




RBANS Validity Indices: a Systematic Review and Meta-Analysis

Robert D. Shura^{1,2,3}  · Timothy W. Brearly^{1,3,4,5} · Jared A. Rowland^{1,3,5,6} · Sarah L. Martindale^{1,3,5} · Holly M. Miskey^{1,2,3} · Kevin Duff⁷

Received: 8 November 2017 / Accepted: 1 May 2018 / Published online: 16 May 2018

© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2018

Abstract

Neuropsychology practice organizations have highlighted the need for thorough evaluation of performance validity as part of the neuropsychological assessment process. Embedded validity indices are derived from existing measures and expand the scope of validity assessment. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) is a brief instrument that quickly allows a clinician to assess a variety of cognitive domains. The RBANS also contains multiple embedded validity indicators. The purpose of this study was to synthesize the utility of those indicators to assess performance validity. A systematic search was completed, resulting in 11 studies for synthesis and 10 for meta-analysis. Data were synthesized on four indices and three subtests across samples of civilians, service members, and veterans. Sufficient data for meta-analysis were only available for the Effort Index, and related analyses indicated optimal cutoff scores of ≥ 1 (AUC = .86) and ≥ 3 (AUC = .85). However, outliers and heterogeneity were present indicating the importance of age and evaluation context. Overall, embedded validity indicators have shown adequate diagnostic accuracy across a variety of populations. Recommendations for interpreting these measures and future studies are provided.

Keywords RBANS · Embedded validity indicators · Effort · Systematic review · Meta-analysis

An integral part of any neurocognitive evaluation is the assessment of performance validity, which refers to the accuracy of a respondent's performance. This is important because assessment interpretation rests on the assumption that an

individual has performed to the best of his or her ability. Assessment of performance validity is considered medically necessary to, and essential for, gold-standard neuropsychological assessment (Bush et al. 2005; Heilbronner et al. 2009). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Rabin et al. 2005; Randolph et al. 1998) is a commonly used brief battery of cognitive function, and a number of validity indicators such as the Effort Scale and the Effort Index have been identified (Novitski et al. 2012; Silverberg et al. 2007). For clinicians and researchers using the RBANS, the presence of an established embedded validity measure is ideal as additional stand-alone measures may not be needed, reducing the cost of test protocols and administration time. If additional validity measures are employed in a given evaluation, an embedded RBANS validity measure allows for a more thorough assessment of validity across the scope of the overall battery. The primary aim of this study was to review and synthesize the body of diagnostic accuracy research on the various RBANS validity indices in order to establish the global utility of these measures to function as indicators of invalid response.

The RBANS was primarily developed as a cognitive screening instrument for geriatric patients evaluated for suspected dementia but has been expanded to evaluate a wider

✉ Robert D. Shura
Robert.Shura2@va.gov

¹ VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center (MIRECC), Salisbury Veterans Affairs Health Care System (SVAHCS), Salisbury, NC 28144, USA

² Mental Health & Behavioral Sciences Service Line, SVAHCS, Salisbury, NC 28144, USA

³ Department of Psychiatry and Behavioral Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27103, USA

⁴ Neuropsychology Assessment, Walter Reed National Military Medical Center, Bethesda, MD 20889, USA

⁵ Research & Academic Affairs Service Line, SVAHCS, Salisbury, NC 28144, USA

⁶ Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA

⁷ Center for Alzheimer's Care, Imaging, & Research, Department of Neurology, University of Utah, Salt Lake City, UT 84108, USA

age range and multiple neurological and psychiatric conditions (Randolph et al. 1998). The measure consists of 12 subtests that contribute to a Total score and Index scores across five domains: Immediate Memory, Visuospatial-Constructional, Language, Attention, and Delayed Memory. Index scores are converted to age-based standard scores ($M = 100$, $SD = 15$). The original normative sample age range was 12 to 89 years, though additional age, education, gender, and race norms were published for the elderly (Duff et al. 2003; Duff and Ramezani 2015; Patton et al. 2003). There are four equivalent forms to allow for serial assessment, and the test is versatile, taking approximately 30 min to administer with minimal materials required (Randolph et al. 1998). Thus, the RBANS fills a niche as an assessment instrument falling between brief dementia screeners and full neuropsychological assessment, further diversifying the assessment abilities of clinical psychologists and expanding the repertoire of repeatable batteries available to neuropsychologists.

In addition to neuropsychological screening for neurodegenerative processes in the elderly, the RBANS can serve as a more general screening instrument of neurocognition (Randolph et al. 1998). Utility of the RBANS has been assessed in a variety of clinical populations outside the context of dementia including individuals with traumatic brain injury (TBI), multiple sclerosis, Parkinson's disease, Huntington's disease, hepatitis C virus, schizophrenia, and bipolar disorder, though research is limited on using the RBANS in these populations (Beatty 2004; Dickerson et al. 2004; Duff et al. 2010; Gold et al. 1999; Gogos et al. 2010; Lippa et al. 2013; McKay et al. 2007; McKay et al. 2008; Umaki and Denney 2013; Schoenberg et al. 2012; Wilk et al. 2004). Additional features such as multivariate base rates and a standardized regression-based change formula facilitate advanced interpretation of the RBANS (Crawford et al. 2012; Duff et al. 2004).

Numerous stand-alone performance validity tests created solely to assess accuracy of a respondent's performance on cognitive tests are available (see Sollman and Berry 2011). However, embedded validity measures derived from established neuropsychological measures are increasing in popularity. Embedded validity measures are sensitive to invalid performance as measured by stand-alone performance validity tests (PVTs) or known groups such as simulated malingerers. For example, Reliable Digit Span is derived from Wechsler Intelligence Scales Digit Span subtest which is a measure of working memory that has also been shown to contain a robust measure of performance validity (Schroeder et al. 2012). Embedded validity measures are convenient because they do not add time to an assessment, and a large number of embedded measures have been validated in the literature (Boone 2013). Clinicians are encouraged to employ multiple validity measures, with a focus on identifying cutoff scores with specificity estimates over 90, typically resulting in decreased sensitivity in comparison to stand-alone measures

(Bush et al. 2005). These multiple validity measures may be temporally spread throughout an evaluation and across different cognitive domains (Boone 2009). The reason for a given patient failing a validity measure is usually unknown, but when identified, such invalid presentations may lead to qualified conclusions regarding an individual's true cognitive ability. When sensitive validity measures are passed, the evaluator has a reliable basis to interpret the remaining scores as an accurate reflection of a respondent's true ability.

Embedded performance validity measures have been derived from various RBANS subtests and two have been researched in multiple studies. The Effort Index (EI) uses scores from the Digit Span and List Recognition subtests to evaluate performance validity (Silverberg et al. 2007). The diagnostic accuracy of the EI was initially evaluated by Silverberg and colleagues in 123 individuals across multiple groups, namely, patients with TBI, a control group, a clinical malingering group per Slick et al. (1999) criteria, and two simulation groups. Overall classification accuracy in this sample was highest using a cutoff score of $EI > 0$ (area under the curve [AUC] = .90; Silverberg et al. 2007). However, the EI has been shown to be affected by age, education, and level of cognitive dysfunction (Duff et al. 2011). A second embedded validity index, the Effort Scale (ES), also uses scores from the List Recognition and Digit Span subtests, but incorporates scores from the List Recall, Story Recall, and Figure Recall subtests (Novitski et al. 2012). In the initial ES study by Novitski and colleagues, the ES outperformed the EI in identifying mild TBI patients with identified poor effort ($n = 25$) compared to amnesic clinical patients ($n = 69$; EI AUC = .61; ES AUC = .91). Additional studies have emerged evaluating both of these indices in a variety of samples, though there is notable variability in recommended cutoff scores (Carter et al. 2016; Dunham et al. 2014; Lippa et al. 2017).

