J Head Trauma Rehabil
Vol. 29, No. 1, pp. 1–10

Screening for Postdeployment Conditions: Development and Cross-Validation of an Embedded Validity Scale in the Neurobehavioral Symptom Inventory

Rodney D. Vanderploeg, PhD; Douglas B. Cooper, PhD; Heather G. Belanger, PhD; Alison J. Donnell, PhD; Jan E. Kennedy, PhD; Clifford A. Hopewell, PhD; Steven G. Scott, DO

Objective: To develop and cross-validate internal validity scales for the Neurobehavioral Symptom Inventory (NSI). **Participants:** Four existing data sets were used: (1) outpatient clinical traumatic brain injury (TBI)/neurorehabilitation database from a military site ($n = 403$), (2) National Department of Veterans Affairs TBI evaluation database ($n = 48\,175$), (3) Florida National Guard nonclinical TBI survey database ($n = 3098$), and (4) a cross-validation outpatient clinical TBI/neurorehabilitation database combined across 2 military medical centers ($n = 206$). **Research Design:** Secondary analysis of existing cohort data to develop (study 1) and cross-validate (study 2) internal validity scales for the NSI. **Main Measures:** The NSI, Mild Brain Injury Atypical Symptoms, and Personality Assessment Inventory scores. **Results:** Study 1: Three NSI validity scales were developed, composed of 5 unusual items (Negative Impression Management [NIM5]), 6 low-frequency items (LOW6), and the combination of 10 nonoverlapping items (Validity-10). Cut scores maximizing sensitivity and specificity on these measures were determined, using a Mild Brain Injury Atypical Symptoms score of 8 or more as the criterion for invalidity. Study 2: The same validity scale cut scores again resulted in the highest classification accuracy and optimal balance between sensitivity and specificity in the cross-validation sample, using a Personality Assessment Inventory Negative Impression Management scale with a T score of 75 or higher as the criterion for invalidity. **Conclusions:** The NSI is widely used in the Department of Defense and Veterans Affairs as a symptom-severity assessment following TBI, but is subject to symptom overreporting or exaggeration. This study developed embedded NSI validity scales to facilitate the detection of invalid response styles. The NSI Validity-10 scale appears to hold considerable promise for validity assessment when the NSI is used as a population-screening tool. **Key words:** *assessment*, *concussions*, *mild traumatic brain injury*, *neuropsychology*, *postconcussion syndrome*, *veterans*

Author Affiliations: *Departments of Mental Health and Behavioral Sciences (Drs Vanderploeg, Belanger, and Donnell), and Physical Medicine and Rehabilitation (Dr Scott) and Health Services Research and Development (HSR&D)/Rehabilitation Research and Development (RR&D) Center of Excellence: Maximizing Rehabilitation Outcomes (Drs Vanderploeg, Belanger, and Scott), James A. Haley Veterans' Hospital, Defense and Veterans Brain Injury Center (Drs Vanderploeg, Belanger, Donnell, and Scott), and Department of Psychology (Drs Vanderploeg and Belanger) and Psychiatry and Neurosciences (Dr Vanderploeg), University of South Florida, Tampa, Florida; Defense and Veterans Brain Injury Center (Drs Cooper and Kennedy) and Neurology Service, Department of Medicine (Drs Cooper and Kennedy), San Antonio Military Medical Center, Fort Sam Houston, San Antonio, Texas; and Traumatic Brain Injury Clinic, Darnall Army Medical Center, Fort Hood, Killeen, Texas (Dr Hopewell).*

This research was supported by the Department of Veterans Affairs, Veterans Health Administration, and Defense and Veterans Brain Injury Center and by a Department of Veterans Affairs Health Service Research & Development grant CCN 06-164 to Steven G. Scott, DO. Further support was provided by the James A. Haley Veterans' Hospital and its HSR&D/RR&D center of excellence.

The views expressed herein are those of the authors and do not necessarily reflect the views or the official policy of the Department of Veterans Affairs or the US government. No potential conflict of interest relevant to this article was reported.

The authors declare no conflicts of interest.

Corresponding Author: *Rodney D. Vanderploeg, PhD, James A. Haley Veterans' Hospital Psychology Service (116B), 13000 Bruce B. Downs Blvd, Tampa, FL 33612 (Rodney.vanderploeg@va.gov).*

DOI: 10.1097/HTR.0b013e318281966e

A SIGNIFICANT CHALLENGE facing the Military Healthcare System and Veterans Health Administration (VHA) during this time of military

conflicts is the identification and treatment of postdeployment conditions, including traumatic brain injury (TBI), depression, posttraumatic stress disorder, other anxiety disorders, and substance-use disorder. To ensure identification of service members potentially needing medical services following deployment in Iraq or Afghanistan, both the Department of Defense and the VHA implemented a series of similar postdeployment screening procedures. Both the Department of Defense and VHA findings suggest that as many as 20% of the service members screen positive for a concussion during their time deployed.[1,2] Although the clinical literature strongly suggests a rapid and complete return to baseline functioning in the vast majority of the service members who experience a single mild TBI (mTBI), there are indications that a minority of individuals report symptoms months or years after the injury event.[3]

