Archives of Clinical Neuropsychology 00 (2020) 1–8

Personality Assessment Inventory Cognitive Bias Scale: Validation in a Military Sample

Patrick Armistead-Jehle[1,*], [†]Paul B. Ingram[2], Nicole M. Morris[2]

[1]Munson Army Health Center, Fort Leavenworth, KS, USA
[2]Department of Psychological Sciences, Texas Tech University, Lubbock, TX, USA

*Correspondence author at: Concussion Clinic, Munson Army Health Center, 550 Pope Ave. Fort Leavenworth KS 66027, USA. Tel: (913) 684-6790; Fax: (913) 684-6766. *E-mail address*: patrick.j.armistead-jehle.civ@mail.mil.

[†]The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Defense Health Agency, Department of Veterans Affairs, or U.S. Government position, policy or decision unless so designated by other official documentation.

Received 5 February 2020; revised 13 April 2020; Accepted 23 June 2020

Abstract

Objective: Recently, in a mixed neuropsychological outpatient sample, a measure of cognitive response bias has been developed for the Personality Assessment Inventory (PAI) called the Cognitive Bias Scale (CBS). This study sought to cross-validate this measure in a military sample.
Method: Retrospective review of 197 active duty soldiers referred to an Army outpatient clinic for neuropsychological evaluation. Groups were created based on the number of failed performance validity tests (0, 1, or 2–3 performance validity testing [PVT] failures).
Results: The magnitude of effect for the 10-item CBS scale was medium-to-large when comparing those with one PVT failure to those with two to three ($d = .98$) and those with no failures ($d = 1.21$); however, effects between the 1 and 2–3 PVT failure groups were less pronounced. In 1 and 2–3 PVT failure groups, a score of ≥ 16 had high specificity (.92 and .95, respectively) and low to moderate sensitivity (.20 and .55, respectively).
Conclusions: In a military sample, the CBS demonstrated high specificity with relatively low sensitivity. The measure operated similarly to the original study and the current data supports the CBS to rule in, but not rule out, over-reported cognitive symptoms on the PAI.

Keywords: PAI CBS; symptom validity; military

Introduction

Validity testing is now a recommended and common component in neuropsychological practice (Bush et al., 2005; Heilbronner et al., 2009; Martin, Schroeder, and Odland, 2015). Generally speaking, such testing falls into two categories: (a) performance validity (defined as the credibility of an examinee's performances on objective cognitive tests) and (b) symptom validity (defined as honesty of an examinee's self-reported symptoms; Larrabee, 2012). Arguably, the most well-developed scales of symptom validity can be found in the MMPI-2 (Butcher et al., 2001), MMPI-2-RF (Ben-Porath and Tellegen, 2008/2011), and Personality Assessment Inventory (PAI; Morey, 1991). With these measures, efforts have been made to develop various scales that tap different aspects of over-reporting (e.g., over-reporting of psychological, somatic, and cognitive symptoms).

On the MMPI-2 and MMPI-2-RF, the over-report of cognitive symptoms is captured with the Response Bias Scale (Gervais, Ben-Porath, Wygant, and Green, 2007), which utilized failure on performance validity testing (PVT) as a criterion for item selection. Accordingly, this scale has performed well at identifying over-reported cognitive symptoms and is widely relied on

doi:10.1093/arclin/acaa049 Advance Access publication 00 Month 0000

in neuropsychological settings. Historically, there has been no similar method to evaluate over-report of cognitive symptoms on the PAI, as the validity scales have focused on misrepresented psychological symptoms. However, Gaasedelen, Whiteside, Altmaier, Welch, and Basso (2019) recently developed the Cognitive Bias Scale (CBS), which is essentially a PAI analog to the MMPI-2/MMPI-2-RF RBS. As part of that validation, they used the methodology employed to create the RBS, including failure on two PVTs as a criterion for item selection. These authors utilized a sample of 306 outpatients with mixed neurologic diagnoses to identify and cross validate the CBS. In these data, the CBS better predicted failure on performance validity testing than the existing PAI SVTs. Likewise, the effect size for discriminating valid versus invalid performance validity was high (Cohen's $d = -0.96$) and AUC was acceptable at 0.72. Classification statistics indicated that a cut score of ≥16 had specificity at the recommended level of at least .90, with the highest available sensitivity (.37). At this cut score, positive and negative predictive powers (with a base rate of 38%) were .70. The authors concluded the CBS might have utility in detecting the over-report of cognitive symptoms, even though further cross validation was necessary.