Given the popularity and versatility of the RBANS, as well as the standard of including performance validity indices in all cognitive assessments, the key question of this review was: what is the evidence that the embedded validity measures from the RBANS adequately assess performance validity? Our aim was to synthesize research on existing validity measures using systematic review and meta-analysis in order to make clinical and research recommendations. We focused on adult studies evaluating the diagnostic accuracy of RBANS embedded validity measures compared to known groups or stand-alone PVTs. For indices that could be evaluated with meta-analysis, a global meta-analysis identifying a general cutoff score was completed. This will be especially useful for readers working with mixed samples. A secondary aim was to synthesize diagnostic accuracy of RBANS validity measures for specific populations, namely civilians, service members, and veterans, as well for young versus older adult groups. Studies with heterogeneous populations suggest different cutoff scores for the Effort Index, indicating a need for subgroup analyses. The current

synthesis will help guide RBANS users to decide which cutoff score is most appropriate for a given respondent. Finally, this paper provides tables of all published papers on RBANS validity indices for clinicians and researchers, with data presented to aid readers in selecting specific studies for individual reference as needed for a given clinical or research situation.

Method

Systematic Review

Search Strategy and Screening This study was not registered and this is not an update to prior systematic reviews or meta-analyses. Study type and scope of interest were determined a priori and defined in the format of PICOS, presented in Table 1. Abstracts were screened, and studies included if they used the RBANS as part of the design, evaluated a performance validity indicator identified within the RBANS, were written in English, and were published in a peer-reviewed journal. Systematic searches were conducted on 04/20/2017 for terms related to the RBANS and performance validity. Table 2 lists the search strategy and results from EBSCO, including the PsycINFO, PsycARTICLES, and Academic Search Premier databases. Table 3 lists the search strategy and results from PubMed. There were no date restrictions used in the searches given the relatively recent emergence of embedded PVTs in the RBANS. Unpublished studies and dissertations were not included (for discussion, see van Driel et al. 2009; Vickers and Smith 2000). Search results were archived using EndNote version 7.5. Two authors (SLM and HMM) independently screened abstracts, with each author recording results in independent Excel files. Discrepancies ($n = 28$) were reviewed by a third author (RDS), with inclusion based on a retain rating by two of three authors.

Selection Full article review was conducted independently by two other authors (TWB and JAR). Review criteria for inclusion were experimental or cross-sectional design, group statistic calculated using a comparison group, report of

Table 2 Search strategy and results for EBSCO: PsycINFO, PsycARTICLES, and Academic Search Premier

Step	Terms	Hits
1	RBANS OR Repeatable Battery for the Assessment of Neuropsychological Status	1,314
2	effort OR malingering	721,394
3	symptom OR performance	2,483,492
4	validity	297,040
5	3 AND 4	58,614
6	2 OR 5	776,816
7	1 AND 6	164
8	Limit: academic journals	137
9	Limit: English	136
10	Duplicates removed	106

Search completed on 04/20/2017. RBANS = Repeatable Battery for the Assessment of Neuropsychological Status

diagnostic accuracy statistics (sensitivity, specificity, or area under the curve), and at least part of the sample being greater than age 18. Record reviews were recorded in independent Excel files. All initial reviews were rated equally, and a third rater was not needed to address discrepancies. Reference lists of the included articles were reviewed by RDS, leading to one additional record. Authors of identified records were contacted with the list of articles in an attempt to identify any additional records not discovered by the search strategies, but no additional records were identified. Table 4 lists excluded records and brief information on each, and Fig. 1 presents the PRISMA flow diagram. Once all included records were identified, data were extracted for synthesis and meta-analysis. We created a table for data extraction, which was populated by one author and checked for accuracy by a second author. Extracted data were compiled in Tables 5 and 6 and included sensitivity and specificity values needed for meta-analysis. Descriptive data from individual studies were also extracted to synthesize, specifically on sample age, population, comparison target for RBANS validity indices, and AUC statistics. Authors of included studies were contacted by e-mail in attempts to obtain missing data.

Table 1 PICOS

Parameter	Inclusion	Exclusion
Population	Adult	Subjects <18 years
Intervention	RBANS	Modified administrations
Comparison	Validated PVTs, Slick criteria, known clinical groups, simulated groups	No comparison
Outcomes	Diagnostic accuracy (AUC, sensitivity, specificity)	No defined outcomes
Study Design	Case-control, RCT, NRCT, cross-sectional	Case reports, editorials, opinions, dissertations/theses, reviews

RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; PVT = performance validity test; AUC = area under the curve; RCT = randomized controlled trial; NRCT = non-randomized controlled trial

Table 3 Search strategy and results for PubMed

Step	Terms	Hits
1	RBANS OR Repeatable Battery for the Assessment of Neuropsychological Status	420
2	effort OR malingering	111,484
3	symptom OR performance	934,582
4	validity	138,034
5	3 AND 4	19,428
6	2 OR 5	130,146
7	1 AND 6	39
9	Limit: English	39

Search completed on 04/20/2017. RBANS = Repeatable Battery for the Assessment of Neuropsychological Status

Meta-Analysis

Meta-Analytic Approach Random effects meta-analysis was conducted according to the procedures recommended by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Macaskill et al. 2010). Bivariate hierarchical summary receiver operating characteristics curve (HSROC) and heterogeneity analyses were conducted in SAS Studio using PROC NLMIXED and the MetaDas macro. Bivariate summary receiver operating characteristics curve (SROC) and publication bias analyses were completed using Stata 15 with the metandi and MIDAS packages. All analyses were conducted using the true positives, false positives, true negatives,

and false negatives for each study. Version 5.3 of Review Manager (RevMan 2014) was used to calculate these variables based on published group sizes, sensitivity, and specificity values as per Cochrane Collaboration guidance (Macaskill et al. 2010). The purpose of heterogeneity analyses is to test whether covariates such as age and sample characteristics are associated with variability in model results beyond what would be expected due to within-study sampling error (for a full explanation refer to Macaskill et al. 2010).

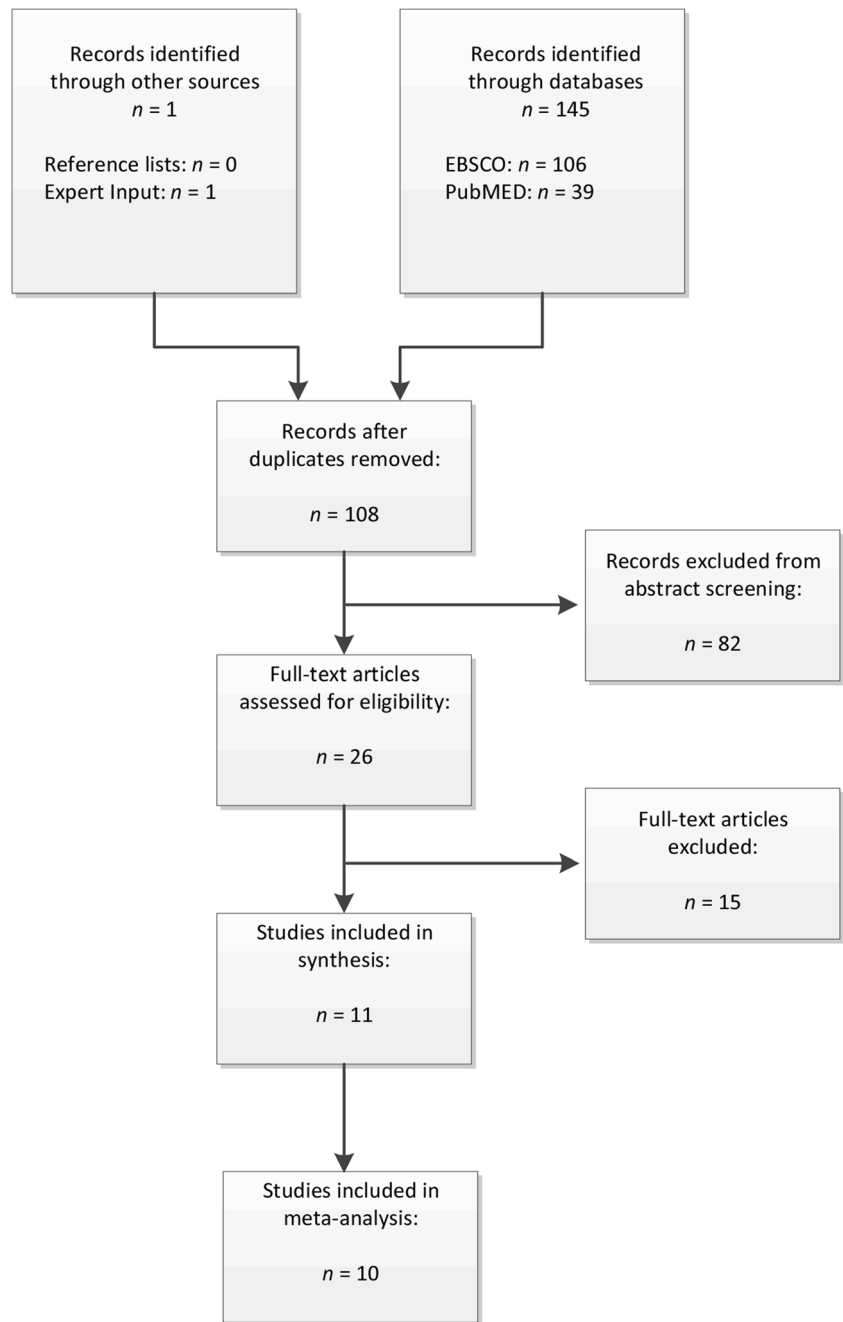
Planned analyses included calculation of an HSROC for each RBANS validity measure with adequate data for meta-analysis. HSROC analysis provides an overall summary of classification accuracy for a validity index and was used to generally evaluate the global diagnostic accuracy of a given validity measure, in line with the primary aim of this study. SROC analysis provides a summary of sensitivity and specificity for specific test cutoff scores and was completed for each measure cutoff with adequate studies available. A minimum of four studies was required to run any analysis due to statistical software limitations and concerns regarding model heterogeneity. For HSROC analysis, the cutoff score with the highest sensitivity for each respective study was included in the analysis to avoid double counting data from a single sample and over-representing studies that reported more cutoffs. Data provided by Barker et al. (2010) was excluded from any analyses containing data from Bortnik et al. (2013) or Paulson et al. (2015) due to sample overlap confirmed by study authors by e-mail communication. Novitski et al. (2012) was excluded from all analyses because data required for meta-analysis was not provided.

