohort, combined with the service needs and potential for iatrogenic consequences previously articulated, the present study sought to further characterize the base rates and underpinnings of SVT failure in the OEF/OIF-era military-service population.

Scope of the Present Study

We address the findings of previous research, particularly the wide range of SVT-failure results in the OEF/OIF cohort, the notably high failure rates observed in association with TBI and the strong effect of evaluation context, using a large, multi-site sample drawn from the three highest levels of the DoD/VA Polytrauma System of Care. Symptom validity is defined by performance on the Green WMT (Green, 2005). The WMT is considered psychometrically more robust and sensitive than the MSVT,
with comparative failure rates in a forensic (VBA disability compensation) sample \( (N = 279) \) of 37% and 29%, respectively, and an overall agreement (pass/fail) rate of 80% (Boone, 2007; Green, 2007).

We propose three primary hypotheses. (1) Due to the use of a larger, prospectively enrolled, more diverse OEF/OIF-era sample, overall SVT rates will be lower than the 58%–68% rate reported by Armistead-Jehle (2010), Nelson and colleagues (2010), and Russo (2012) and more comparable with civilian non-criminal forensic rates. (2) Replication will be observed for Nelson and colleagues’ (2010) finding that SVT performance differs by context of evaluation, with a lower rate of SVT failure observed in participants serving solely in a research protocol than those providing dual consent for both research and clinical uses of their data. As a corollary of this hypothesis, it is expected that participants with TBI tested in the dual context will show a higher SVT failure rate than participants with TBI tested in the research context. (3) We hypothesize that within the research context, participants with TBI will show a higher failure rate than participants without TBI.

We secondarily examine relationships between variables previously explored in the OEF/OIF literature by Whitney and colleagues (2009), Armistead-Jehle (2010), and Russo (2012): age, sex, education, ethnicity, active versus veteran military status, presence versus absence of SC and SC percentage if awarded, and presence versus absence of probable depression and PTSD.

**Methods**

**Participants**

Participants consisted of 214 OEF/OIF-era military personnel participating between June 2006 and September 2010 in a multisite protocol conducted by the Neurocognition Laboratory of the Mid-Atlantic Mental Illness Research, Education, and Clinical Center (MA-MIRECC). Institutional Review Board approval was obtained at each individual site. All participants provided written informed consent and were remunerated.

The sample was drawn from institutions within the first three levels of the DoD/VA Polytrauma System of Care and comprised both Active military and Veteran research volunteers, with most of the Veteran participants enrolled in VA outpatient medical care. Inpatient participants \( (n = 44) \) were recruited from a Level I PRC/PTRP in the McGuire VA Medical Center (VAMC) in Richmond, Virginia, where they were receiving either inpatient or residential treatment (i.e., polytrauma; Belanger, Scott, Scholten, Curtiss, & Vanderploeg, 2005; Lew, 2005; Lew et al., 2007; Sayer et al., 2008). Non-Inpatient (VA outpatient and non-VA-affiliated research-volunteer) participants \( (n = 170) \) were recruited from the Level II McGuire VAMC PNS and Level III PSCTs at the Hefner VAMC in Salisbury and the Durham VAMC in North Carolina. The latter participants were acquired from the MA-MIRECC Recruitment Database, which is a research registry open to all personnel serving in the U.S. Armed Forces at any time post 11 September 2001 and includes an optional authorization for re-contact for additional research participation after the index assessment visit (see Dedert et al., 2009). MA-MIRECC Recruitment Database participants were re-contacted and offered enrollment in a study involving several hours of neurocognitive and personality testing to advance understanding of post-deployment mental health and for which they would be remunerated. Excluded from re-contact efforts were individuals with evidence of psychosis, current substance abuse or dependence, or exclusively non-military-related TBI, as well as those who declined authorization for re-contact. Participants were also oversampled to ensure adequate cell sizes with TBI and PTSD. Specifically, we examined the MA-MIRECC Recruitment Database so as to emphasize acquisition of participants with these primary diagnostic classifications of interest. Data on selective neuropsychological, depression, and PTSD measures on up to 54 of the 214 participants have been reported previously (Belanger et al., 2009a; Nelson, Yoash-Gantz, Pickett, & Campbell, 2009).

Context of evaluation was determined prior to neuropsychological testing. Participants consented for their neuropsychological data to either be used (a) specifically for Research purposes, consistent with the design of the Research Core of the MA-MIRECC, or (b) for Dual research and clinical purposes (released to a staff neuropsychologist to generate a physician-requested clinical report to inform patient care). Thus, the latter participants consented to both completion of a Polytrauma System of Care clinical consult and the pooling of their test findings with research data on post-deployment mental health. A total of 136 participants (64%) were consented under a Research context; 78 participants (36%) provided Dual consent for clinical and research applications of their data.

**Procedures and Measures**

For Inpatient participants, medical records were reviewed for demographic data on age, sex, education, ethnicity, and military-service history. TBI classification was also obtained from medical records, including date and description of the trauma event (e.g., blast, vehicular crash); duration of loss of consciousness, duration of posttraumatic amnesia, and/or Glasgow Coma Scale (GCS) score; and neuroradiology reports. For Non-Inpatient participants, the MA-MIRECC Recruitment Database provided...
demographic data on age, sex, education, ethnicity, SC status and percentage, and military-service history, all reacquired at the time of neurocognitive evaluation. TBI history was obtained from a MA-MIRECC Recruitment Database self-report questionnaire, adapted from a Defense and Veterans Brain Injury Center (DVBI; Salazar, Zitnay, Warden, & Schwab, 2000) screening instrument. Self-report of date, trauma-event description, level and duration of alteration of consciousness, and presence and duration of anterograde amnesia were obtained for up to six injury events (Ivins et al., 2003). For cases in which the subject indicated an inability to recall whether or not he/she had suffered a head injury \( (n = 19, 9\%) \), further determination proceeded through clinician-administered interview (McCormick TBI Interview; McCormick, 2009) and/or medical-record review.