The current study sought to cross validate the CBS in a sample of active duty military personnel seen for outpatient neuropsychological evaluation. We hypothesized, given the similar diagnostic makeup of the current sample to relative Gaasedelen et al's sample, that the CBS would adequately discriminate between valid versus invalid groups and a cut score at or around 16 would provide a specificity of over .90 with lower sensitivity.

Methods

Participants

The study included 197 Active Duty soldiers referred for evaluation to an outpatient neuropsychology clinic at a U.S. Army Health Center. The average age of the sample was 38.1 years (SD = 8.2), with an average education of 15.8 years (SD = 2.5). In general, the sample was male, white, and contained slightly more officers than enlisted (20% E1-E5 and 22.5% E6-E9). All participants were Army active duty and 65.5% had a remote history of mild traumatic brain injury (mTBI; months since most recent injury M = 89.5, SD = 84.0). Diagnosis of mTBI was operationally defined via the DoD/VA criterions. More specifically, following the injury event: loss of consciousness 30 min or less; post-traumatic amnesia 24 hr or less; self-reported alteration of consciousness/mental state lasting a moment to 24 hr; or Glasgow Coma Scale score ≥ 13. Nonmutually exclusive diagnoses for those with psychiatric conditions included post-traumatic stress disorder (10.2%), anxiety disorder (26.9%), and depressive disorder (10.2%). Approximately 33% had no psychiatric diagnosis. Diagnoses were made by the first author (a board-certified neuropsychologist) based on the totality of the data available at the time of evaluation. Concussion diagnosis was based on clinical interview and medical records review. Psychiatric diagnoses were made on the basis of clinical interview, medical record review, and psychological test data. Extended demographic characteristics are provided in Table 1, along with comparisons between groups. There were no differences between the PVT failure groups in terms of gender, diagnosis, or rank. The group that failed 2–3 PVTs had a slightly higher number of African Americans, but were otherwise similar.

Procedures and Measures

All patients were tested by a trained neuropsychology technician under the supervision of the first author. All participants were administrated a battery of neuropsychological tests, which included the Medical Symptom Validity Test (MSVT) and Nonverbal Medical Symptom Validity Test (NV-MSVT), Repeatable Battery of Neuropsychological Status (RBANS), and the Personality Assessment Inventory (PAI; see Table 1).

Repeatable Battery of Neuropsychological Status. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) contains the effort index (EI; Silverberg, Wertheimer, & Fichtenberg, 2007); a scaled composite index compiled from two battery subtests that has demonstrated good specificity in classifying those with extra-test performance validity testing failure within military samples (Armistead-Jehle and Hansen, 2011). Of the 132 who took the RBANS, 35 (17.5%) failed the EI.

Medical Symptom Validity Test. The MSVT (Green, 2004) is a brief automated verbal memory screening with several subtests designed to measure performance validity. In addition to data presented in the manual, several studies have demonstrated the utility of this measure in the discrimination between those with genuine memory impairment and those simulating impairment