Table 4 Full text papers excluded

Authors	Year	Exclusion	Sample	Notes
Armistead-Jehle et al.	2012	No DA	M, T, OP	Compared MCI to various PVTs
Bayan et al.	2018	No DA (used estimate from prior study)	F, IP	RBANS EI in psychotic inpatients
Benitez et al.	2011	Did not assess RBANS embedded indices	V, T, OP	Depression and cognition
Burton et al.	2014	No DA	C, T, OP	EI/ES failure rates in dementia
Carter et al.	2016	No DA	C, T, OP	EI/ES failure rates in Parkinson's
Duff et al.	2011	No DA	C, mixed	EI failure rates in elderly
Hook et al.	2009	No DA	C, T, OP	EI failure rates in elderly
Lippa et al.	2014	No DA	C, T, IP	Validity in acute moderate/severe TBI
Moore et al.	2013	No DA to validity criterion	C, T, OP	EI to predict treatment in psychotic patients
Morra et al.	2014	Poster	unclear	EI in schizophrenia
Morra et al.	2015	No DA; appears same as poster	C, T, OP/IP	EI in schizophrenia
O'Mahar et al.	2011	No DA	C, R	Test-retest reliability of EI
Peck et al.	2014	Poster	C, T, IP	DA to TOMM with inpatients
Sieck et al.	2012	No DA	C, T, OP	Validity in Huntington's patients
Zimmer et al.	2017	No DA	V, T, OP	EI and other PVTs in mild TBI

For Sample: C = community; S = student; V = veteran; M = military; T = treatment (clinical) sample; F = forensic; R = research; OP = outpatient; IP = inpatient. No DA = no diagnostic accuracy (AUC, sensitivity, specificity) in methods; MCI = mild cognitive impairment; PVT = performance validity test; EI = Effort Index; ES = Effort Scale; TBI = traumatic brain injury; TOMM = Test of Memory Malingering

Fig. 1 PRISMA flow diagram



Investigation of Heterogeneity Visual inspection of data suggested heterogeneity among studies, even within the same cutoff. For instance, sensitivity values for EI listed in Table 5 ranged from .14 to .89 for $a \geq 4$ cutoff score. One possible explanation for this is that seven of the nine included EI studies were conducted in clinical samples, whereas Crighton et al. (2015) analyzed a forensic sample, and Dunham et al. (2014) utilized a college student simulation sample. Another possible explanation is that EI may not be an appropriate validity index in older samples due to increased incidences of neurocognitive dysfunction (e.g., Dunham et al. 2014). Finally, it was observed that the majority of included

studies examined active duty military and veteran samples, with the remaining three studies being performed in non-military contexts.

As per recommendations of the Cochrane Collaboration, HSROC analyses were re-run with a dichotomous covariate to account for each hypothesized source of heterogeneity, namely, clinical versus non-clinical samples, age ($M_{\text{age}} \leq 64$, $M_{\text{age}} \geq 65$), and civilian versus military (including veteran). There was no statistical evidence that non-clinical samples significantly affected model heterogeneity ($\chi^2 = 3.4$, $p = 0.18$). However, Dunham et al. (2014) appeared to be an outlier upon visual inspection. EI performance appeared to be less

Table 5 Extracted data of interest from included studies: Effort Index

Record	Age	<i>n</i>	Sample	Grouping	Outcome	Diagnostic accuracy	Notes
	<i>M</i>	Range				Cutoff	
Armistead-Jehle and Hansen 2011	34.0	22–46	85	M, T, OP	MSVT NV-MSVT TOMM	AUC = .74 [.59, .90]	SP .96 .92 .92 AUC based on MSVT failure
Barker et al. 2010	71.7	44–94	357	V, T, OP	MSVT NV-MSVT TOMM	AUC = .60 [.44, .77]	.24 .31 .30 .97 .97 .96
Bortnik et al. 2013	76.7	59–94	128	V, T, OP	TOMM plus behavioral indicators	AUC = .86 [.81, .91]	.93 .82 .64 .51 .42 .70
Crighton et al. 2015	42.5	19–67	105	F, C	Slick Criteria	$d = 0.64$ (ANOVA)	.50 .81 .85 .92 .96 .41
Dunham et al. 2014	76.4 ^a 27.8 ^b	52–89 22–61	90	C, S	Simulation plus MSVT	AUC = .84	.48 .43 .29 1.00
Jones 2016	31.4	--	462	M, T, OP	Slick Criteria	$d = 0.85$.97 .97 1.00
Lippa et al. 2017	28.5 ^c 27.9 ^d	--	250	M, T, OP	TOMM1		.84 .89 .94 .99 .99 1.00
Novitski et al. 2012	49.0 ^e 80.6 ^f	--	94	T, OP	mTBI fail WMT vs. MCI	AUC = .61	--
Paulson et al. 2015	68.2	41–87	239	V, T, OP	TOMM	AUC = .88 [.83, .94]	.93 .63
Silverberg et al. 2007	39.9 ^g --	17–68 18–48 ^h	47 76	T, S	mTBI, clinical malingering (Slick), simulators	AUC = .90 [.84, .96]	.93 ⁱ .96 ^j .86 ^k .78

Table 5 (continued)

Record	Age	<i>n</i>	Sample	Grouping	Outcome	Diagnostic accuracy	Notes
Young et al. 2012	54.3	10.4	--	85	V, T, F, OP	WMT	–
					≥ 2	.67 ⁱ .92 ^j .75 ^k	.81
					≥ 3	.67 ⁱ .79 ^j .50 ^k	.91
					≥ 4	.53 ⁱ .71 ^j .46 ^k	1.00
					≥ 3	.54	.81
					≥ 4	.31	.94
					≥ 5	.19	.94

For Sample: C = community; S = student; V = veteran; M = military; T = treatment (clinical) sample; F = forensic; OP = outpatient; IP = inpatient. MAL = malingering; EI = Effort Index; ES = Effort Score; mTBI = mild traumatic brain injury; AUC = area under the curve; WMT = Word Memory Test; MCI = mild cognitive impairment; TOMM = Test of Memory Malingering; TOMM1 = TOMM Trial 1; MSVT = Medical Symptom Validity Test; NV-MSVT = Non-Verbal Medical Symptom Validity Test