Within the VHA, a positive TBI screen automatically generates a consult to a TBI specialist or specialty clinic for a Comprehensive TBI Evaluation if the veteran agrees to further assessment and care. This evaluation is to be completed by TBI physician experts and/or interdisciplinary polytrauma/TBI rehabilitation teams. The purposes of the follow-up evaluation are to (1) confirm the diagnosis of TBI, even if the presenting symptoms are felt to be secondary to other factors such as posttraumatic stress disorder, stress, depression, or chronic pain, and (2) institute an appropriate plan for follow-up care (eg, other evaluations or diagnosis-based or symptom-based treatment). The evaluation consists of a review of blast exposures and TBI events, a targeted review of systems, and a physical examination conducted by a licensed medical practitioner with expertise in TBI, typically a physiatrist or a neurologist. Embedded within the Comprehensive TBI Evaluation is a 22-item postconcussive symptom questionnaire called the Neurobehavioral Symptom Inventory (NSI).[4] These 22 symptoms are each rated on a scale of 0 to 4 (none, mild, moderate, severe, and very severe) with 3 different types of symptoms: affective/psychological/stress, somatic/physical, and cognitive. The NSI is completed by patients during the evaluation for the purposes of assessing for the presence of postconcussion symptoms and patient distress surrounding these symptoms. A VHA TBI treatment algorithm has been developed for the 22 symptoms of the NSI.

Unfortunately, because the NSI is a self-report measure, it is particularly susceptible to exaggeration, either due because of intentional efforts to malinger or because of subconscious processes such as somatoform disorders. As the diagnosis of mTBI is frequently also reliant upon self-report, this leaves the Comprehensive TBI Evaluation particularly susceptible to bias. This may adversely affect treatment planning, as well as disability-based evaluations such as military medical evaluation boards and Department of Veterans Affairs (VA) compensation and pension assessments. Unfortunately, the NSI has no validity scale or other mechanism of assessing response bias and has the potential to be manipulated by respondents for a number of reasons.

Evaluation of response bias in veterans and military personnel with self-reported histories of mTBI has been the subject of several recent publications examining performance validity on cognitive measures. Invalid cognitive performance has ranged from 17%[5] to 59%[6,7] in the VA samples. In contrast, evaluation of response bias in veteran and military populations using self-report symptom measures is not as well studied. Whitney and colleagues[8] examined the Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2)[9] and the MMPI-2 Restructured Form[10] Response Bias Scale (RBS) in a sample of 45 veterans involved in outpatient neuropsychological evaluations conducted for clinical purposes. The authors compared RBS scores in veterans who passed with that in those who failed a commonly administered measure of neurocognitive symptom validity, the Test of Memory Malingering.[11] Although the RBS raw score was significantly higher in those who failed the Test of Memory Malingering, both groups showed elevated RBS scores suggestive of exaggerated complaints.

These findings are consistent with Larrabee's distinction between cognitive "performance validity" and "symptom validity" reporting.[12] Invalid cognitive performance does not necessarily translate into invalid symptom reporting, and vice versa. Therefore, these 2 types of validity assessment are not interchangeable but are clearly important to correct interpretation of cognitive performance and symptom self-report.

Interian and colleagues[13] demonstrated that individuals who endorsed an elevated number of pseudoneurological symptoms were more likely to present with more severe levels of anxiety, depression, somatic complaints, and physical dysfunction. Given the nonspecific nature of postconcussive complaints, Cooper and colleagues[14] developed the Mild Brain Injury Atypical Symptoms (mBIAS) scale to assist military and veteran healthcare providers in evaluating self-report in individuals presenting to TBI clinics for postdeployment evaluation. They identified 5 rationally developed atypical neurological items selected from a larger pool of atypical symptoms. The final 5 items were selected by an expert panel consisting of a board-certified neurologist, a board-certified physiatrist, and a senior neuropsychologist, each with extensive experience in working with mTBI. Items were written in a style consistent with the NSI, using a comparable Likert scale. Analysis of psychometric properties demonstrated that a score of 8 or more on these 5 mBIAS items was optimal for the detection of symptom overreporting, as defined by endorsement of symptoms on

both the NSI and the Posttraumatic Stress Disorder Checklist (PCL-M) 1.5 standard deviations above the mean on both scales in their clinical sample. Although the authors acknowledged that such high scores (+1.5 SD on both measures) could reflect severe neurological and psychiatric symptoms, they reasoned that responses in this range across both measures warranted some degree of scrutiny as they could reflect a response bias. This validity approach is similar to the Negative Impression Management (NIM) scale on the Personality Assessment Inventory (PAI) that is composed of items reflecting highly unlikely or bizarre symptoms.[15]

An alternative approach to assessing symptom overreporting is an infrequent item-endorsement approach used on the F scale of the Minnesota Multiphasic Personality Inventory and the MMPI-2 where items were selected on the basis of endorsement by fewer than 10% of a clinical sample. Some of the F-scale items overlap with one of the clinical scales, such as scales 6 or 8.

Administration of psychological test batteries, such as the MMPI-2 Restructured Form or PAI, is not practical for the purpose of screening. In addition, the VHA healthcare system has already implemented the NSI as part of its core battery of brief, self-reported symptom questionnaires. This article describes the development and cross-validation of internal validity scales for the NSI, using these 2 approaches (ie, unlikely or unusual symptom items and low base-rate items), as well as a validity scale combining these 2 approaches.

STUDY 1: DEVELOPMENT OF INTERNAL VALIDITY SCALES FOR THE NSI

Methods

Participants

To identify potentially unusual or highly unlikely items from the NSI, we examined correlations between the NSI and the original 5 items from the mBIAS scale, using the clinical sample at San Antonio Military Medical Center (SAMMC; $n = 443$) that was used for the original mBIAS validation.[14] Identification of low-frequency NSI items was conducted, using 3 independent samples: (*a*) the SAMMC clinically presenting original mBIAS development sample, (*b*) a sample of the Florida National Guard (FLNG) that completed an online survey designed to evaluate the effects of military deployment in Iraq or Afghanistan ($n = 3127$),[16] and (*c*) the VA national data from the Comprehensive TBI Evaluation completed after positive TBI screens ($n = 48\ 175$).[17] Each sample is briefly described.