All participants received a standardized, fixed battery of neurocognitive, personality, psychopathology, and symptom-validity measures administered in accordance with their respective manuals. The WMT was presented on Dell Latitude D410 laptop computers identically issued across the three VAMC sites. The psychometrician (possessing at least graduate-level psychology training) was seated in the exam room distant from the subject and outwardly oblivious to the computer screen, quietly scoring test forms acquired earlier in session. The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) and PTSD Checklist-Military version (PCL-M; Keen, Kutter, Nikeles, & Krinsley, 2008) self-report instruments were administered in paper-and-pencil format.

**Group Assignments and Statistical Analyses**

Scoring criteria for SVT outcome provided in the Green (2005) WMT manual were applied, using the established cutoffs for pass, fail, and caution. Assignment to a Probable Depression group was defined by a BDI-II total score cutoff of 19 or greater, which has been shown to provide optimal sensitivity (87%) and specificity (79%) in the TBI population for a diagnosis of major depressive disorder using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-IV; First, Spitzer, Gibbon, & Williams, 1995; Homaifar et al., 2009). Assignment to a Probable PTSD group was defined by a PCL-M total score of 50 or greater, the predominant cutoff score in the literature (Riviere, Kendall-Robbins, McGurk, Castro, & Hoge, 2011). Although we are unaware of studies of PCL-M diagnostic accuracy conducted with OEF/OIF veterans, Yeager, Magruder, Knapp, Nicholas, and Frueh (2007) reported a sensitivity of 0.53 and specificity of 0.95 for a large multi-service-era sample \( (N = 840) \) of veterans receiving VA primary care. Estimating a 19% base rate of PTSD among OEF/OIF veterans (Tanielian & Jaycox, 2008), approximately 87% of participants will be correctly classified using this method (71% of those screening positive will truly have PTSD, and 90% of those screening negative for PTSD will not have PTSD).

Analyses were performed using SAS version 9.1.3 (SAS Institute Inc., Cary, NC, USA). Chi-square or Fisher’s exact tests were used to examine independence between WMT classification and discrete variables: evaluation context, Inpatient versus Non-Inpatient status, Polytrauma System of Care level, VAMC site, sex, ethnicity (Caucasian vs. minority), military status (Active vs. Veteran), and presence/absence of SC, Probable Depression, and Probable PTSD. Differences between proportions failing the WMT were examined using z-scores. Analyses of variance (ANOAs) were used to examine continuous dependent measures (age, education, and SC-disability percentage), with the WMT outcome (pass, fail, caution) and context (Dual, Research) as class variables and examination of the WMT outcome \( \times \) context interaction.

**Results**

**Demographic, Military/VA-System, and Clinical Characteristics of the Total Sample**

Demographic and military/VA-system variables for the total sample \( (N = 214) \) are presented in the upper portion of Table 1. The mean \( (SD) \) age and education were 34.0 (9.7) and 13.7 (1.8) years, respectively. The sample was 93% men, 64% Caucasian, and 52% still Active in the military, either in Active Duty, National Guard, or Ready Reserve components. The modal service branch in which participants served was the Army, followed in frequency by the Marine Corps, Navy, and Air Force. Forty-four percent of the sample had SC; participants with SC averaged a 43% disability rating. In the context of SC awarded predominately for musculoskeletal conditions, 35 participants (37%) had SC for PTSD, 7 participants (7%) for another anxiety diagnosis, 7 participants (7%) for TBI, 3 participants (3%) for major depression, and 1 participant (1%) for bipolar disorder. Ten participants (11%) had SC for two of these psychiatric disorders; one participant (1%) had SC for three.

The lower portion of Table 1 provides the clinical variables of number positive and negative for TBI history and severity of most recent TBI based on chart review for Inpatients (duration of LOC and/or GCS score) and the MA-MIRECC-adapted DVBI TBI screen and McCormick Interview for Non-Inpatients. Fifty-seven percent of the sample was positive for TBI history, with 57% of injuries attributed to blasts. Most recent TBI was mild in 83% of participants, moderate in 7%, and severe in 7%. Screening-cutoff outcomes for diagnostic variables yielded 36% with Probable Depression and 42% with Probable PTSD.
Main Hypotheses.

Examination of performance on the Green WMT supported hypothesis 1, with an overall SVT failure rate of 25% in the total sample. Pass and caution rates were 57% and 17%, respectively. No differences in SVT performance were observed for Inpatients versus Non-Inpatients, \( \chi^2(2, N = 214) = 3.29, \Phi = 0.21, p = .19 \), Polytrauma System of Care level, \( \chi^2(4, N = 214) = 7.35, \Phi = 0.22, p = .12 \), or VAMC site, \( \chi^2(4, N = 214) = 6.39, \Phi = 0.20, p = .17 \). Subsequent analyses were therefore considered justified combining across these subsamples.

Table 1. Demographic, military/VA-system, and clinical variables for the OEF/OIF-era sample (\( N = 214 \))