- ^a memory impairment group
- ^b coached simulators group
- ^c PVT pass
- ^d PVT fail
- ^e mTBI
- ^f amnesic MCI or probable Alzheimer's
- ^g mTBI groups
- ^h control and simulator groups
- ⁱ clinical malingerers (used for all meta-analyses; *n* = 15)
- ^j simulated naïve malingerers (*n* = 28)
- ^k simulated coached malingerers (*n* = 24)

Some studies include additional diagnostic accuracy data for other cutoff scores not included here

Table 6 Extracted data of interest from included studies: other indices

Record	Index	Age	<i>M</i>	<i>SD</i>	Range	<i>n</i>	Sample	Grouping	Outcome	Cutoff	SEN	SP
Barker et al. 2010	LR	71.7	9.3	44–94	357	V, T, OP	TOMM	AUC = .83 [.77, .89]	< 13	.38	.91	
	DS							AUC = .79 [.72, .85]	< 7	.42	.92	
Crighton et al. 2015	PN				105	F, C	Slick Criteria	AUC = .70 [.61, .79]	< 9	.42	.91	
	ES	42.5	10.7	19–67				<i>d</i> = 0.71 (ANOVA)	–	–	–	
Dunham et al. 2014	ES	76.4 ^a	10.5	52–89	90	C, S	Simulation plus MSVT	AUC = .92	< 12	.88	.81	
	ES	27.8 ^b	9.0	22–61				–	< 10	.53	.84	
Lippa et al. 2017	ES	28.5 ^c	7.9	--	250	M, T, OP	TOMM1	–	< 11	.53	.84	
	ES	27.9 ^d	6.6	--				–	< 12	.53	.84	
Novitski et al. 2012	ES	49.0 ^e	10.1	--	94	T, OP	mTBI fail WMT, vs. MCI	AUC = .91	< 12	–	–	
	ES	80.6 ^f	6.3	41–87	239	V, T, OP	TOMM	AUC = .71 [.65, .77]	< 12	.42	.71	
Paulson et al. 2015	ES	68.2	9.9					AUC = .90 [.85, .94]	< 42	.82	.77	
	PVI							AUC = .94 [.91, .97]	< 24	.84	.90	
Silverberg et al. 2007	DS	39.9 ^e	--	17–68	123	T, S	mTBI, clinical malingering	AUC = .84 [.77, .92]	–	–	–	
	LR						(Slick), simulators	AUC = .91 [.85, .97]	–	–	–	
Young et al. 2012	LR	54.3	10.4	--	85	V, T, F, OP	WMT	<i>d</i> = 1.28 (0.73–1.83)	< 15	.33	.97	
								(<i>t</i> test)	< 16	.43	.90	
									< 17	.52	.90	

LR = List Recognition; DS = Digit Span; PN = Picture Naming; ES = Effort Scale; PVI = Performance Validity Index; CRIER = Charleston Revised Index of Effort for the RBANS. For Sample: C = community; S = student; V = veteran; M = military; T = treatment (clinical) sample; F = forensic; OP = outpatient; IP = inpatient. TOMM = Test of Memory Malingering; TOMM1 = TOMM Trial 1; MSVT = Medical Symptom Validity Test; WMT = Word Memory Test; MCI = mild clinical impairment; mTBI = mild traumatic brain injury; AUC = area under the curve

^a memory impairment group

^b coached simulators group

^c PVT pass

^d PVT fail

^e mTBI

^f amnesic MCI or probable Alzheimer's

^g mTBI groups

Some studies include additional diagnostic accuracy data not included here for other cutoff scores

consistent in older samples ($\chi^2 = 9.0, p = 0.01$), and there was significant variation noted between civilian and military samples ($\chi^2 = 13.9, p < .001$).

Results

Systematic Review

Effort Index Eleven papers met inclusion criteria following screening and full text review, which are presented in Table 5. All 11 studies evaluated the EI, four using civilian samples, four using veteran samples, and three using military samples. The initial EI paper (Silverberg et al. 2007) is divided into two experiments. The first focused on development of the EI in comparison to the Digit Span and List Recognition subtests. The second experiment used a hybrid civilian sample including a clinical mild TBI group, a clinical malingering group (defined by Slick Criteria; Slick et al. 1999), three groups of undergraduate students for a control, and two simulated malingering groups (coached and naïve). Diagnostic accuracy was best using a cutoff score of $EI > 0$ for all three malingering groups compared to the mild TBI group. Strengths of this study included use of both mild TBI and control groups, three different malingering groups, and inclusion of additional information on the Digit Span and List Recognition subtests. One limitation is the small subsample size of the five groups, ranging from 15 to 32.

Novitski et al. (2012) used a sample of mild TBI patients who failed the Word Memory Test compared to a geriatric sample with cognitive impairment, resulting in the lowest AUC value of the civilian papers. In contrast, Dunham et al. (2014) used a sample of psychology graduate student simulated malingerers compared to a community geriatric sample comprised of those with memory impairment, resulting in a higher AUC score, but with unacceptable specificity using the cutoff of $EI > 3$. Additionally, the study evaluated both EI and ES specificities across various levels of cognitive impairment using the RBANS Total Score. EI specificity decreased with increasing cognitive impairment, though the ES showed an opposing trend. Finally, Crighton et al. (2015) used a forensic sample with groups based on Slick Criteria. There was a medium effect size of group on EI scores, though diagnostic accuracy was somewhat poorer across different cutoff scores than other studies. In summary, the EI showed AUC scores ranging from .61 to .90 at various cutoff scores in civilian samples. The civilian studies ranged in age from 19 to 89 years.

Three studies used military outpatient treatment samples to evaluate the EI. Armistead-Jehle and Hansen (2011) found a cutoff score of $EI \geq 1$ produced the best diagnostic accuracy using failure on three separate stand-alone PVTs (with Medical Symptom Validity Test used for AUC calculation)

in an outpatient clinical sample of service members, most with a history of mild TBI. Similarly, Jones (2016) used a clinical sample, most of whom had a history of mild TBI, and found malingering groups defined by Slick Criteria to have a large effect on EI. A cutoff score of $EI \geq 1$ demonstrated the best accuracy. In the third study using mild TBI patients (Lippa et al. 2017), a cutoff score of $EI \geq 1$ also resulted in acceptable accuracy. In general, the EI is useful as an adjunct PVT in military treatment samples presenting with a history of mild TBI. Only Armistead-Jehle et al. (2012) reported AUC values, which were .74 and .60 using cutoff scores of 1 and 4, respectively. As expected, the military samples were young compared to the civilian samples, with mean ages ranging from 28 to 34 years.

In veteran samples, a hybrid treatment-forensic sample by Young et al. (2012), accuracy was commensurate with other samples based on predicting failure on the Word Memory Test. A cutoff score of $EI \geq 3$ showed the best sensitivity (.54), but a cutoff of $EI \geq 4$ was needed to improve specificity to the $> .90$ range, at the expense of sensitivity (.31). Three other studies were published by the same research group, with overlapping clinical veteran samples. In the earliest study, Barker et al. (2010) found a cutoff score of $EI \geq 4$ best at predicting Test of Memory Malingering failure in a clinical sample of veterans with a mean age of 72 years. Also using the $EI \geq 4$ cutoff score, Bortnik et al. (2013) found much poorer specificity using a dementia sample overlapping with Barker et al. (2010). Unique to that study was the evaluation of specificities for EI across four different types of dementia, with specificity range spanning .47 to .78. In the final veteran study, Paulson et al. (2015) found a cutoff score of $EI \geq 2$ adequate when predicting Test of Memory Malingering performance, but specificity was poor. Two studies reported AUC values of .86 and .88. In contrast to the military studies, veteran studies tended to use more conservative cutoff scores for EI, and to focus more on dementia rather than mild TBI. Veteran samples also tended to use older participants than military studies, with a range of 41 to 94 years in the referenced studies.