Table 1. Demographic and Descriptive Summary

Demographic Variables	Full Sample ($n = 197$)			0 PVT Failures ($n = 131$)			1 PVT Failure ($n = 30$)			2–3 PVT failures ($n = 36$)			χ^2
	n	%		*n*	%		*n*	%		*n*	%		
Gender (male)	166	83.0%		112	85.5%		21	70.0%		30	83.3%		4.12^{ns}
Ethnicity													28.82∗
White	135	67.5%		98	74.8%		16	53.3%		20	55.6%		
African American	38	19.0%		18	13.7%		5	16.7%		13	36.1%		
Hispanic	19	9.5%		12	9.2%		6	20.0%		1	2.8%		
Asian	4	2.0%		2	1.5%		1	3.3%		1	2.8%		
Native American/Pacific Islander	1	0.5%		0	0.0%		0	0.0%		1	2.8%		
Other	3	1.5%		1	0.8%		2	6.7%		0	0.0%		
Diagnosis													6.57^{ns}
History of concussion	120	63.0%		76	58.0%		19	63.3%		25	69.4%		
Psychiatric diagnoses	43	21.5%		25	19.1%		6	20.0%		10	27.8%		
Other	19	9.5%		16	12.2%		3	10.0%		0	0.0%		
None	18	9.0%		14	10.7%		2	6.7%		1	2.8%		
Rank													8.40^{ns}
Enlisted	85	42.5%		47	35.9%		16	53.3%		20	55.6%		
Officer	115	57.5%		84	64.1%		14	46.7%		16	44.4%		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD	ANOVA
Years of education	197	15.8	2.5	131	16.0	2.4	30	15.4	2.7	36	15.6	2.5	$F(2,196) = 1.09^{ns}$
Age	197	38.1	8.2	131	38.1	7.9	30	38.0	9.4	36	38.7	8.2	$F(2, 196) = 0.09^{ns}$
Cognitive Test Scores													
WTAR predicted FSIQ	132	107.2	9.0	84	108.6	9.2	17	104.2	9.1	31	105.2	7.9	$F(2, 131) = 2.72^{ns}$
RBANS													
Total	132	92.8	14.5	84	98.0	11.9	17	91.8	11.0	31	79.0	13.7	$F(2, 129) = 27.34$∗
Immediate Memory	132	96.8	15.6	84	99.7	14.6	17	95.5	15.2	31	89.4	16.7	$F(2, 129) = 5.29$∗
Visuospatial/Construction	132	95.5	16.1	84	98.8	15.1	17	95.7	15.4	31	86.7	16.2	$F(2, 129) = 6.90$∗
Language	132	97.5	13.1	84	99.8	12.0	17	100.6	10.0	31	89.5	14.7	$F(2, 129) = 8.31$∗
Attention	132	93.7	16.0	84	99.2	14.3	17	88.9	6.9	31	81.2	16.3	$F(2, 129) = 19.57$∗
Delayed Memory	132	90.0	17.7	84	96.7	12.2	17	90.1	17.2	31	71.7	17.9	$F(2, 129) = 34.19$∗
COWA	191	43.4	9.5	127	45.0	9.9	29	41.5	7.9	35	38.8	7.7	$F(2, 188) = 12.08$∗
Trail Making Test A	194	44.5	14.1	129	47.3	13.8	29	43.8	12.3	36	34.9	12.4	$F(2, 191) = 13.11$∗
Trail Making Test B	194	46.2	11.3	129	48.8	10.7	29	44.1	10.0	36	38.7	11.1	$F(2, 191) = 6.97$∗

Note: ∗ $p < .01$, ns = nonsignificant. WTAR = Weschler Test of Adult Reading, mTBI = mild Traumatic Brain Injury, COWA = Controlled Oral Word Association.