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years; mean ± SD)</td>
<td>34.0 ± 9.7</td>
</tr>
<tr>
<td>Sex (men/women)</td>
<td>199/15</td>
</tr>
<tr>
<td>Education (years; mean ± SD)</td>
<td>13.7 ± 1.8</td>
</tr>
<tr>
<td>Ethnicity (Caucasian/minority)</td>
<td>136/78</td>
</tr>
<tr>
<td><strong>Military/VA-system status</strong></td>
<td></td>
</tr>
<tr>
<td>Military status (Active/Veteran)</td>
<td>112/102</td>
</tr>
<tr>
<td><strong>Most recent service branch</strong></td>
<td></td>
</tr>
<tr>
<td>Air Force</td>
<td>9</td>
</tr>
<tr>
<td>Active Component</td>
<td>7</td>
</tr>
<tr>
<td>National Guard</td>
<td>2</td>
</tr>
<tr>
<td>Army</td>
<td>153</td>
</tr>
<tr>
<td>Active Component</td>
<td>79</td>
</tr>
<tr>
<td>National Guard</td>
<td>54</td>
</tr>
<tr>
<td>Ready Reserves</td>
<td>20</td>
</tr>
<tr>
<td>Marine Corps</td>
<td>35</td>
</tr>
<tr>
<td>Active Component</td>
<td>34</td>
</tr>
<tr>
<td>Ready Reserves</td>
<td>1</td>
</tr>
<tr>
<td>Navy</td>
<td>17</td>
</tr>
<tr>
<td>Active Component</td>
<td>13</td>
</tr>
<tr>
<td>Ready Reserves</td>
<td>4</td>
</tr>
<tr>
<td><strong>Deployment</strong></td>
<td></td>
</tr>
<tr>
<td>Afghanistan</td>
<td>11</td>
</tr>
<tr>
<td>Afghanistan/Iraq</td>
<td>10</td>
</tr>
<tr>
<td>Iraq</td>
<td>109</td>
</tr>
<tr>
<td>Iraq region</td>
<td>1</td>
</tr>
<tr>
<td>Iraq/Kuwait</td>
<td>68</td>
</tr>
<tr>
<td>Kuwait</td>
<td>3</td>
</tr>
<tr>
<td>Other Middle Eastern country</td>
<td>2</td>
</tr>
<tr>
<td>None (Continental U.S.)</td>
<td>10</td>
</tr>
<tr>
<td>Service connected (+/−)</td>
<td>95/119</td>
</tr>
<tr>
<td>Service connection (%; mean ± SD)</td>
<td>42.6 ± 28.6</td>
</tr>
<tr>
<td><strong>Clinical status</strong></td>
<td></td>
</tr>
<tr>
<td>TBI (+/−)</td>
<td>122/92</td>
</tr>
<tr>
<td>TBI severity</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>101</td>
</tr>
<tr>
<td>Dazed/confused</td>
<td>52</td>
</tr>
<tr>
<td>LOC 1–20 min</td>
<td>46</td>
</tr>
<tr>
<td>GCS score 13–15</td>
<td>3</td>
</tr>
<tr>
<td>Moderate</td>
<td>8</td>
</tr>
<tr>
<td>LOC 21–59 min</td>
<td>4</td>
</tr>
<tr>
<td>GCS score 9–12</td>
<td>4</td>
</tr>
<tr>
<td>Severe</td>
<td>9</td>
</tr>
<tr>
<td>LOC ≥60 min</td>
<td>8</td>
</tr>
<tr>
<td>GCS score &lt;9</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
</tr>
<tr>
<td>Blast/other TBI event</td>
<td>69/53</td>
</tr>
<tr>
<td>Probable Depression (+/−)</td>
<td>77/133</td>
</tr>
<tr>
<td>Probable PTSD (+/−)</td>
<td>89/125</td>
</tr>
</tbody>
</table>

Notes: GCS = Glasgow Coma Scale; LOC = loss of consciousness; PTSD = posttraumatic stress disorder; TBI = traumatic brain injury; VA = Department of Veterans Affairs.

**Green WMT Classification Outcomes**

Main Hypotheses. Examination of performance on the Green WMT supported hypothesis 1, with an overall SVT failure rate of 25% in the total sample. Pass and caution rates were 57% and 17%, respectively. No differences in SVT performance were observed for Inpatients versus Non-Inpatients, \( \chi^2(2, N = 214) = 3.29, \Phi = 0.21, p = .19 \), Polytrauma System of Care level, \( \chi^2(4, N = 214) = 7.35, \Phi = 0.22, p = .12 \), or VAMC site, \( \chi^2(4, N = 214) = 6.39, \Phi = 0.20, p = .17 \). Subsequent analyses were therefore considered justified combining across these subsamples.
Green WMT results for Dual and Research groups and subclassified by TBI status are presented in Table 2. Confirming hypothesis 2, a significant association was observed for context of the evaluation, with participants in the Dual context almost equally likely to pass or fail, whereas those in the Research group were over four times more likely to pass than fail ($p < .0001$). Specifically, of participants tested in the Dual context, 40% passed, 42% failed, and 18% were classified as caution, whereas in the Research context, 68% passed, 15% failed, and 17% were classified as caution. The difference between failure rates was also highly significant ($p < .0001$). Examination of the corollary of hypothesis 2 predicting worse performance for Dual than Research TBI participants fell short of significance for the test of independence ($p = .08$) but was significant comparing failure rates ($p = .03$). Support was not observed for hypothesis 3: participants in the Research context with TBI were statistically independent from those without TBI, and the failure rates were not significantly different between groups.

Secondary Examination of Associations with WMT Failure. Secondary analyses of additional variables potentially associated with WMT failure are presented in Table 3. In view of the previous findings establishing differences in the failure rate as a function of evaluation context but non-significant differences in the Research context between those with and without TBI, results are provided for the entire sample and subdivided by Dual and Research contexts for all variables except sex due to an insufficient number of females. Overall, demographic, military/VA-system, and clinical contributions to SVT failure yielded mixed support for findings previously reported in the literature for this cohort.

Examination of demographic variables yielded no differences in WMT performance in relation to age, education, or ethnicity. However, we did observe an association with sex ($p = .001$): all females passed the WMT, whereas only slightly more than half (54%) of males passed the WMT, 27% failed, and 19% were classified as caution, resulting in a significant difference in the failure rate ($p = .02$). For participants still Active in the military, 54% passed, 27% failed, and 19% were classified as caution, while for Veterans, 61% passed, 24% failed, and 16% were classified as caution. Differences between Active and Veteran participants were not significant. For participants with SC, 56% passed, 26% failed, and 18% were classified as caution, while in participants without SC, 59% passed, 24% failed, and 17% were classified as caution. Differences with respect to SC status were not significant. Likewise, for those with SC, no differences were observed in the mean percentage of disability awarded between WMT-classification outcomes. For Probable Depression, 42% passed, 34% failed, and 25% were classified as caution, whereas in the participants without depression, 67% passed, 20% failed, and 14% were classified as caution. These results yielded a significant association with depression ($p = .002$), with depressed participants significantly more likely to fail ($p = .02$). For Probable PTSD, 44% passed, 34% failed, and 22% were classified as caution, whereas in participants without PTSD, 67% passed, 19% failed, and 14% were classified as caution. These results yielded a significant association with PTSD status ($p = .003$), with Probable PTSD participants significantly more likely to fail ($p = .02$).