Other Indices All of the studies that evaluated validity measures other than the EI ($n = 8$; see Table 6) continued to include the EI in the analyses, typically as a basis of comparison for the newer indices. The ES (Novitski et al. 2012) was evaluated in five of those studies. In the initial study to validate the ES, the authors compared the ES in middle-aged mild TBI patients with invalid Word Memory Test profiles to elderly patients with cognitive impairment. A cutoff score of $ES < 12$ improved diagnostic accuracy over the EI. Dunham et al. (2014) also noted improved accuracy over the EI using an ES cutoff score of < 12 in their simulation study. In the third civilian study of ES, Crighton et al. (2015) found a greater effect size for ES than EI when comparing forensic groups based on Slick Criteria. In all three of these civilian samples, the ES

outperformed the EI, with two studies reporting AUCs of .91 and .92.

Only one study evaluated ES in a military sample (Lippa et al. 2017). Although the ES held acceptable sensitivity, no cutoff score achieved specificity of $\geq .90$. However, at the same specificity level (.84), $EI \geq 1$ had a higher sensitivity (.58) than $ES < 14$ (.54). One veteran study (Paulson et al. 2015) also evaluated ES. Like the military sample, ES in the veteran sample performed poorer at $ES < 12$ compared to $EI \geq 2$. The Paulson paper was the only one to report an AUC, which was .71. In general, the ES outperformed the EI in civilian samples, but not in military or veteran samples.

Two other composite indices have been established, but only evaluated in a single study. Paulson et al. (2015) evaluated both the Performance Validity Index (PVI) and the Charleston Revised Index of Effort for the RBANS (CRIER) compared to the EI and ES in an outpatient veteran sample. Both new measures outperformed the EI and ES, with the CRIER showing the best diagnostic accuracy using a cutoff score of $CRIER < 24$. AUCs for the PVI and CRIER were .90 and .94 respectively. Although the indices show promise, replication will be needed before considering clinical use.

In addition to composite indices, three RBANS subtests were evaluated in isolation with regard to utility as validity measures. Three studies evaluated the List Recognition subtest. In one, List Recognition was significantly poorer in veterans who failed the Word Memory Test compared to those who passed, at large effect size, $d = 1.28$ (Young et al. 2012). A study using a veteran sample resulted in an AUC of .83 using List Recognition (Barker et al. 2010), though a higher AUC of .91 was found in a civilian sample (Silverberg et al. 2007). Barker et al. (2010) and Silverberg et al. (2007) also evaluated the Digit Span subtest, resulting in relatively similar AUCs of .79 and .84, respectively. Finally, Picture Naming was analyzed in Barker et al. (2010), resulting in an AUC of .70, the lowest of the three subtests evaluated.

Meta-Analytic Results

Due to the limited number of studies available, analysis was only feasible for EI. HSROC analysis was not conducted for ES due to insufficient variation in cutoff scores (all studies except for one provided data at the < 12 cutoff). Tests for publication bias were conducted for each analyzed EI cutoff using the process recommended by Deeks et al. (2005). Regression analysis of funnel plot symmetry shown in Fig. 2 did not indicate the presence of publication bias, as study outcome was not associated with study sample size. Specifically, effective sample size, which accounts for between group variance in target group size relative to total study sample, and study diagnostic odds ratio (DOR), which is the ratio of true to false positives identified, were not related. However, the Dunham et al. (2014) study was an outlier and

was the only study that utilized a simulated invalid performance group. For this reason, an additional HSROC analysis was conducted with this study excluded. In all ROC figures presented in Fig. 3, studies are plotted as circles sized relative to the sample n . Dotted lines around ROC curves indicate the 95% prediction region, a visual representation of the area within which future study results would be expected. Because the HSROC curve is estimated using data from variable cutoffs, no summary sensitivity and specificity is calculated. However, this type of analysis provides an overall estimate of diagnostic accuracy and power represented by the diagnostic odds ratio and AUC. Each HSROC plot in Fig. 3 represents a separate summary analysis for EI as per the results of heterogeneity analyses, namely, all EI studies, all EI studies with Dunham et al. (2014) excluded, studies with a mean age below 65, studies with a mean age below 65 with Dunham et al. (2014) excluded, and studies from military-veteran samples.

The summary diagnostic odds ratio for EI across samples ($n = 9$) was 9.94 [5.74, 17.21] with an AUC of .83 [.76, .88]. After excluding the simulation study ($n = 8$), the summary diagnostic odds ratio was 10.57 [5.87, 19.06] with an AUC of .83 [.77, .88]. In samples with a mean age under 65, the summary diagnostic odds ratio was 9.70 [5.33, 17.66] with an AUC of .82 [.76, .88] ($n = 7$). After excluding the simulation study, the summary diagnostic odds ratio for younger samples was 9.87 [5.35, 18.21] with an AUC of .83 [.76, .88] ($n = 6$). The summary diagnostic odds ratio for military and veteran samples across age groups was 11.03 [5.77, 21.10] with an AUC of .84 [.77, .89] ($n = 6$). Results of the bivariate analysis of all EI cutoff scores with four or more available studies are presented below, along with their relevant subgroup analyses: ≥ 1 , ≥ 2 , ≥ 3 , ≥ 4 , and ≥ 5 (Fig. 4). The ≥ 4 cutoff analyses were both conducted with and without Dunham et al. (2014) simulation study where applicable. In each figure, the summary point is indicated by a solid square, with individual studies presented as circles relative to sample size surrounding the summary SROC curve. The dashed line within the 95% prediction region indicates the 95% confidence interval of the summary point. The summary sensitivity and specificity along with the diagnostic odds ratio and AUC for each cutoff are listed in Table 7. Additional subgroup analyses according to sample characteristics associated with increased model heterogeneity (where sufficient data was available) are also presented in Table 7.

Overall, meta-analysis of current studies indicated that EI is effective for the detection of invalid performance on the RBANS, correctly identifying approximately 10 invalid performances for every one misclassification (increasing to an 11:1 ratio for military and veteran samples). Table 7 suggests that a cutoff score of ≥ 1 appears best suited to identifying invalid performance, achieving the highest possible sensitivity without letting specificity fall below .90. However, Fig. 4