SAMMC clinically presenting original mBIAS sample

Participants were consecutive referrals to a brain injury clinic at SAMMC. The sample was primarily composed of active duty service members (including activated reservists and members of the National Guard). The sample was composed of 2 patient categories: (*a*) service members with documented or suspected mTBI and (*b*) neurorehabilitation patients. As determined by record review or careful clinical interviews, the majority of patients had mTBIs ($n = 268$), 73 had no neurological impairment, 39 had moderate to severe or penetrating TBIs, and 23 had other neurological conditions requiring rehabilitation services (tumor [$n = 7$], stroke [$n = 6$], electrical injury [$n = 3$], subarachnoid hemorrhage [$n = 2$], anoxia [$n = 2$], encephalitis [$n = 1$], Parkinson disease [$n = 1$], and frontal-temporal dementia ($n = 1$)). This archival study was approved by the SAMMC institutional review board. As part of standard operating procedures, all individuals referred to the clinic completed self-report symptom questionnaires on a computer kiosk before their initial encounter with a medical provider. Participants completed the NSI, the PCL-M, and the mBIAS self-report questionnaires (see the original article for additional participant description).[14]

Department of Veterans Affairs administrative database of the Comprehensive TBI Evaluation in clinically presenting patients

After approval from all regulatory committees (ie, both local and national institutional review boards and Research and Development committees), data were obtained from VA's patient care services centralized database on the TBI Screen and Evaluation results between the dates of October 9, 2007, and June 16, 2010. The results of the Comprehensive TBI Evaluations, including results from the NSI, as reported by the patient during the evaluation, were also captured. Only those cases with both a positive TBI Clinical Reminder Screen and a corresponding Comprehensive TBI Evaluation were included. Of the 48 175 participants, 36 655 reported an alteration or loss of consciousness and/or penetration of the brain (76%). See Belanger and colleagues[17] for additional participant description.

Florida National Guard nonclinical survey sample

In 2009 to 2010, an anonymous online survey was conducted of the currently active members of the FLNG. Approximately 10 400 letters were mailed from the FLNG headquarters to their members inviting participation in a Web-based research survey. The protocol was approved by the institutional review board of the University of South Florida. Of 10 400 letters mailed, approximately 700 letters were returned for insufficient addresses, and 4005 individuals responded for a response rate of 41.3%. Of the 4005 individuals who responded, 423 did not fully complete the survey, 371 completed the survey more than 1 time, and 113 gave inconsistent

or impossible responses, resulting in a final sample of 3098 respondents. Of these respondents, 1655 (53.4%) were not deployed and 1443 (46.6%) were deployed. Of those deployed, they had returned from Operation Enduring Freedom/Operation Iraqi Freedom deployment an average of 31.8 (range, 0-95 months) months before the survey. Only 146 participants reported a probable deployment-related TBI, all of which were categorized as mild with memory gaps of less than a day. Exposures to blasts were reported by 51.5% of the deployed sample, about equally divided between those reporting exposure to a primary blast wave and those reporting blast exposure without primary blast wave contact. See Vanderploeg and colleagues[16] for additional participant description.

Measures

All participants in each of the 3 samples completed the NSI.[4] The NSI is a 22-item self-report inventory of common postconcussive symptoms. Participants rated the presence of each symptom within the past 2 weeks on a 5-item Likert-type scale ranging from 0 (none) to 4 (very severe).

The SAMMC sample also completed the 5-item mBIAS scale of atypical neurological symptoms.[14] Items were written in a style consistent with the NSI, using a comparable Likert-type rating scale. The content of these items were temporary complete deafness, seeing only in black and white, losing one's voice for more than a minute, complete loss of feeling in both arms, and difficulty swallowing because of a lump in the throat.

Statistical analyses and study design

The overall plan was to attempt to develop an atypical neurological NSI validity scale and an infrequency NSI validity scale, as well as a validity scale combining both types of NSI items. First, the SAMMC sample was used to determine the correlations between the total 5-item mBIAS score and each of the 22 NSI items. The NSI items having the highest correlations with the mBIAS scale were considered as potentially unlikely or bizarre symptoms (ie, akin to the PAI's NIM items).

Each of the 3 samples was used to determine the mean severity rating of each item as well as the frequency of item endorsement at a severity level of moderate or higher (ie, a rating of 2 or higher). Within the FLNG sample, mean item ratings and severity frequencies were examined separately for: (*a*) those nondeployed (ie, a potential normative reference group), (*b*) those deployed reporting a mTBI, (*c*) those deployed reporting combat exposure, blast exposure, or both, and (*d*) those deployed who report no TBI, no blast, no combat exposure, and no physical injuries. No a priori severity or frequency ratings were identified for possible use as infrequent or a MMPI-2 F-like scale items. Instead, consistency across all data sets and type of participants was considered in potential selection as an infrequent or MMPI-2 F-like scale item.

Using the SAMMC mBIAS data, correlations between each of these 3 NSI validity scales and the mBIAS scale were examined and compared with the correlations between the individual mBIAS items and the mBIAS total score. Using the criterion standard of mBIAS raw scores of 8 or more as invalid,[14] valid versus invalid group differences on these 3 NSI validity scales were examined. Next, the total raw scores from these NSI validity scales were examined to determine the best cut scores to correspond with mBIAS raw scores of 8 or more.