Table 2. Classification Statistics for Scores on the PAI CBS-10

0 PVT vs. 1 PVT Failure											
			20% Base Rate			30% Base Rate			40% Base Rate		
Cut Score	Sens.	Spec.	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.02	1.00	.80	1.00	.80	.70	1.00	.70	.61	1.00	.60
25	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
24	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
23	.05	.99	.80	.60	.81	.71	.72	.71	.61	.70	.61
22	.05	.99	.80	.60	.81	.71	.72	.71	.61	.70	.61
21	.06	.99	.81	.66	.81	.71	.77	.71	.62	.74	.61
20	.06	.98	.79	.40	.81	.70	.53	.71	.61	.64	.61
19	.06	.97	.79	.33	.80	.70	.46	.71	.61	.58	.61
18	.09	.97	.79	.43	.81	.71	.56	.71	.62	.66	.62
17	.14	.95	.79	.43	.82	.71	.56	.72	.63	.66	.62
16	.20	.95	.80	.48	.83	.72	.61	.73	.65	.71	.64
15	.23	.93	.79	.45	.83	.72	.59	.74	.65	.69	.64
14	.32	.92	.80	.51	.84	.74	.64	.76	.68	.74	.67
13	.39	.86	.77	.42	.85	.72	.55	.77	.68	.66	.68
12	.47	.76	.70	.33	.85	.68	.46	.77	.65	.57	.68
11	.50	.72	.67	.31	.85	.65	.43	.77	.63	.54	.68
10	.59	.67	.66	.31	.87	.65	.44	.79	.64	.55	.71

0 PVT vs. 2–3 PVT Failures

			20% Base Rate			30% Base Rate			40% Base Rate		
Cut Score	Sens.	Spec.	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.03	1.00	.81	1.00	.81	.71	1.00	.71	.61	1.00	.61
25	.03	1.00	.81	1.00	.81	.71	1.00	.71	.61	1.00	.61
24	.07	1.00	.81	1.00	.81	.72	1.00	.71	.63	1.00	.62
23	.07	1.00	.81	1.00	.81	.72	1.00	.71	.63	1.00	.62
22	.07	1.00	.80	.57	.81	.71	.69	.71	.62	.78	.61
21	.10	.99	.81	.66	.81	.72	.77	.72	.63	.84	.62
20	.13	.99	.81	.66	.81	.72	.77	.72	.63	.84	.62
19	.23	.96	.80	.47	.82	.71	.60	.72	.63	.70	.62
18	.32	.95	.81	.54	.83	.73	.66	.74	.66	.76	.65
17	.42	.95	.83	.62	.85	.76	.74	.77	.70	.82	.68
16	.55	.92	.83	.59	.83	.77	.71	.79	.72	.79	.71
15	.55	.89	.82	.56	.89	.79	.68	.82	.75	.77	.75
14	.71	.80	.75	.41	.88	.73	.55	.81	.70	.65	.73
13	.74	.78	.77	.45	.91	.76	.58	.86	.75	.68	.80
12	.77	.74	.74	.42	.92	.74	.55	.87	.74	.66	.81
11	.77	.68	.70	.38	.92	.71	.51	.88	.72	.62	.82
10	.81	.63	.67	.36	.93	.69	.49	.88	.70	.60	.83

0 PVT vs. Any PVT Failure

			20% Base Rate			30% Base Rate			40% Base Rate		
Cut Score	Sens.	Spec.	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.00	1.00	.80	1.00	.80	.70	1.00	.70	.60	1.00	.60
25	.02	1.00	.80	1.00	.80	.70	1.00	.70	.61	1.00	.61
24	.03	1.00	.81	1.00	.80	.71	1.00	.71	.61	1.00	.61
23	.03	1.00	.81	1.00	.80	.71	1.00	.71	.61	1.00	.61
22	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
21	.05	.99	.80	.60	.81	.71	.72	.71	.61	.80	.61
20	.06	.99	.81	.66	.81	.71	.77	.71	.62	.84	.61
19	.08	.97	.79	.38	.81	.70	.52	.71	.61	.62	.61

(Continued)