Examination of secondary contributions to SVT failure separately by context (Table 3) revealed that those variables non-significant for the entire sample remained non-significant under both Dual and Research contexts, and ANOVA models examining the WMT outcome $\times$ context interactions were non-significant for the three continuous variables of age, education, and SC percentage. In addition, failure rates for both values of all dichotomous variables were uniformly higher under the Dual context than the Research context at the $p = .02$ level or below (data not shown), except for the analysis examining participants without PTSD described below. For depression status, an association narrowly failed to be observed under the Dual context, $\chi^2(2, N = 76) = 5.79, \Phi = 0.28, p = .06$, but a high level of significance was noted under the Research context, $\chi^2(2, N = 134) = 10.03, \Phi = 0.27, p = .007$. Whereas in the Dual context, the difference in the failure rate between those with (58%) and without (31%) Probable TBI versus TBI-negative participants ($\chi^2(2, N = 134) = 4.94, \Phi = 0.20, p = .08; z = 3.01, p = .03.$

Table 2. Frequencies (percentages) of pass, caution, and fail outcomes on the Green WMT and statistical tests comparing the Dual versus Research context (*), Dual versus TBI-positive participants (+TBI) in the Research context ($\dagger$), and +TBI versus TBI-negative participants (-TBI) in the Research context ($\ddagger$)

<table>
<thead>
<tr>
<th>WMT outcome</th>
<th>Pass</th>
<th>Caution</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>OEF/OIF-era sample</td>
<td>123 (57%)</td>
<td>37 (17%)</td>
<td>54 (25%)</td>
</tr>
<tr>
<td>+TBI combined</td>
<td>56 (46%)</td>
<td>23 (19%)</td>
<td>43 (35%)</td>
</tr>
<tr>
<td>+TBI Dual*</td>
<td>31 (40%)</td>
<td>14 (18%)</td>
<td>33 (42%)</td>
</tr>
<tr>
<td>+TBI Research$\dagger$</td>
<td>25 (57%)</td>
<td>9 (20%)</td>
<td>10 (23%)</td>
</tr>
<tr>
<td>−TBI Research$\dagger$</td>
<td>67 (73%)</td>
<td>14 (15%)</td>
<td>11 (12%)</td>
</tr>
<tr>
<td>Research combined*</td>
<td>92 (68%)</td>
<td>23 (17%)</td>
<td>21 (15%)</td>
</tr>
</tbody>
</table>

Notes: TBI = traumatic brain injury; WMT = Word Memory Test; OEF/OIF = Operation Enduring Freedom/Operation Iraqi Freedom.

$\ddagger$ $\chi^2(2, N = 134) = 3.85, \phi = 0.17, p = .15; z = 1.63, p = .10.$

Downloaded from https://academic.oup.com/acn/article-abstract/28/8/808/5044/Performance-on-the-Green-Word-Memory-Test by guest on 03 October 2017
<table>
<thead>
<tr>
<th>WMT outcome</th>
<th>Pass</th>
<th>Caution</th>
<th>Fail</th>
<th>z-value, p-value</th>
<th>Statistical test, effect size, p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years; mean ± SD)</td>
<td>34.2 ± 10.0</td>
<td>33.8 ± 10.5</td>
<td>33.8 ± 8.8</td>
<td></td>
<td>F(2,211) = 0.05, η² = 0.00, p = .95</td>
</tr>
<tr>
<td>Dual</td>
<td>27.2 ± 5.9</td>
<td>28.1 ± 11.1</td>
<td>31.8 ± 7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>36.6 ± 9.9</td>
<td>37.3 ± 8.6</td>
<td>36.9 ± 9.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (men/women)</td>
<td>108/15</td>
<td>37/0</td>
<td>54/0</td>
<td>2.33, .02</td>
<td>Fisher’s exact p = .001</td>
</tr>
<tr>
<td>n</td>
<td>54/100</td>
<td>19/0</td>
<td>27/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (years; mean ± SD)</td>
<td>13.9 ± 1.9</td>
<td>13.5 ± 1.5</td>
<td>13.4 ± 1.8</td>
<td></td>
<td>F(2,211) = 1.93, η² = 0.02, p = .15</td>
</tr>
<tr>
<td>Dual</td>
<td>12.9 ± 1.2</td>
<td>13.0 ± 1.1</td>
<td>13.5 ± 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>14.3 ± 1.9</td>
<td>13.7 ± 1.6</td>
<td>13.3 ± 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity (Caucasian/minority)</td>
<td>82/41</td>
<td>23/14</td>
<td>31/23</td>
<td>-1.08, .28</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>60/53</td>
<td>17/18</td>
<td>23/29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military status (Active/Veteran)</td>
<td>61/62</td>
<td>21/16</td>
<td>30/24</td>
<td>0.55, .58</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>54/61</td>
<td>19/16</td>
<td>27/24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC (+/-)</td>
<td>53/70</td>
<td>17/20</td>
<td>25/29</td>
<td>0.33, .74</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>56/59</td>
<td>18/17</td>
<td>26/24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression (+/-)</td>
<td>32/89</td>
<td>19/18</td>
<td>26/26</td>
<td>2.30, .02</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>42/67</td>
<td>25/14</td>
<td>34/20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD (+/-)</td>
<td>39/84</td>
<td>20/17</td>
<td>30/24</td>
<td>2.41, .02</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>44/67</td>
<td>22/14</td>
<td>34/19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: WMT = Word Memory Test; SC = service connection; PTSD = posttraumatic stress disorder. Data are provided for the total sample and separately for Dual and Research contexts, except for sex.
Depression was significant ($z = 2.34, p = .02$), in the Research context those with Probable Depression (17%) were not significantly different from those without the disorder (14%; $z = 0.58, p = .56$). For PTSD, a high level of significance was observed for both Dual, $\chi^2(2, N = 78) = 10.69$, $\Phi = 0.37$, $p = .005$, and Research, $\chi^2(2, N = 136) = 9.80$, $\Phi = 0.27$, $p = .007$, contexts. However, whereas in the Dual context, failure rates were significantly higher for those with Probable PTSD (64%) than without PTSD (27%; $z = 3.27, p = .001$), in the Research context, those with Probable PTSD (16%) were not significantly different from those without the disorder (15%; $z = 0.17, p = .86$). Furthermore, although the failure rates were significantly different between Dual (64%) and Research (16%) contexts for participants with Probable PTSD ($z = 4.59, p < .0001$), the failure rates were not significantly different between Dual (27%) and Research (15%) contexts for those without PTSD ($z = 1.59, p = .11$).