Results

Correlations between the 22 NSI items and the mBIAS total score ranged from 0.289 to 0.476. Five items had correlations higher than 0.41. These items were hearing problems, noise sensitivity, change in taste or smell, difficulty making decisions, and slowed thinking. Four additional items had correlations between 0.40 and 0.41. The correlations among the 5 individual mBIAS items and the mBIAS total score were also examined. Correlations between individual mBIAS items and the mBIAS total score ranged from 0.519 to 0.668. However, the intercorrelations among the 5 mBIAS items were lower, ranging from 0.171 to 0.257. On the basis of these analyses, the 5 NSI items with correlations above 0.41 were considered as unlikely or bizarre symptoms to compose a NIM scale, NIM5.

The NSI item mean scores and frequencies of reporting moderate or higher levels of severity are reported in Table 1. Six different cohorts of participants were examined and arranged in the table from lowest levels of symptom reporting on the left to highest levels on the right. The FLNG nonclinical sample clearly had lower levels of reporting than the clinically presenting samples. Those FL guard members who were deployed in Iraq or Afghanistan but did not experience physical injuries, mTBIs, or blast exposures had similar levels of symptoms severity as the nondeployed guard members. Across cohorts, 6 items consistently had the lowest severity ratings and frequencies: dizziness, balance problems, coordination difficulties, nausea, vision problems, and change in taste or smell. Although these 6 symptoms may be symptoms of concussion, particularly in the acute phase, these symptoms clearly are reported with the lowest frequency in the chronic phase in all of our samples. If the NIM5 was used to determine validity in the FLNG sample (ie, NIM5 <14), then within the valid reporters, only 1 of these 6 symptoms differed between those deployed with (*a*) other physical injuries (not a mTBI) and (*b*) a deployment-related mTBI. The

TABLE 1 *NSI item mean scores, standard deviations, and frequencies of reporting moderate or higher levels of severity*[a]

	FL National Guard nondeployed			FL National Guard deployed no trauma			FL National Guard blast or combat			FL National Guard deployed TBI			SAMMC mBIAS development			VA national		
	N = 1655			*N* = 256			*N* = 711			*N* = 146			*N* = 443			*N* = 48 175		
NSI item	**Mean**	**SD**	**Freq.**	**Mean**	**SD**	**Freq.**	**Mean**	**SD**	**Freq.**	**Mean**	**SD**	**Freq.**	**Mean**	**SD**	**Freq.**	**Mean**	**SD**	**Freq.**
Dizzy	0.12	0.41	2.3	0.09	0.39	1.6	0.33	0.68	7.9	0.60	0.99	15.1	0.89	0.94	24.2	1.16	0.96	35.1
Balance	0.07	0.31	1.1	0.06	0.33	0.8	0.30	0.70	7.7	0.59	1.03	17.8	1.01	0.98	28.0	1.10	0.99	33.0
Coordination	0.09	0.35	1.9	0.07	0.35	1.2	0.30	0.68	7.0	0.67	1.09	20.5	0.98	0.99	28.9	1.11	1.04	34.0
Headaches	0.39	0.75	8.8	0.52	0.90	15.2	1.00	1.13	32.2	1.60	1.39	48.6	1.70	1.27	54.9	2.12	1.17	71.0
Nausea	0.08	0.36	1.6	0.07	0.37	1.6	0.21	0.58	5.8	0.47	0.93	15.1	0.69	0.91	17.6	0.87	1.04	25.8
Vision	0.11	0.40	2.5	0.11	0.39	2.7	0.45	0.82	12.9	0.68	1.03	17.8	1.01	1.08	30.2	1.24	1.14	39.7
Light sensitive	0.14	0.47	3.7	0.17	0.52	4.7	0.57	0.97	17.7	0.96	1.23	29.5	1.13	1.25	34.8	1.53	1.25	49.8
Hearing	0.13	0.46	3.2	0.27	0.64	6.2	0.84	1.05	26.0	1.08	1.20	34.9	1.21	1.19	37.9	1.77	1.18	60.5
Noise sensitive	0.09	0.41	2.4	0.13	0.47	2.7	0.58	0.96	17.3	0.95	1.22	30.8	1.17	1.23	37.2	1.63	1.27	53.7
Numbness	0.18	0.56	4.7	0.16	0.48	3.9	0.60	0.95	18.0	0.88	1.17	28.1	1.39	1.26	44.5	1.43	1.26	46.3
Taste/smell	0.03	0.22	0.4	0.04	0.27	1.2	0.18	0.58	5.6	0.40	0.88	12.3	0.50	0.87	14.7	0.59	0.98	17.8
Appetite	0.15	0.50	3.1	0.14	0.48	4.7	0.44	0.89	12.8	0.74	1.13	22.6	1.12	1.14	35.4	1.32	1.25	43.5
Concentration	0.23	0.64	6.2	0.25	0.66	7.0	0.89	1.19	27.4	1.45	1.41	47.9	1.76	1.30	55.8	2.11	1.23	69.6
Memory	0.24	0.62	5.6	0.20	0.57	4.7	0.96	1.18	28.7	1.50	1.38	48.6	1.98	1.30	64.8	2.32	1.19	76.0
Decision making	0.13	0.47	3.1	0.09	0.35	1.6	0.57	0.99	16.9	1.03	1.29	32.9	1.22	1.20	41.8	1.52	1.28	49.6
Slow thinking	0.16	0.53	3.6	0.14	0.45	2.7	0.67	1.06	20.0	1.18	1.38	37.7	1.40	1.27	44.9	1.77	1.28	57.3
Fatigue	0.27	0.67	3.6	0.32	0.71	2.7	0.93	1.15	20.0	1.37	1.35	37.7	1.73	1.24	57.6	1.98	1.27	64.6
Sleep	0.36	0.81	9.5	0.38	0.88	11.7	1.13	1.31	36.0	1.59	1.44	50	2.13	1.35	66.6	2.61	1.25	81.3
Anxiety	0.28	0.73	6.9	0.25	0.67	7.0	1.03	1.23	31.5	1.62	1.41	51.4	1.59	1.31	49.4	2.37	1.23	76.4
Depressed	0.24	0.65	5.9	0.21	0.60	6.2	0.77	1.11	22.2	1.27	1.32	39.7	1.10	1.19	33.0	1.89	1.33	60.0
Irritable	0.25	0.66	5.1	0.19	0.59	5.1	1.04	1.28	29.0	1.51	1.48	45.2	1.90	1.32	61.4	2.60	1.18	82.2
Frustration	0.24	0.67	6.5	0.21	0.62	5.1	0.93	1.23	29.3	1.44	1.48	45.2	1.47	1.29	46.0	2.23	1.29	70.7