Table 2. Continued

Cut Score	Sens.	Spec.	20% Base Rate			30% Base Rate			40% Base Rate		
			Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
18	.12	.95	.78	.36	.81	.70	.49	.72	.62	.60	.62
17	.20	.95	.80	.48	.83	.72	.61	.73	.65	.71	.64
16	.27	.93	.80	.50	.84	.73	.63	.75	.67	.73	.66
15	.36	.89	.78	.44	.85	.73	.58	.76	.68	.68	.68
14	.38	.82	.73	.34	.84	.69	.47	.75	.64	.58	.66
13	.47	.80	.74	.37	.86	.70	.50	.78	.67	.61	.69
12	.50	.76	.70	.34	.86	.68	.47	.78	.65	.58	.69
11	.55	.69	.66	.31	.86	.65	.43	.78	.63	.54	.70
10	.64	.63	.63	.30	.87	.63	.42	.80	.63	.53	.72

0 PVT vs. 1 PVT Failure

Cut Score	Sens.	Spec.	20% Base Rate			30% Base Rate			40% Base Rate		
			Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.02	1.00	.80	1.00	.80	.70	1.00	.70	.61	1.00	.60
25	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
24	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
23	.05	.99	.80	.60	.81	.71	.72	.71	.61	.70	.61
22	.05	.99	.80	.60	.81	.71	.72	.71	.61	.70	.61
21	.06	.99	.81	.66	.81	.71	.77	.71	.62	.74	.61
20	.06	.98	.79	.40	.81	.70	.53	.71	.61	.64	.61
19	.06	.97	.79	.33	.80	.70	.46	.71	.61	.58	.61
18	.09	.97	.79	.43	.81	.71	.56	.71	.62	.66	.62
17	.14	.95	.79	.43	.82	.71	.56	.72	.63	.66	.62
16	.20	.95	.80	.48	.83	.72	.61	.73	.65	.71	.64
15	.23	.93	.79	.45	.83	.72	.59	.74	.65	.69	.64
14	.32	.92	.80	.51	.84	.74	.64	.76	.68	.74	.67
13	.39	.86	.77	.42	.85	.72	.55	.77	.68	.66	.68
12	.47	.76	.70	.33	.85	.68	.46	.77	.65	.57	.68
11	.50	.72	.67	.31	.85	.65	.43	.77	.63	.54	.68
10	.59	.67	.66	.31	.87	.65	.44	.79	.64	.55	.71

0 PVT vs. 2–3 PVT Failures

Cut Score	Sens.	Spec.	20% Base Rate			30% Base Rate			40% Base Rate		
			Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.03	1.00	.81	1.00	.81	.71	1.00	.71	.61	1.00	.61
25	.03	1.00	.81	1.00	.81	.71	1.00	.71	.61	1.00	.61
24	.07	1.00	.81	1.00	.81	.72	1.00	.71	.63	1.00	.62
23	.07	1.00	.81	1.00	.81	.72	1.00	.71	.63	1.00	.62
22	.07	1.00	.80	.57	.81	.71	.69	.71	.62	.78	.61
21	.10	.99	.81	.66	.81	.72	.77	.72	.63	.84	.62
20	.13	.99	.81	.66	.81	.72	.77	.72	.63	.84	.62
19	.23	.96	.80	.47	.82	.71	.60	.72	.63	.70	.62
18	.32	.95	.81	.54	.83	.73	.66	.74	.66	.76	.65
17	.42	.95	.83	.62	.85	.76	.74	.77	.70	.82	.68
16	.55	.92	.83	.59	.83	.77	.71	.79	.72	.79	.71
15	.55	.89	.82	.56	.89	.79	.68	.82	.75	.77	.75
14	.71	.80	.75	.41	.88	.73	.55	.81	.70	.65	.73
13	.74	.78	.77	.45	.91	.76	.58	.86	.75	.68	.80
12	.77	.74	.74	.42	.92	.74	.55	.87	.74	.66	.81
11	.77	.68	.70	.38	.92	.71	.51	.88	.72	.62	.82
10	.81	.63	.67	.36	.93	.69	.49	.88	.70	.60	.83

0 PVT vs. Any PVT Failure

(Continued)