**Discussion**

The present study sought to advance understanding of symptom validity in the OEF/OIF-era cohort by exploring observed SVT failure rates and the relevance of evaluation context. We addressed limitations of previous studies (e.g., small clinical samples drawn retrospectively), describing SVT performance in OEF/OIF veterans and active military personnel (Armistead-Jehle, 2010; Nelson et al., 2010; Russo, 2012; Whitney et al., 2009) using a large ($N = 214$), predominantly post-deployed sample recruited into a well controlled, prospectively designed study conducted across three levels of the DoD/VA Polytrauma System of Care and at three VAMC sites using identical experimental protocols. All evaluations consisted of a fixed neurocognitive battery employing the Green WMT as the operational indicator of level of effort.

As expected, our data yielded a rate of WMT failure for the total sample on the low end of the range reported in the current literature. The Whitney and colleagues (2009), Armistead-Jehle (2010), Russo (2012), and Nelson and colleagues (2010) studies reported sample SVT failure rates ranging from 17% to 68%. In the present study, we observed a failure rate of 25%. This finding supported our primary hypothesis that SVT failure in our sample would be consistent overall with civilian non-criminal forensic rates (Larrabee, 2003; Mittenberg et al., 2002). Our observations are also consistent with the unremarkable OEF/OIF-cohort findings of Nelson and colleagues (2010) showing no difference in WMT performance between OEF/OIF deployed and veterans drawn from an earlier military era. Overall, our findings are closest to the 22% figure reported for the entire sample by Nelson and colleagues (2010), which contained the greatest power of the four studies previously published and included participants without TBI. In the Dual context, which was composed of clinical cases satisfying Polytrauma System of Care consult requests, the observed failure rate for our study was 42%. Importantly, our overall findings are discernibly lower than the 54% criminal-forensic SVT failure rate (Ardolf et al., 2007) and particularly the 58%–68% failure rates reported by Armistead-Jehle (2010), Nelson and colleagues (2010), and Russo (2012) for OEF/OIF retrospective clinical samples with TBI. The contrast between findings is even more noteworthy when considered in light of the greater sensitivity of the WMT compared with the MSVT used by Whitney and colleagues (2009) and Armistead-Jehle (2010). Less suspicion of symptom invalidity therefore arises than might otherwise be applied to the larger OEF/OIF cohort from data published in prior literature.

Our results also support the Nelson and colleagues (2010) findings showing the context of evaluation to be significantly associated with SVT performance: individuals participating in neuropsychological evaluations identified as carrying potential clinical implications performed worse on the WMT than those whose evaluations were conducted solely under the auspices of research. For participants in the Research context, the failure rate was 15%. Nelson and colleagues’ (2010) OEF/OIF sample exhibited a failure rate of 67% for the forensic (concussion) group and 8% for the research group, approximately half of which had history of concussion. It is possible that the lower figure of 42% we obtained for the Dual group represents an intermediate value reflecting the summation of both clinical (increasing) and research (decreasing) contributions arising out of the combined contextual influences. It is also important to acknowledge that Nelson and colleagues (2010) used a different measure to assess symptom validity. Although not addressing OEF/OIF participation per se, Armistead-Jehle and Buican (2012) recently reported that active-duty personnel examined in the context of Medical Evaluation Board (MEB) disability assessments failed the WMT at a rate of 54%, whereas those examined from the same Army Medical Facility for non-MEB clinical purposes failed at a lower rate of 35%. Such results suggest that context is a matter not only of clinical versus research application of the findings but, in a purely clinical setting, the implications for future military status. Similarly, Armistead-Jehle and Hansen (2011) reported in their sample of 335 active-duty military drawn from an Army Medical Facility (OEF/OIF deployment unspecified) that whereas personnel with a modal rank of Sergeant (E-5) failed the MSVT at a rate of 30%, those enrolled in a training course required of field officers at the rank of Major (O-4) failed at a rate of only 8%. These findings suggest that in addition to the type of clinical versus research context, select subsamples or separate military cohorts may be identified manifesting differential associations with failing SVT performance (Armistead-Jehle & Hansen, 2011). Likewise, knowledge of the purposes of the assessment (e.g., neurocognitive vs. psychiatric) may influence whether cognitive versus psychological symptoms carry exaggeration (Ruocco et al., 2008).

Considered as a whole, these findings indicate that investigators cannot naively assume that neurocognitive findings directly capture neurologic compromise but instead must recognize that data may reflect feedback-related inputs to the subject. Such
potential inputs include whether test efforts may influence future clinical care or career status, hold potential for disability payments or other types of secondary gain, contribute to general research knowledge pertaining to other service personnel but confer no personal benefit, or yield only monetary remuneration in exchange for time in study participation. Our findings speak specifically to the rate of SVT failure: however, further research could beneficially explore potential contextual associations with neuropsychological test findings themselves in those with passing and, especially, caution WMT outcomes. Investigation is also warranted of the possible benefits to performance following from an enhanced motivational effect (e.g., as a research subject representing one’s military cohort) over other contexts and feedback-related inputs to the test participant.