Abbreviations: ; Freq., frequency; SAMMC, San Antonio Military Medical Center; VA, Department of Veterans Affairs

[a]Six different cohorts of participants were examined and are arranged in the table from lowest levels of symptom reporting on the left to highest levels on the right. Shaded rows indicate the lowest-frequency items.

"impaired coordination" symptom was higher in the other physical injury group than in the mTBI group, otherwise groups did not differ on these symptoms. This suggests that in the chronic phase, these symptoms are not uniquely related to mTBI and may serve as a low-frequency validity scale.

One of these items (ie, change in taste or smell) was also one of the items associated with the mBIAS total score (see earlier). Noise sensitivity was a low-frequency item in the 2 noninjured FLNG cohorts, but not in the other cohorts. On the basis of these analyses, these 6 low-frequency NSI items were considered as MMPI-2 F-like scale items to compose a low-frequency NSI validity scale, LOW6.

In the SAMMC mBIAS data set, the correlation of the mBIAS total score with these NSI validity indicators were moderate: NIM5 scale = 0.566, LOW6 scale = 0.500, and the 10 items combined into a Validity-10 scale = 0.561 (1 item overlapped between the NIM5 and the LOW6 scales). These were comparable to the correlations between the 5 mBIAS items and the mBIAS total score, which were 0.659, 0.518, 0.519, 0.613, and 0.668. Using an mBIAS raw score of 8 or more as invalid,[14] the 3 potential NSI validity scales were compared across valid versus invalid groups. All 3 scales differed significantly across groups: NIM5, $F_{1,441} = 172.37$, $P < .001$; LOW6, $F_{1,441} = 114.79$, $P < .001$; Validity-10, $F_{1,441} = 163.52$, $P < .001$. Table 2 shows group differences on these 3 NSI validity measures.

Logistic regression analyses were conducted to identify potential cut scores on these 3 NSI validity scales, as well as the NSI total raw score, that resulted in the highest validity classification rates, using an mBIAS score of 8 or more as the criterion for invalidity. Analyses revealed the following best cut scores: NIM5, more than 12; LOW6, more than 13; Validity-10, more than 22; and the NSI total score, more than 58. Sensitivity and specificity for these validity cut scores are presented in Table 3. As can be seen, the Validity-10 index provided the best balance of sensitivity and specificity at 0.81 and 0.94, respectively.

TABLE 2 *mBIAS valid versus invalid group differences on the NSI validity scales in the SAMMC mBIAS development data set*

NSI validity scale	Valid, $n = 392$		Invalid, $n = 51$	
	Mean	SD	Mean	SD
NIM5	4.65	3.72	12.02	4.18
LOW6	4.38	3.71	10.45	4.52
Validity-10	8.66	6.33	20.94	7.32

Abbreviations: mBAIS, Mild Brain Injury Atypical Symptoms; NSI, Neurobehavioral Symptom Inventory; SAMMC, San Antonio Military Medical Center.

STUDY 2: CROSS-VALIDATION OF THE NIM5, LOW6, AND VALIDITY-10 NSI VALIDITY SCALES

Methods

Participants

The military cross-validation sample ($n = 206$) was composed of patients presenting to the TBI clinics at SAMMC and Darnall Army Medical Center at Fort Hood. Participants included patients across the spectrum of TBI severity. Individuals were administered the PAI as part of more comprehensive neuropsychological evaluation. The mean age of the sample was 30.9 (SD = 7.2) years. The sample was composed of active duty service members: 6.8% were officers, the remainder were enlisted. As determined by record review or careful clinical interviews, all participants sustained a TBI: 69% mild, 22% moderate, 6% severe, and 3% penetrating. Mechanisms of injuries were blast (54%), motor vehicle accident (16%), fall (12%), bullet (2%), or unknown (16%). Time since injury ranged from 1 to 299 (mean = 31.6, SD = 36.7) weeks. This multisite archival study was approved by the SAMMC institutional review board.

Measures

As part of standard operating procedures for the SAMMC patients with TBI, all individuals referred to the clinic completed self-report symptom questionnaires on a computer kiosk before their initial encounter with a medical provider, including the NSI. They all also completed the PAI as part of a more comprehensive neuropsychological test battery. For the patients with TBI at Darnall Army Medical Center, both self-report checklists and PAI were completed as part of their comprehensive neuropsychological evaluation.