Table 2. Classification Statistics for Scores on the PAI CBS-10

			20% Base Rate			30% Base Rate			40% Base Rate		
Cut Score	Sens.	Spec.	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
26	.00	1.00	.80	1.00	.80	.70	1.00	.70	.60	1.00	.60
25	.02	1.00	.80	1.00	.80	.70	1.00	.70	.61	1.00	.61
24	.03	1.00	.81	1.00	.80	.71	1.00	.71	.61	1.00	.61
23	.03	1.00	.81	1.00	.80	.71	1.00	.71	.61	1.00	.61
22	.03	.99	.80	.50	.80	.70	.63	.70	.61	.73	.61
21	.05	.99	.80	.60	.81	.71	.72	.71	.61	.80	.61
20	.06	.99	.81	.66	.81	.71	.77	.71	.62	.84	.61
19	.08	.97	.79	.38	.81	.70	.52	.71	.61	.62	.61
18	.12	.95	.78	.36	.81	.70	.49	.72	.62	.60	.62
17	.20	.95	.80	.48	.83	.72	.61	.73	.65	.71	.64
16	.27	.93	.80	.50	.84	.73	.63	.75	.67	.73	.66
15	.36	.89	.78	.44	.85	.73	.58	.76	.68	.68	.68
14	.38	.82	.73	.34	.84	.69	.47	.75	.64	.58	.66
13	.47	.80	.74	.37	.86	.70	.50	.78	.67	.61	.69
12	.50	.76	.70	.34	.86	.68	.47	.78	.65	.58	.69
11	.55	.69	.66	.31	.86	.65	.43	.78	.63	.54	.70
10	.64	.63	.63	.30	.87	.63	.42	.80	.63	.53	.72

Note. CBS-10 = Cognitive Bias Scale-10; Sens = Sensitivity, Spec = Specificity; PPP = Positive Predictive Power; NPP = Negative Predictive Power.

in a range of patient samples (see Carone, 2009 for review). In this sample, 54 (27.6%) of 196 possible participants failed the MSVT.

Nonverbal Medical Symptom Validity Test. The NV-MSVT (Green, 2008) is a brief automated nonverbal memory screening with several subtests designed to measure performance validity. The NV-MSVT as a measure of performance validity has been widely validated (see Wager & Howe, 2010 for review). Of the 194 possible cases, 35 (18.0%) failed the NV-MSVT.

Personality Assessment Inventory. The PAI (Morey, 1991) is an actuarial measure of personality and emotional functioning that consists of 344 items answered on a 4-point Likert format that render 22 nonoverlapping scales. The measure renders four primary validity scales (Inconsistency [ICN], Infrequency [INF], Positive Impression Management [PIM], and Negative Impression Management [NIM]). The PAI also renders 18 clinical scale and 31 clinical subscales that were not employed in the current study. The psychometric properties of the PAI have been well established (see Kurtz and Blais, 2007; Morey, 1991). As noted above, Gaasedelen et al. (2019) established the CBS, which includes 10 items from the PAI (items 33, 72, 113, 166, 206, 209, 242 [F], 252 [F], 274, and 304 [F]). Per Gaasedelen et al, “The scale itself is unit-weighted, with true items being scored as 0 = false, 1 = somewhat true, 2 = mainly true, and 3 = very true. Items that are keyed as false are reverse scored.”

Statistical Analyses

Participants were excluded if their scores on INF ($n = 5$) or ICN ($n = 1$) exceeded the recommended values in the PAI’s technical manual (Morey, 1991). Participants were grouped according the number of failed PVTs concurrently administered with the PAI (e.g., NV-MSVT, MSVT, and RBANS EI). Analysis of variance (ANOVA) was utilized to examine between-group differences across these groups for CBS scores. Effect size estimates were also calculated. Lastly, receiver operator curve (ROC) characteristics and estimated sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP) for CBS scores were calculated.