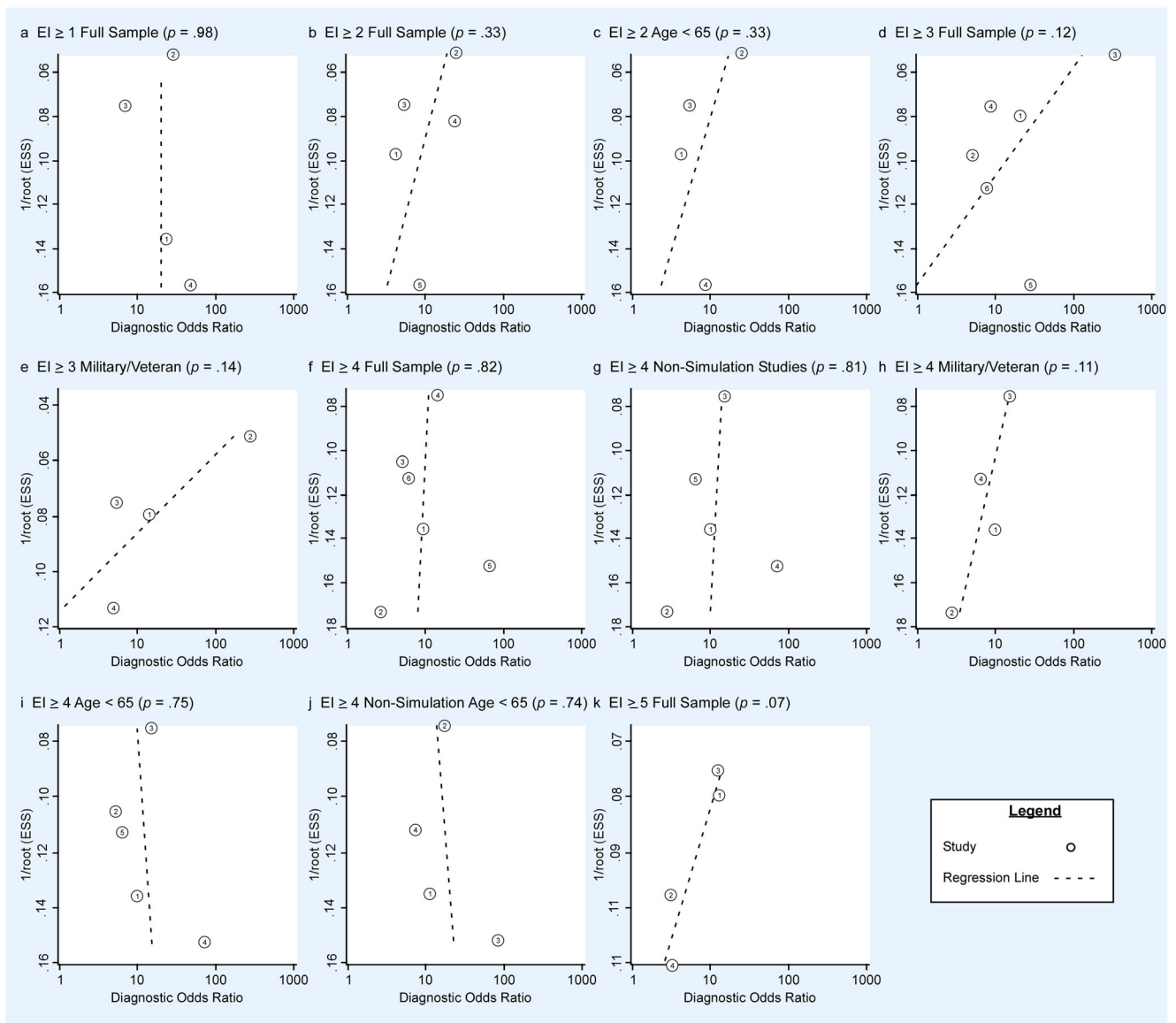


Fig. 2 Deeks’ funnel plot asymmetry tests for various cutoff scores of the Effort Index (EI). Full sample studies include the simulation study, non-simulation exclude that study but are otherwise the same. Plots are also

presented for cutoff scores using only studies with military and veteran samples, and for studies with samples less than 65 years old, when at least 4 studies were available in those categories

indicates the presence of an outlier study (Silverberg et al. 2007) with sensitivity and specificity values outside of the prediction region for this cutoff, suggesting that it is biasing the summary results with findings that may not be reproduced in future studies. This could not be empirically tested due to the limited number of studies available at this cutoff. The ≥ 2 cutoff indicated good accuracy across studies, however, this was unexpectedly driven by an elderly sample, with AUC values falling in the unacceptable range when older samples were excluded from the analysis. Both of these cutoffs show promise and may be appropriate for some settings, as described in systematic review results.

Cutoff scores of ≥ 3 and ≥ 4 performed relatively well, with ≥ 3 performing best in military samples. A cutoff of ≥ 3 also

provided the most stable and reliable results across samples with specificity appropriate for a validity test (i.e., $> .90$) and adequate sensitivity. Higher cutoffs ($\geq 4, \geq 5$) resulted in the expected relative decreases in sensitivity with marginal increases in specificity, and this effect was more pronounced at the ≥ 4 cutoff when simulation studies were excluded.

Discussion

The primary aim of this study was to evaluate the utility of RBANS embedded validity indices for detecting invalid performance. The Effort Index (EI) was the most researched index, and the only index for which meta-analysis could be

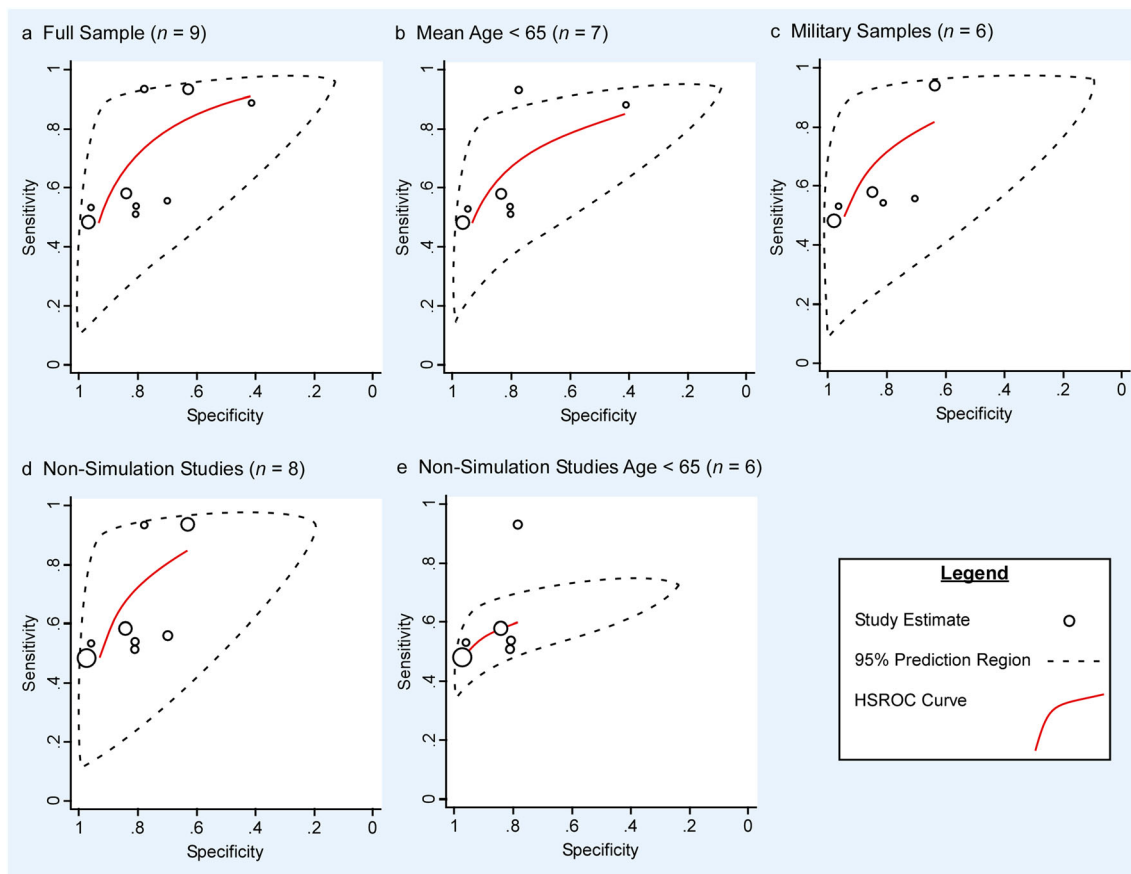


Fig. 3 Bivariate hierarchical summary receiver operating characteristics curve (HSROC) plots based on heterogeneity analyses of the Effort Index (EI). In all HSROC figures studies are plotted as circles sized relative to

the study sample size. Dotted lines around ROC curves indicate the 95% prediction region

completed. The EI correctly identified approximately 10 invalid performances for every one misclassification, increasing to 11 for military and veteran samples. The total AUC across all samples was .83 (CI = .76–.88), which is comparable to many stand-alone PVTs. These results were consistent regardless of whether or not a simulator study (Dunham et al. 2014) was included in analyses. Cutoffs of ≥ 1 and ≥ 3 performed best from a purely quantitative perspective, with AUCs of .86 and .85 respectively. Of note, all cutoff scores had AUC values of .79 or greater. For comparison, one study evaluating diagnostic accuracy to a group identified as probable malingering neurocognitive dysfunction found AUCs ranging from 0.88 to .90 on the Word Memory Test effort subtests and .79 to .84 on the Test of Memory Malingering trials, which are two of the most commonly used stand-alone PVTs (Fazio et al. 2015). Our AUCs from meta-analysis are comparable to those stand-alone test results, highlighting the overall efficacy of the EI more broadly. For use in heterogeneous samples, such as making determinations for excluding invalid responders from research studies, the choice to use a cutoff score of 1 or 3 will likely depend on how conservative one wishes to be in making such decisions.