Statistical analyses and study design

The PAI NIM scale with a T score of 75 or higher was used as the criterion standard for invalid symptom reporting (PAI NIM Valid $n = 161$; NIM Invalid $n = 45$).[15] Similar to the aforementioned development study, logistic regression analyses were conducted to identify the best cut scores for the 3 new NSI validity scales (NIM5, LOW6, and Validity-10), as well as the NSI total raw score, that resulted in the highest validity classification. These were compared with the cut scores derived from the development study using the mBIAS to derive the NSI validity cut scores. Sensitivity and specificity of validity classifications were then calculated.

TABLE 3 *Sensitivity and specificity of the NSI validity indices in the SAMMC mBIAS development and military cross-validation data sets*

	Development sample using the mBIAS scale				Cross-validation sample using the PAI NIM scale			
Scale	**NIM5**	**LOW6**	**Valdity-10**	**NSI-Total**	**NIM5**	**LOW6**	**Validity-10**	**NSI-Total**
Cut score	>12	>13	>22	>58	>12	>13	>22	>58
Sensitivity	0.74	0.66	0.81	0.79	0.48	0.62	0.61	0.59
Specificity	0.93	0.91	0.94	0.93	0.82	0.83	0.85	0.84

Abbreviations: mBAIS, Mild Brain Injury Atypical Symptoms; NSI, Neurobehavioral Symptom Inventory; PAI, Personality Assessment Inventory; SAMMC, San Antonio Military Medical Center.

Results

The correlation between the PAI and the new NSI validity scales are presented in Table 4. As can be seen, the correlation between the PAI NIM scale and the new NSI validity scales were moderate, ranging from 0.56 to 0.62. The highest correlation was with the Validity-10 scale. Importantly, the pattern of correlations between both the NIM scale and the new NSI validity scales with the various PAI validity and clinical scales was similar. This suggests that the new NSI validity scales, particularly the Validity-10 scale, have relationships with aspects of psychopathology that are similar to those reflected by the NIM scale, although given the nature of the NSI items, with lower associations with more psychotic conditions. This provides evidence of convergent and discriminate validity for the new NSI validity scales.

TABLE 4 *Correlations between the PAI NIM and NSI validity scales with the PAI clinical scales in the military cross-validation data set*

	NSI validity scales and PAI NIM scale			
PAI scale	**NIM5**	**LOW6**	**Validity-10**	**NIM**
INC	0.11	0.14	0.13	0.28
INF	−0.02	0.04	0.02	0.05
NIM	0.61	0.56	0.63	1.00
PIM	−0.47	−0.33	−0.44	−0.56
SOM	0.68	0.67	0.73	0.70
ANX	0.69	0.56	0.67	0.76
ARD	0.71	0.59	0.70	0.74
DEP	0.71	0.57	0.69	0.76
MAN	0.29	0.28	0.30	0.46
PAR	0.47	0.35	0.44	0.70
SCZ	0.67	0.52	0.64	0.81
BOR	0.59	0.47	0.57	0.77
ANT	0.05	−0.03	0.02	0.24
ALC	−0.02	−0.09	−0.06	0.10
DRG	0.10	0.09	0.04	0.21
AGG	0.46	0.28	0.41	0.59
SUI	0.32	0.24	0.30	0.51
STR	0.45	0.34	0.43	0.58
NON	0.42	0.28	0.38	0.54
RXR	−0.60	−0.44	−0.56	−0.66
DOM	−0.02	0.02	−0.00	−0.07
WRM	−0.36	−0.21	−0.31	−0.48

Abbreviations: NIM, Negative Impression Management; NSI, Neurobehavioral Symptom Inventory; PAI, Personality Assessment Inventory.

The same validity scale cut scores again resulted in the highest classification accuracy and optimal balance between sensitivity and specificity in the cross-validation sample. Sensitivity and specificity for these cut scores are presented in Table 3. As can be seen, the Validity-10 index again provided the best balance of sensitivity and specificity at 0.61 and 0.85, respectively. These cross-validation values are lower than in the original development study, as is the case with most cross-validation studies, especially if a different validity criterion standard is used.

Finally, using these cut scores for the Validity-10 index, percentages of each of the 6 samples that would be classified as invalid were calculated. Invalid percentages were as follows: FLNG nondeployed = 0.1%, FLNG deployed but no trauma experiences = 0%, FLNG with blast or combat exposure = 1.7%, FLNG with deployment-related mTBI = 8.2%, SAMMC mBIAS development sample = 7.0%, military cross-validation sample = 15.2%, and VA national data = 11.6%. Using the cumulative frequency distribution, normalized Validity-10 T scores for these 6 samples are presented in Table 5. Scores were transformed to a normalized distribution by assigning raw scores at a given cumulative percentile to the T score that represents that the percentage in a normal distribution (eg, the most infrequent 5% [ie, 95th cumulative percentile] would correspond to a T score of 66, while the 50th cumulative percentile would correspond to a T score of 50). This table can serve as normative data for clinicians who may chose to use this validity scale.