Results

The CBS had an overall mean of 9.3, a standard deviation of 4.8, and scores ranging from 2–31, consistent with scores reported during the CBS’s development (Gaasedelen et al., 2019). CBS scores were compared using a univariate ANOVA. Based on ANOVA results using Tukey post-hoc testing, there were significant differences ($F(2, 196) = 16.12$, $p < .001$) between those failing zero PVTs ($M = 8.3$, SD = 4.2) and those who failed one (M = 9.5, SD = 4.0 $p < .01$) and two-three

($M = 13.0$, SD $= 5.8$, $p < .001$) PVT(s). One PVT and two-three PVT failure groups were also significantly different ($p < .01$). A separate ANOVA comparing the zero failure and any PVT failure group was likewise significant ($M = 11.9$ SD $= 4.7$, $p < .001$). Effect size estimates were calculated to aid in interpreting the magnitudes of these differences (zero Failures vs. one Failure $d = .29$; one Failure vs. two-three Failures $d = .98$; zero Failures vs. two-three Failures $d = 1.21$; zero Failure vs. Any PTV failures $d = .50$). These results indicate that as the number of PVT failures increase, so do scores on the CBS. Specifically, those with two-three PVT failures have large effect differences compared to both the zero and the one PVT failure groups. When post-hoc testing is examined, collateral neurocognitive assessment scores display a similar pattern across validity response groups (see Table 1). Trails A, as well as RBANS Delayed Memory, Language, and Total Scores significantly differentiated between all three groups. The remaining neurocognitive test scores (e.g., Trials B, COWA, and RBANS Immediate Memory, Visual Spatial, and Attention) significantly differed between the zero and one failure as well as 1 and 2–3 failure group(s), but not between the 1 and 2–3 failure groups.

ROC analysis was calculated, along with sensitivity and specificity estimates (Table 2), for CBS raw scores between: (a) zero and one PVT failure (area under curve [AUC] = .60; 95% confidence interval [CI] = .49–.71); (2) 0 and 2–3 failures (AUC = .79; 95% CI = .69–.88); and (3) 0 and any PVT failure (AUC = .65; 95% CI = .56–.73). Differentiating those with 0 and 1 PVT failure, a CBS score of ≥16 demonstrated sensitivity of .20 and specificity of .95, while a cut score of ≥14 was associated with a sensitivity of .32 and specificity of .92. These two scores most closely approximate the commonly accepted .95 and .90 specificity values, respectively. PPP and NPP metrics are also provided for PVT failure base rates of 20, 30, and 40%. Across base rate estimates, a CBS raw score of 14 provided slightly higher PPP and NPP than a score of 16 when classifying those who passed all or failed one PVT. In contrast, ROC analysis between the 0 and 2–3 failure groups found a much better classification capacity. A CBS score of ≥16 had a specificity of .95 and a sensitivity of .32, while a score of ≥15 had a specificity of .89 and a sensitivity of .55. The PPP and NPP were slightly better with a score ≥ 16. As anticipated, the classification statistics for zero versus any PVT failures fall in between the zero and one PVT failure and the 0 and 2–3 PVT failures. A CBS cut score of ≥16 had adequate specificity of .93, but limited sensitivity at .27.

Discussion

The current study aimed to cross validate the PAI CBS in a sample of active duty soldiers. As evidenced by medium to large effect sizes, the CBS discriminated well between zero failed PVTs and those with 1 and 2–3 PVT failures. Effect sizes in the current data between the zero and one failure (Cohen's $d = 1.21$) and 0 and 2–3 failure groups (Cohen's $d = .98$) were similar to the original data obtained by Gaasedelen et al. (2019) (Cohen's $d = -0.96$). Likewise, descriptive statistics of the CBS were similar to those in the initial validation. ROC analyses provided further evidence of the measures utility. Although the AUC of the zero and one PVT failure was somewhat lower (.60) than the original study (.72), the AUC observed between the zero and two-three PVT failure groups (.79) is better. These differences in AUC values may reflect methodological differences as Gaasedelen et al. (2019) compared those who failed two PVTs (at least one of which was a stand-alone PVT).