The 11 included EI studies spanned settings (clinical and forensic), populations (civilian, military, and veteran), designs (known groups, simulation studies, and groups based on validated PVTs), and age ranges (19 to 80+). Samples included healthy controls, cognitively impaired groups, mild TBI, and various psychiatric presentations. Most available studies focused on validating the EI in younger samples, suggesting a need for further research to more firmly establish guidelines for the use of EI with older populations. For service members, younger age groups, and mild TBI samples, a qualitative review of published research concluded that a cutoff score of $EI \geq 1$ is most appropriate. Although meta-analysis did not contradict this conclusion, it did suggest that a more conservative cutoff score of $EI \geq 3$ may be best for elderly and military or veteran samples. Silverberg et al. (2007), the first study to propose the EI, was a relative outlier in SROC analyses. The study was the only one to use Slick Criteria within a non-military and non-forensic sample suggesting that findings may have been sample-specific and indicating the need for additional research across subpopulations where the RBANS is used. In the meantime, Table 5 can be consulted to identify which individual study might be a best reference for guiding

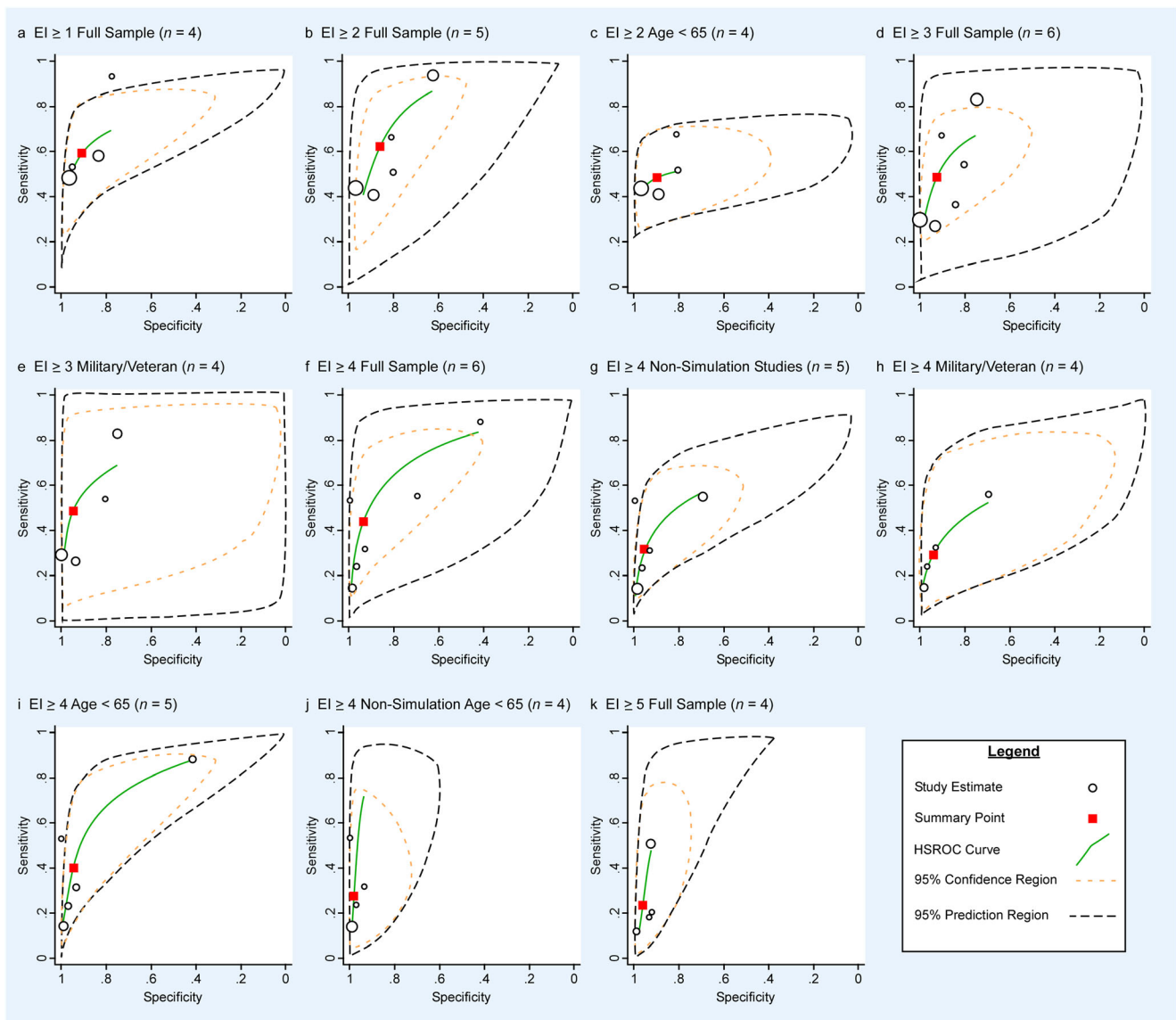


Fig. 4 Bivariate summary receiver operating characteristics curve (SROC) plots, corresponding to bias funnel plots presented in Fig. 2 for various Effort Index (EI) cutoff scores. The summary point is indicated by a solid square, with individual studies presented as circles relative to

sample size surrounding the summary SROC curve. The dashed line within the 95% prediction region indicates the 95% confidence interval of the summary point

classification decisions and EI cutoff scores for specific samples or patient characteristics.

The ES was created to address some of the limitations of the EI, particularly possible high false positive rates in elderly samples with frank memory impairment (Novitski et al. 2012). Meta-analysis could not be completed on the ES due to limitations in available research. Qualitatively, in older civilian samples, the ES outperformed the EI on diagnostic accuracy using a cutoff score of $ES < 12$. However, in the two studies that used military and veteran samples, the ES was poorer than the EI. The fewer number of studies present on the ES, and the variable results in comparison to the EI, indicate that this index should be considered as a secondary

measure to the EI pending additional research. Two other embedded indices were found (the PVI and CRIER) but were reported in only one paper. In that study (Paulson et al. 2015), both measures outperformed the EI and ES in a sample of veterans, with the CRIER performing best at an AUC of 0.94 compared to failure on the Test of Memory Malingering. However, additional studies will be necessary to further validate the utility and generalizability of those measures. Additionally, a number of studies also evaluated various single-subtest embedded measures, notably List Recognition, Digit Span, and Picture Naming subtests, though the subtests typically did not perform as well as the composite scores.

Table 7 Effort Index Summary
ROC analysis results

Cutoff score	Studies	Sensitivity [95% CI]	Specificity [95% CI]	DOR [95% CI]	AUC [95% CI]
EI ≥ 1	4	.59 [.47, .70]	.91 [.80, .96]	14.72 [7.57, 28.64]	.86 [.80, .91]
EI ≥ 2	5	.62 [.40, .80]	.86 [.73, .94]	10.32 [5.37, 19.81]	.83 [.76, .89]
Age < 65	4	.48 [.40, .55]	.89 [.79, .95]	7.88 [3.68, 16.88]	.80 [.71, .87]
EI ≥ 3	6	.48 [.31, .66]	.93 [.78, .98]	12.08 [3.98, 36.70]	.85 [.72, .92]
Military	4	.48 [.26, .71]	.95 [.68, .99]	17.42 [3.13, 97.10]	.88 [.68, .96]
EI ≥ 4	6	.44 [.22, .70]	.92 [.73, .98]	11.73 [2.98, 46.17]	.84 [.68, .94]
No Simulators	5	.32 [.19, .48]	.95 [.85, .99]	9.80 [3.72, 22.78]	.82 [.71, .90]
Age < 65	5	.40 [.17, .69]	.95 [.77, .99]	11.83 [5.13, 27.27]	.84 [.75, .91]
Age < 65 No Simulators	4	.27 [.16, .42]	.98 [.95, .99]	17.16 [6.37, 46.29]	.88 [.78, .94]
Military*	4	.28 [.15, .47]	.95 [.80, .99]	7.08 [2.51, 19.99]	.79 [.65, .89]
EI ≥ 5	4	.23 [.12, .40]	.96 [.93, .98]	7.76 [4.02, 14.97]	.80 [.72, .86]

* Contains a study with $M_{age} \geq 65$. EI = Effort Index; DOR = Diagnostic Odds Ratio

Based on these results, the following algorithm might be used when considering RBANS validity indices. For patients who are younger than 65 and without severe neurological impairments (e.g., severe TBI, acute stroke, advanced Huntington's), using a conservative EI cutoff of $EI \geq 1$ should suffice in the identification of possible invalid responding. For those who fail, conclusions regarding cognitive ability might be deferred or further evaluation incorporating stand-alone validity measures might be considered, and comments regarding true cognitive deficits delayed until that time. In this situation, other RBANS scores within the average range might be interpreted to reflect a minimal level of performance. For those over 65 or with likely true memory deficits (based on history, presentation, collateral information, etc.), interpreting the ES in the context of a failed EI might be considered. Clinicians and researchers using the RBANS should not rely on a single embedded measure cutoff score. Other information such as inconsistencies in reported history, knowledge of known brain-behavior relationships, and profile of performance across RBANS indices and subtests may also be helpful in informing decisions regarding validity of cognitive test performance.