Contrary to our third primary hypothesis, we did not observe in the Research context an effect for TBI compared with participants without TBI: TBI-positive participants failed at a rate of 23% compared with 12% for those without TBI, which was not statistically significant. These findings therefore suggest that prior reports implicating an increased likelihood of cognitive-symptom exaggeration in TBI (Mittenberg et al., 2002) may have instead been governed by the clinical context of testing. Other potential influences on SVT outcomes from concomitant aspects of TBI history cannot also be excluded, such as the presence of injury in addition to the head, hospitalization, residual pain, and comorbidities such as depression and PTSD.

In recognition of the latter consideration, we examined additional possible contributions to WMT failure by performing secondary analyses of variables explored in the prior OEF/OIF literature. Consistent with the study of Armistead-Jehle (2010), we observed a higher WMT failure rate for participants classified with depression than those without the disorder based on BDI-II cutoff score. It is possible, as Armistead-Jehle (2010) suggests, that depression contributes to symptom exaggeration. Depression is associated with cognitive biases and increased subjective complaints of memory disorder and other neurocognitive and somatic symptoms that frequently have little if any relationship to objective conditions (Everaert, Koster, & Derakshan, 2012; Hoerger, Quirk, Chapman, & Duberstein, 2012; Rohling, Green, Allen, & Iverson, 2002; Tylee & Gandhi, 2005). A segment of these individuals may be offering overt demonstration of their perceived impairment to the clinician (or researcher) to emphasize, embellish, or dramatize their mood status. The presence of depression therefore requires particular attention to the SVT outcome and the dissociation of subjective from objective findings in those passing SVTs, particularly in a clinical context. In this regard, whereas in the Dual context the rate of failure for those with depression (58%) was nearly twice the failure rate for those without depression (31%), in the Research context it was only slightly higher for those with (17%) than without depression (14%).

In contrast to the findings of Armistead-Jehle (2010), we additionally observed a significantly higher failure rate for participants classified with PTSD (34%) relative to those without the disorder (19%) based on well-established PCL-M cutoff score. It is possible that Armistead-Jehle (2010) examined an insufficient number of participants without PTSD (n = 4) to demonstrate an effect. As with depression, WMT performance was modified by context: whereas in the Dual context more than twice failed the WMT with PTSD (64%) than without (27%), the rate in the Research context was negligibly higher for those with PTSD (16%) than without (15%). As posited in association with depression, the same possibility exists for symptom exaggeration in PTSD, particularly in a clinical context (Freeman, Powell, & Kimbrell, 2008; Frueh, Grubaugh, Elhai, & Buckley, 2007; McHugh & Treisman, 2007; Rosen & Taylor, 2007). To our knowledge, only Demakis, Gervais, and Rohling (2008) have dedicated investigation to cognitive SVT outcomes in PTSD, reporting a 20% failure rate on the WMT Delayed Recognition section in a sample of patients referred for worker’s compensation claims. Consideration of SVT failure has been absent from the four prior meta-analyses linking PTSD to lower intelligence (Brewin, Andrews, & Valentine, 2000; Ozer, Best, Lipsey, & Weiss, 2003) and verbal-memory impairments (Brewin, Kleiner, Vasterling, & Field, 2007; Johnsen & Asbjørnsen, 2008), and SVT exclusions have frequently been neglected in subsequent publications despite examination of clinical or dual-consenting patients (Aupperle et al., 2012; Burriss, Ayers, Ginsberg, & Powell, 2008; Mackin, Lesselyong, & Yaffe, 2012). The omission of SVT testing in PTSD research, as is the case for depression (Rohling et al., 2002), risks contamination of the knowledge base tying PTSD to compromised neurocognition (Frueh et al., 2007; Rosen & Taylor, 2007).

In addition to observing markedly higher failure rates in the Dual context, we incidentally noted slightly higher percentages of participants failing in the caution classification for Probable Depression (20%) and Probable PTSD (22%), particularly compared with their controls, but also relative to the other variables of interest and the overall sample rate of 18%. These findings are intriguing and raise the general issue of optimal consideration of caution cases in further evaluation of the remaining neuropsychological protocol. One explanation of this finding may be that psychiatric distress per se is directly impairing cognitive efficiency—not to a level that produces overt WMT failure but one that nevertheless compromises performance. Studies have found high WMT failure rates in psychiatric groups with little external incentive to perform poorly, such as those with schizophrenia not involved in any litigation (Gorissen, Sanz, & Schmad, 2005). Gorissen and colleagues (2005) furthermore reported that 25% of their non-psychotic psychiatric controls, over half of which had primary mood or anxiety disorders, failed the WMT. In this regard, the specific cognitive components of the WMT are important to consider: The WMT presents in a paired-associate format numerous related word pairs assumed to be strongly encoded into item-level and associative memory following two learning trials amounting to mere seconds per pair. The Immediate Recognition and Delayed Recognition sections subsequently call for
identification of single items paired with semantically related distractors (similar to pocketknife vs. jackknife). Error feedback is briefly flashed, followed immediately by the next pairing. In addition to the added difficulty possibly ensuing from unanticipated retrieval requirements (two-choice recognition rather than associative recall following encoding), deliberating between similar distractors will entail considerable linguistic discrimination, contextual memory, and frontal-executive search and comparative operations. These task demands, in turn, will necessitate cognitive effort and the integrity of mediating neural systems (Manning, Sperling, Sharan, Rosenberg, & Kahana, 2012; Raposo, Han, & Dobbin, 2009; Telling, Meyer, & Humphreys, 2010). In this regard, PTSD represents one such disorder by phenomenology and experimental evidence that may embody a fundamental disruption of associative memory (Guez et al., 2011).

In support of the contention that the test consumes a non-trivial quantity of cognitive resources and manifests oversensitivity to nervous-system impingement, examination of the WMT using functional magnetic resonance imaging (fMRI) has demonstrated activation in areas tied to the mobilization of mental effort: dorsolateral prefrontal cortex, anterior insula, superior parietal cortex including medial portions, and retrosplenial cortex (Allen, Bigler, Larsen, Goodrich-Hunsaker, & Hopkins, 2007). Likewise, Dean, Victor, Boone, and Arnold (2008) observed that patients with low intelligence quotients (IQs) on the Wechsler Adult Intelligence Scale-Third Edition exhibited higher WMT failure rates than those with higher IQs. Merten, Bossink, and Schmand (2007) reported that patients with mild Alzheimer’s disease failed the WMS Consistency Scale at a rate of 75%. The latter authors furthermore noted that the WMT correlated most highly in a neuropsychological battery with measures of learning and memory, thus indicating the presence of a significant mnemonic load. The anergia of depression, perhaps secondary to anhedonia, may represent one neural-system abnormality in which the cognitive effort available to expend is exceeded by the amount required by the test (Brinkmann, Schüpbach, Joye, & Gendolla, 2009; Poland-Ross et al., 2013; Kessels, Ruis, & Kappelle, 2007).