The optimal cut score in the current data for one PVT failure was ≥14. For any PVT failures or two-three PVT failures the optimal cut score was ≥16, as these scores maximized sensitivity at an acceptable level of specificity (.92 or greater). Gaasedelen et al. (2019) identified cut scores of ≥16 and ≥ 19 as demonstrating the highest levels of specificity (.90 and .96, respectively). In their data, sensitivity was .37 for ≥16 and .31 for ≥19. In the current data, a cut score of ≥16 had high specificity (.95, .92, and .93) with marginal to acceptable sensitivity (.20, .55, and .27), whereas a cut score of ≥19 also had excellent specificity (.97 and .96), but low sensitivity (.06, .08, and .23). Trends in the current data are similar to Gaasedelen et al. (2019), in that the CBS (at the appropriate cut score) has minimal risk for false positives, but (owing to relatively lower sensitivity) will more frequently fail to identify many individuals who will fail PVTs. To this end, the current data echo that of Gassedelen and colleagues. More specifically, the CBS can be effectively employed to rule in over-reporting of cognitive symptoms, but owing to limited sensitivity and positive predictive power it cannot be reliably employed to rule out such exaggeration.

The current study is not without limitations. First, the sample was predominantly male, with higher levels of education and as such, the current results cannot necessarily be generalized to females and those with lower educational attainment. Next, as with the Gaasedelen et al. (2019) study, the criterion groups were constructed primarily by memory based PVTs (e.g., MSVT and NV-MSVT) and alternative PVTs based on attention, executive functioning, and/or processing speed paradigms could generate different results. Although further research into the CBS is indicated, the current data validate the CBS and provide initial support for the scale as a measure of cognitive over-reporting in a military sample.

References

Armistead-Jehle, P., & Hansen, C. L. (2011). Comparison of the repeatable battery for the assessment of neuropsychological status effort index and stand-alone symptom validity tests in a military sample. *Archives of Clinical Neuropsychology*, *26*, 592–601.

Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form) manual for administration, scoring, and interpretation*. Minneapolis: The University of Minnesota Press.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H. et al. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy and planning committee. *Archives of Clinical Neuropsychology*, *20*, 419–426.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2: Manual for administration and scoring (Rev)*. Minneapolis: University of Minnesota Press.

Carone, D. A. (2009). Test review of the medical symptom validity test2009. *Applied Neuropsychology*, *16*, 309–311.

Gaasedelen, O. J., Whiteside, D. M., Altmaier, E., Welch, C., & Basso, M. R. (2019). *The construction and initial validation of the Cognitive Bias Scale for the Personality Assessment Inventory*. The Clinical Neuropsychologist.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a response bias scale for the MMPI-2. *Assessment*, *14*, 196–208.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual*. Edmonton, Canada: Green's Publishing.

Green, P. (2008). *Manual for the Nonverbal Medical Symptom Validity Test*. Edmonton, Canada: Green's Publishing.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. L., Millis, S. R., & Conference Participants (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093–1129.

Kurtz, J. E., & Blais, M. A. (2007). Introduction to the special issue on the personality assessment inventory. *Journal of Personality Assessment*, *88*, 1–4.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 1–7.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of north American professions. *The Clinical Neuropsychologist*, *29*, 741–776.

Morey, L. C. (1991). *Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources, Inc.

Randolph, C. (1998). *Repeatable battery for the assessment of neuropsychological status manual*. San Antonio, TX: The Psychological Corporation.

Silverberg, N. D., Wertheimer, J. C., & Fitchtenberg, N. L. (2007). An effort index for the repeatable battery for the assessment of neuropsychological status (RBANS). *The Clinical Neuropsychology*, *2007*(*21*), 841–854.

Wager, J. G., & Howe, L. L. S. (2010). Nonverbal medical symptom validity test: Try faking now! *Applied Neuropsychology*, *17*, 305–309.