TABLE 5 *Normalized T scores corresponding to raw Validity-10 scores for different samples*[a]

	T score						
Raw score	FLNG nondeployed, *n* = 1655	FLNG deployed no trauma, *n* = 256	FLNG deployed blast/combat, *n* = 711	FLNG deployed mTBI, *n* = 146	SAMMC mBIAS data, *n* = 443	Military PAI cross-validation, *n* = 206	VA national data, *n* = 48 175
0	55	54	44	43	37	33	30
1	58	58	47	46	39	35	32
2	61	61	50	48	41	38	35
3	63	63	52	49	42	39	37
4	65	64	54	50	44	41	39
5	66	65	55	51	46	43	41
6	67	67	56	52	47	44	42
7	69	70	58	53	48	45	44
8	70	71	59	54	49	45	45
9	71	71	60	55	50	46	47
10	72	72	61	56	52	47	48
11	73	73	62	56	53	47	49
12	74	73	62	57	54	48	50
13	75	74	63	57	55	49	51
14	75	75	65	58	56	50	52
15	76	76	65	59	57	52	54
16	77	77	66	60	58	53	55
17	78	81	67	61	59	54	56
18	78	85	68	61	60	55	57
19	79	87	69	62	61	57	58
20	80	90	71	63	62	58	60
21	80	>90	72	64	63	59	61
22	81		72	64	65	60	62
23	82		72	65	66	62	63
24	>90		72	65	67	63	64
25			73	66	68	64	65
26			75	67	69	66	66
27			76	67	70	69	67
28			76	68	71	70	69
29			77	69	72	72	70
30			77	70	74	74	71
31			77	71	76	75	72
32			77	72	77	76	74
33			77	72	78	>90	75
34			78	73	79		76
35			79	73	>90		78
36			80	74			79
37			80	74			81
38			81	75			86
39			85	83			>90
40			≥90	≥90			

Abbreviations: FLNG, Florida National Guard; mBAIS, Mild Brain Injury Atypical Symptoms; mTBI, mild traumatic brain injury; NSI, Neurobehavioral Symptom Inventory; PAI, Personality Assessment Inventory; SAMMC, San Antonio Military Medical Center, VA, Department of Veterans Affairs.

[a]Raw Validity-10 scores of 23 or higher were considered invalid in the current studies.

GENERAL DISCUSSION

The NSI is a commonly used tool to assess symptoms following concussions or more severe TBIs. The VHA uses the NSI as part of its Comprehensive TBI Evaluation, which has been administered to more than 50 000 veterans to date. However, the NSI does not have an internal validity scale and as a result, is subject to symptom overreporting or exaggeration. The current studies developed and evaluated possible internal validity scales for the NSI. Two alternative approaches,

(*a*) reporting of unlikely symptoms and (*b*) reporting of low frequency but not necessarily unlikely symptoms, were explored independently and then in a combined manner. The combined approach, the Validity-10 scale, resulted in the highest sensitivity and specificity outcomes. As is the case for most scale development and cross-validation studies, sensitivity and specificity were higher in the development sample (0.81 and 0.94, respectively) than in the cross-validation study (0.61 and 0.85, respectively). However, the same Validity-10 scale cut score ($\geq$23) resulted in the best overall group classification (ie, sensitivity/specificity balance) in both studies. It is important to realize that with a sensitivity of 0.61, the Validity-10 scale would incorrectly identify 4 of 10 NSI profiles as valid when they may actually represent exaggeration of symptom reporting. Similarly, a specificity of 0.85 would mean that 1.5 individuals of 10 identified as symptom overreporters may be accurately reporting their symptoms.

This study developed the NSI Validity-10 scale, using the atypical neurological items composing the mBIAS scale as the validity standard, and then cross-validated the Validity-10 scale against the PAI NIM scale. Future research should examine correspondence between the NSI Validity-10 and the various MMPI-2 overreporting scales including F, Fp (Infrequency-Psychopathology), RBS, and FBS (Fake Bad Scale) scales. It remains possible that overendorsement on the NSI Validity-10 scale reflects somatization rather than overt exaggeration. However, detection of either response style is helpful for the treating or evaluating clinician.

As might be expected, symptom reporting was far higher in clinically presenting samples than in an epidemiological sample of the FLNG. Validity-10 scale, using a cut score of 23 or higher, resulted in virtually no one without a history of trauma being classified as invalid in nonclinical sample. Nonclinical participants reporting a deployment-related mTBI had an 8% invalidity rate, while those in the 3 clinical samples had invalidity rates ranging from 7% to 15.2%. These symptom invalidity rates are somewhat lower than expected, given that prior studies have reported performance invalidity rates ranging from 17% to 59%. However, the correspondence between performance invalidity on cognitive measures and symptom invalidity on self-report measures in mTBI samples is unclear. Nevertheless, the cut score needs to be reevaluated in future research of the NSI Validity-10 scale to determine if it could be lower without sacrificing specificity.

Reinforcing the recommendation to further examine the Validity-10 cut score was the finding that the mean score on the Validity-10 scale in the mBIAS invalid group was 20.94. This is lower than the "23 or higher" cut score that resulted in the best sensitivity/specificity balance in this study. However, it is certainly possible that participants may overreport symptoms to a greater extent on the NSI than they did on the mBIAS items, or vice versa. The mBIAS and PAI NIM scores used as the invalidity criterion standards in the current investigation may not be the best standards for determining NSI invalidity. Participants clearly can overreport more on one measure than another and overreport certain types of symptoms more so than others. Furthermore, this NSI invalidity cut score requires an overall rating of "moderate" (ie, 2) or higher on each of the 10 items or requires ratings of "severe" or higher on many items (if some are not rated as problematic). This is a relatively high standard for invalidity, given that these are low-frequency or unusual-symptom items and that the definition of "severe" states that the symptom is "frequently present and disrupts activities; I can only do things that are fairly simple or take little effort."