However, the findings of Green’s group do not support a neurocognitive effect of depression once participants with SVT failure are removed from the sample (Rohling et al., 2002), and, as noted above, the consideration of SVT outcomes has generally been neglected in connection with PTSD (i.e., examination of neurocognitive performance after exclusion of participants failing SVT assessment). Further investigation is warranted of WMT caution cases, including the influence of psychiatric presentation and the specific WMT subtests preferentially involved (e.g., immediate or delayed recognition vs. the consistency measure). More rigorous, independent operationalization of cognitive effort is also required, such as that achieved by fMRI (Barnes, Bullmore, & Suckling, 2009; Browndyke et al., 2008), pupillometry (Kahneman, 1973; Karatekin, Bingham, & White, 2010), systolic blood pressure (Hess & Ennis, 2012), contingent negative variation (Ansari & Derakshan, 2011), and other physiological, psychophysiological, and electrophysiological measures.

Implicit in the assertion that psychiatric distress might contribute to a subtle but discernible decrement in cognitive performance is the assumption that these disorders satisfy the same rigorous standard of validity as imposed by neuropsychological testing—that is, non-biased responding is observed. Notably, our self-report measures resulting in Probable Depression (BDI-II) and PTSD (PCL-M) classifications do not incorporate psychological symptom-validity indices. For the literature more broadly, whereas numerous studies have examined the veracity of cognitive presentations, few have examined the role of psychological variables (e.g., Boone & Lu, 1999), and even fewer have introduced psychological symptom-validity measures to optimize diagnostic fidelity. For example, many of the reports linking symptoms of mTBI to PTSD (Hoge et al., 2004, 2008; Hoge, Aucander, & Miliken, 2006; Hoge, Terhakopian, Castro, Messer, & Engel, 2007; Vasterling et al., 2006) make exclusive use of similar self-report measures that do not incorporate symptom-validity assessment. Future studies should therefore include both cognitive and psychological symptom-validity measures when asserting the role that psychiatric presentation may play in effort and neuropsychological performance (Armistead-Jehle & Buican, 2012; Jones et al., 2012; Lange et al., 2012; Rohling et al., 2002).

Further examination of military-service-related variables additionally yielded findings frequently at odds with prior OEF/OIF-era literature. We examined individuals with and without SC status and found the WMT failure and pass rates to be nearly equivalent at approximately 25% and 58%, respectively. Of those in our sample with TBI and SC, 40% failed. This finding is consistent with Russo (2012) but inconsistent with Armistead-Jehle’s (2010) failure rate of 96% among those with SC and implication of a culture of entitlement within the VA system. However, our sample included nearly equal representation of active and veteran military personnel within three levels of the Polytrauma System of Care, whereas the Armistead-Jehle (2010) sample was composed solely of veterans from a PNS clinic. The possibility also exists that a substantial number of our sample planned but had yet to obtain SC or that the time since injury or reason for SC differed between samples. Likewise, with regard to Whitney and colleagues’ (2009) observation that only active-duty military personnel failed SVT assessment, the current study does not support this finding. Although we are unaware of additional reports of active-duty military specifying OEF/OIF status, Lange and colleagues (2012) recently observed for 143 polytrauma patients assessed for TBI at an echelon V medical center that only 19.4% failed the WMT and one other SVT. Armistead-Jehle and Hansen (2011) noted a similar 20% failure rate for the MSVT in 85 active-duty soldiers mostly (85%) with reported history of concussion evaluated in an outpatient neuropsychology clinic at a U.S. Army health center. However, in a sample of 335 active-duty outpatients examined in a TBI clinic in an Army medical facility, Armistead-Jehle and Buican (2012) reported a failure rate over twice as high on the WMT: 42%.

Jones and colleagues (2012)
observed failure rates of 45% on the WMT and 55% on the VSVT in 502 active-duty personnel assessed at either a TBI clinic or outpatient neuropsychology service at two Army medical centers. Rates of SVT failure in these studies between 19% and 55% thus appear to be comparable to those reported for OEF/OIF veterans. The absence of replication in the current study, evidence from contemporaneous literature, and the low power \((N = 23)\) adopted by Whitney and colleagues (2009) therefore do not support the pursuit of secondary gain among active-duty military personnel that is implied throughout their article. Furthermore, until more is understood about SVT failure in this cohort, application of the Slick and colleagues (1999) criteria, as well as inferences of military personnel deliberately exploiting the system, is likely premature.

In view of the lack of significant findings relating to SC, it is possible that Dual participants with SVT failure could be attempting to over-demonstrate their impairment as more of a cry for help than for maintenance of disability status as posited by Armistead-Jehle (2010). Conversely, those in the Research group who failed the WMT could also be seeking secondary gain, although it is less apparent how this gain would be realized. While symptom validity can be adequately determined, assessing one’s intent remains elusive (Sweet, 1999). Examination of SVT failure and self-perception of current and pre-deployment functioning suggests a relationship between SVT failure and change in cognitive and emotional status and warrants further investigation (McCormick, Rowland, & Yoash-Gantz, 2011).

Consistent with Jones and colleagues (2012), Armistead-Jehle and Buican (2012), and Russo (2012), we did not observe significant effects for education or ethnicity. Similarly, consistent with Jones and colleagues (2012) and Armistead-Jehle and Buican (2012) but unlike Russo (2012), we did not observe age effects. Finally, unlike Armistead-Jehle (2010), Jones and colleagues (2012), Armistead-Jehle and Buican (2012), and Russo (2012), we observed a significant sex effect, but our subsample was small. Given the inconclusive nature of our findings, together with the recent ascension of women into combat and other occupational specialties with increased TBI risk, the topic of male versus female SVT performance in military cohorts requires further study.