Although this study found support for using the Validity-10 scale in assessing whether or not NSI symptom overreporting could be a problem, clinicians must also consider other methods of examining validity to corroborate any suspicions raised by the NSI validity measures with additional measures (eg, validity scales on other psychological measures) or behavioral observations. For example, clinicians should consider whether or not symptom self-reporting is consistent with behavioral observations or known day-to-day functioning. If participants report severe sensitivity to noise but have no adverse response to loud and disrupting noises during the evaluation (eg, a loud buzzing noise or a disrupting conversation in the hallway outside of the assessment office), the self-reported difficulty should be considered with a degree of skepticism. Similarly, if on the NSI they report severe problems with slowed thinking or difficulty making decisions but respond quickly to evaluation questions and easily make decisions regarding evaluation or appointment options, the severity of their self-reported problems should be suspected.

On the NSI, ratings of "very severe" are operationally defined as "almost always present and I have been unable to perform at work, school or home due to this problem; I probably cannot function without help." If ratings are very severe on several cognitive NSI items, yet the person recently started a new full-time job (or a difficult academic course load) and manages a family life with several small children, then there would appear to be a discrepancy between the person's self-reported symptom severity and actual day-to-day behavioral abilities suggestive of invalid symptom overreporting. These types of common sense symptom validity considerations should always supplement scale-based validity assessment (eg, an elevated Valdity-10 scale).

While the literature of symptom validity testing in military and veteran populations continues to evolve, assessment of response bias (whether due to overt

exaggeration or somatization) is an important consideration, and one that is largely overlooked in current postdeployment screening procedures. The current studies describe the development and validation of an embedded validity measure, using a tool (ie, the NSI) that is currently in wide use throughout the Department of Defense and VA healthcare systems. The NSI Validity-10 scale appears to hold considerable promise for validity assessment when the NSI is used as a population-screening tool. Specifically, this brief scale appears to be useful for identifying those individuals who might require more comprehensive assessment as part of their evaluation (eg, using the MMPI-2 Restructured Form or other psychological measures with internal validity scales as part of the assessment of self-reported difficulties and symptoms). In addition, it may alert clinicians to rely upon other objective measures, such as injury severity and functional status, rather than relying upon individual self-report.

Finally, symptom underreporting is also possible. That is why the PAI includes a Positive Impression Management scale to evaluate for possible underreporting. The validity indices developed in this study do not address this issue.

REFERENCES

1. Terrio H, Brenner LA, Ivins BJ, et al. Traumatic brain injury screening: preliminary findings in a US army brigade combat team. *J Head Trauma Rehabil*. 2009;24:14–23.
2. Sayer NA, Nelson D, Nugent S. Evaluation of the Veterans Health Administration traumatic brain injury screening program in the upper Midwest. *J Head Trauma Rehabil*. 2011;26:454–467.
3. Vanderploeg RD, Belanger HG, Curtiss G. Mild traumatic brain injury and posttraumatic stress disorder and their associations with health symptoms. *Arch Phys Med Rehabil*. 2009;90:1084–1093.
4. Cicerone KD, Kalmar K. Persistent postconcussion syndrome: the structure of subjective complaints after mild traumatic brain injury. *J Head Trauma Rehabil*. 1995;10:1–17.
5. Whitney KA, Shepard PH, Williams AL, Davis JL, Adams KM. The Medical Symptom Validity Test in the evaluation of Operation Iraqi Freedom/Operation Enduring Freedom soldiers: a preliminary study. *Arch Clin Neuropsychol*. 2009;24:145–152.
6. Armistead-Jehle P. Symptom validity test performance in US veterans referred for evaluation of mild TBI. *Appl Neuropsychol*. 2010;17:52–59.
7. Nelson NW, Hoelzle JB, McGuire KA, Ferrier-Auerbach AG, Choarlesworht MJ, Sponheim SR. Evaluation context impacts neuropsychological performance of OEF/OIF veterans with reported combat-related concussion. *Arch Clin Neuropsychol*. 2010;25:713–723.
8. Whitney KA, Davis JJ, Shepard PH, Herman SM. Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Arch Clin Neuropsychol*. 2008;23:777–786.
9. Butcher JN, Dahlstrom WG, Graham JR, Tellegen A, Kaemmer B. *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration and Scoring*. Minneapolis, MN: University of Minnesota Press; 2001.
10. Ben-Porath YS, Tellegen A. *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Manual for Administration, Scoring, and Interpretation*. Minneapolis, MN: University of Minnesota Press; 2008.
11. Tombaugh TN. *The Test of Memory Malingering*. Los Angeles, CA: Western Psychology Corporation; 1996.
12. Larrabee GJ. Performance validity and symptom validity in neuropsychological assessment. *J Int Neuropsychol Soc*. 2012;18:1–7.
13. Interian A, Gara MA, Díaz-Martínez AM, et al. The value of pseudoneurological symptoms for assessing psychopathology in primary care. *Psychosom Med*. 2004;66:141–146.
14. Cooper DB, Nelson L, Armistead-Jehle P, Bowles AO. Utility of the Mild Brain Injury Atypical Symptoms Scale as a screening measure for symptom over-reporting in Operation Enduring Freedom/Operation Iraqi Freedom service members with postconcussive complaints. *Arch Clin Neuropsychol*. 2011;26:718–727.
15. Morey LC. *Personality Assessment Inventory Professional Manual*. 2nd ed. Odessa, FL: Psychological Assessment Resources; 2007.
16. Vanderploeg RD, Belanger HG, Horner RD, et al. Health outcomes associated with military deployment: mild traumatic brain injury, blast, trauma, and combat associations in the Florida National Guard. *Arch Phys Med Rehabil*. 2012;93:1887–1895.
17. Belanger HG, Vanderploeg RD, Soble JR, Richardson M, Groer S. Validity of the Veterans Health Administration's traumatic brain injury screen. *Arch Phys Med Rehabil*. 2012;93:1234–1239.