Overall, our findings suggest that false-positive diagnosis can be reduced and the process for assessing TBI strengthened by careful attention to validation procedures across the full spectrum of subjective and objective impairments (Spencer, Dragg, Walker, & Bieliauskas, 2010). Research has established that subjective memory complaints have little if any statistical relationship to objective performance and that cognitive SVT failure is associated with psychological symptom exaggeration, particularly complaints of memory impairment (Armistead-Jehle & Buican, 2012; Armistead-Jehle, Gervais, & Green, 2012a, 2012b; Jones et al., 2012; Lange et al., 2012). Careful evaluation of the contribution of comorbidities after application of validity indices, both for neurocognition and for psychopathology, can help to better distinguish factitious, psychogenic, somatoform, and malingered findings from true neurologic contributions to case presentation. With this systematic diagnostic approach, combined with education promoting positive expectations for recovery in cases of positive mTBI, TBI is no longer “occult” and iatrogenesis can be minimized. This procedural path is particularly indicated for contexts holding clinical implications, but, as we have shown, parallel steps are also necessary for studies of TBI in the research context, which does not guarantee SVT performance impervious to failure. In point of fact, these findings argue for consideration of any research report in relation to the test participants’ understanding of the potential uses of the data acquired.

Limitations of the study include the examination of only one primary SVT instrument. Consideration of more than one measure, including embedded indices, would have strengthened identification procedures, although requiring failure of more than one SVT would have effectively lowered our observed rates even further. An additional weakness included the creation of depression and PTSD groups based on cutoff scores from self-administered instruments lacking psychological-validity scales rather than from diagnostic interviews. Thus, a yet undiscovered variable could explain findings for both SVT outcome and self-reported psychiatric status. However, it can be noted that, even in the more ideal situation of standardized interviewer-administered diagnostic classification such as provided by the SCID-IV, no formal procedures for assessing symptom validity are incorporated, and clinicians are not necessarily adept at discerning dissimulation (Binder & Rohling, 1996; Rosen & Taylor, 2007). A further limitation included our use of the PCL-M, which may have resulted in under-detection of PTSD, as only military-related trauma symptoms are assessed. TBIs in the Non-Inpatient sample were self-reported, predominantly mild, variable in time since injury, and not always verifiable through medical-record reviews. Separate analyses were conducted for Probable Depression and PTSD classifications, yet diagnostic overlap existed between groups, and thus the two subsamples were not statistically independent. Our study sample was a complex one, combining inpatients, VA outpatients, non-VA-affiliated veteran research volunteers, and active military personnel without medical concerns. Additionally, while we have described clinical implications of evaluations as a component of the Dual context, it is important to acknowledge that all participants were enrolled in a research protocol entailing self (rather than random or consecutive) selection, informed consent, and remuneration. For these and other reasons, external validity is limited, and the sample may not be fully representative of the active-duty or veteran OEF/OIF-era military cohort or VA patients.

In summary, our findings provide a well powered estimate of base rates of SVT failure, caution, and passing performance in the OEF/OIF cohort and underscore the importance of the test participant’s understanding as to the uses of the data forthcoming from
effort applied to neuropsychological testing. Knowledge that results have clinical application significantly inflates the probability of observing an invalid SVT result in those with TBI, depression, and PTSD. Overall, SVT inclusion and characterization is critical in future studies of the OEF/OIF cohort due to the high and variable rates of SVT performance observed; methodological differences between studies; the complex clinical comorbidity of TBI history, depression, and PTSD; and the substantial implications for VA. Future studies should also include psychological symptom-validity measures when asserting the role psychopathology plays in effort and cognitive performance.

Funding

Preparation of this manuscript was supported in part by the Department of Veterans Affairs Mid-Atlantic Mental Illness Research, Education, and Clinical Center (MIRECC) for the Study of Post-Deployment Mental Health, the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Mental Illness Research and Treatment, and a Department of Veterans Affairs Rehabilitation Research and Development Career Development Award-2 (1IK2RX000703-01 to S.D.M.).

Conflict of Interest

None declared.

Acknowledgements

We thank those who kindly volunteered to participate in this study. The authors also wish to thank members of the Mid-Atlantic MIRECC Registry Workgroup who include: Jean C. Beckham, Mira Brancu, Patrick S. Calhoun, Rita M. Davison, A. Meade Eggleston, Kimberly T. Green, Angela C. Kirby, Harold S. Kudler, John A. Fairbank, Jessica J. Fulton, Jeffrey M. Hoerle, Christine E. Marx, Gregory McCarthy, Scott D. Moore, Rajendra A. Morey, Victoria M. Payne, Mary C. Pender, Jennifer C. Naylor, Jennifer L. Strauss, Kristy Straits-Tröster, Cindy Swinkles, H. Ryan Wagner II, and Richard D. Weiner from the Durham Veterans Affairs Medical Center, Durham, NC, USA; Marinell Miller Mumford from the Hampton Veterans Affairs Medical Center, Hampton, VA, USA; Anthony Fernandez, Adrienne Kinne, Robin Lumpkin, Treven C. Pickett, and William C. Walker from the Hunter Holmes McGuire Veterans Affairs Medical Center, Richmond, VA, USA; and Robin Hurley, Katherine H. Taber, and Jared A. Rowland from the W. G. Hefner Veterans Affairs Medical Center, Salisbury, NC. Additionally, we would like to thank Perry Whitted of the Mid-Atlantic MIRECC for administrative support and Jennifer Y. Hong of Trinity College, Duke University, for assistance with the manuscript. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References


Sigfusdottir, B. J. (2008). "To care for him who shall have borne the battle and for his widow and his orphan" (Abraham Lincoln): The Department of Veterans Affairs (DVA). *Archives of Physical Medicine and Rehabilitation, 89*, 160–162.