There were some limitations to this study. First, there were a relatively small number of available studies, resulting in limited power to investigate underlying sources of heterogeneity (e.g., diagnostic group) and the effect of outliers. Heterogeneity analyses in meta-analysis of diagnostic accuracy studies with a small number of studies are underpowered and important characteristics contributing to heterogeneity may not be identified. Although the importance of accounting for differences in military and older samples were identified, this does not rule out the possibility that other sample specific characteristics may have contributed to model variability. Too few studies evaluated each EI cutoff score, limiting the ability to use bivariate approaches. For the ES, many studies only

reported performance at the established cutoff score ($ES < 12$), which did not allow for meta-analysis. Additionally, there was significant heterogeneity across studies. Further research is needed to more firmly establish cutoff scores for embedded validity measures within the RBANS and expand the use of such across additional populations, most notably geriatric populations. Research validating the promising CRIER and PVI indices and comparing them with the EI and ES would be a significant contribution. Finally, external validation of related RBANS cutoff scores is needed (e.g., related to aspects of daily functioning).

Compliance with Ethical Standards

Conflict of Interest There are no conflicts of interest to disclose.

Financial Disclosure This work was supported by resources of Salisbury Veterans Affairs Health Care System; the Mid-Atlantic Mental Illness Research Education and Clinical Center; and the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Mental Illness Research and Treatment.

Disclaimer The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs, the Department of Defense, or the U.S. Government.

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(7), 592–601. <https://doi.org/10.1093/arclin/acr049>.
- Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012). Memory complaints inventory and symptom validity test performance in a clinical sample. *Archives of Clinical Neuropsychology*, 27(7), 725–734.

- Barker, M. D., Horner, M. D., & Bachman, D. L. (2010). Embedded indices of effort in the repeatable battery for the assessment of neuropsychological status (Rbans) in a geriatric sample. *The Clinical Neuropsychologist*, 24(6), 1064–1077. <https://doi.org/10.1080/13854046.2010.486009>.
- Bayan, S. M., Nitch, S., Kinney, D., & Kaminetskaya, M. (2018). Exploratory analysis of CVLT-II forced-choice recognition and RBANS EI base rates for forensically-committed, psychotic-disordered inpatients: Implications on clinical decision-making. *Applied Neuropsychology: Adult*, 25(1), 71–81. <https://doi.org/10.1080/23279095.2016.1247092>.
- Beatty, W. W. (2004). RBANS analysis of verbal memory in multiple sclerosis. *Archives of Clinical Neuropsychology*, 19(6), 825–834. <https://doi.org/10.1016/j.acn.2003.12.001>.
- Benitez, A., Horner, M. D., & Bachman, D. (2011). Intact cognition in depressed elderly veterans providing adequate effort. *Archives of Clinical Neuropsychology*, 26(3), 184–193.
- Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23(4), 729–741. <https://doi.org/10.1080/13854040802427803>.
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. New York: Guilford.
- Bortnik, K. E., Horner, M. D., & Bachman, D. L. (2013). Performance on standard indexes of effort among patients with dementia. *Applied Neuropsychology: Adult*, 20(4), 233–242. <https://doi.org/10.1080/09084282.2012.695757>.
- Burton, R. L., Enright, J., O'connell, M. E., Lanting, S., & Morgan, D. (2014). RBANS embedded measures of suboptimal effort in dementia: Effort scale has a lower failure rate than the effort index. *Archives of Clinical Neuropsychology*, 30(1), 1–6.
- 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 & planning committee. *Archives of Clinical Neuropsychology*, 20(4), 419–426. <https://doi.org/10.1016/j.acn.2005.02.002>.
- Carter, K. R., Scott, J. G., Adams, R. L., & Linck, J. (2016). Base rate comparison of suboptimal scores on the RBANS effort scale and effort index in Parkinson's disease. *The Clinical Neuropsychologist*, 30(7), 1118–1125. <https://doi.org/10.1080/13854046.2016.1206145>.
- Crawford, J. R., Garthwaite, P. H., Morrice, N., & Duff, K. (2012). Some supplementary methods for the analysis of the RBANS. *Psychological Assessment*, 24(2), 365–374. <https://doi.org/10.1037/a0025652>.
- Crighton, A. H., Wygant, D. B., Holt, K. R., & Granacher, R. P. (2015). Embedded effort scales in the repeatable battery for the assessment of neuropsychological status: Do they detect neurocognitive malingering? *Archives of Clinical Neuropsychology*, 30(3), 181–185.
- Deeks, J. J., Macaskill, P., & Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, 58(9), 882–893. <https://doi.org/10.1016/j.jclinepi.2005.01.016>.
- Dickerson, F., Boronow, J. J., Stallings, C., Origoni, A. E., Cole, S. K., & Yolken, R. H. (2004). Cognitive functioning in schizophrenia and bipolar disorder: Comparison of performance on the repeatable battery for the assessment of neuropsychological status. *Psychiatry Research*, 129(1), 45–53. <https://doi.org/10.1016/j.psychres.2004.07.002>.
- van Driel, M. L., De Sutter, A., De Maeseneer, J., & Christiaens, T. (2009). Searching for unpublished trials in Cochrane reviews may not be worth the effort. *Journal of Clinical Epidemiology*, 62(8), 838–844.e833. <https://doi.org/10.1016/j.jclinepi.2008.09.010>.
- Duff, K., & Ramezani, A. (2015). Regression-based normative formulae for the repeatable battery for the assessment of neuropsychological status for older adults. *Archives of Clinical Neuropsychology*, 30(7), 600–604. <https://doi.org/10.1093/arclin/acv052>.
- Duff, K., Patton, D., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2003). Age- and education-corrected independent normative data for the RBANS in a community dwelling elderly sample. *The Clinical Neuropsychologist*, 17(3), 351–366. <https://doi.org/10.1076/clin.17.3.351.18082>.
- Duff, K., Schoenberg, M. R., Patton, D., Mold, J., Scott, J. G., & Adams, R. L. (2004). Predicting change with the RBANS in a community dwelling elderly sample. *Journal of the International Neuropsychological Society*, 10(6), 828–834.
- Duff, K., Beglinger, L. J., Theriault, D., Allison, J., & Paulsen, J. S. (2010). Cognitive deficits in Huntington's disease on the repeatable battery for the assessment of neuropsychological status. *Journal of Clinical and Experimental Neuropsychology*, 32(3), 231–238. <https://doi.org/10.1080/13803390902926184>.
- Duff, K., Spering, C. C., O'Bryant, S. E., Beglinger, L. J., Moser, D. J., Bayless, J. D., et al. (2011). The RBANS effort index: Base rates in geriatric samples. *Applied Neuropsychology*, 18(1), 11–17. <https://doi.org/10.1080/09084282.2010.523354>.
- Dunham, K. J., Shadi, S., Sofko, C. A., Denney, R. L., & Calloway, J. (2014). Comparison of the repeatable battery for the assessment of neuropsychological status effort scale and effort index in a dementia sample. *Archives of Clinical Neuropsychology*, 29(7), 633–641. <https://doi.org/10.1093/arclin/acu042>.
- Fazio, R. L., Sanders, J. F., & Denney, R. L. (2015). Comparison of performance of the test of memory malingering and word memory test in a criminal forensic sample. *Archives of Clinical Neuropsychology*, 30(4), 293–301. <https://doi.org/10.1093/arclin/acv024>.
- Gogos, A., Joshua, N., & Rossell, S. L. (2010). Use of the repeatable battery for the assessment of neuropsychological status (RBANS) to investigate group and gender differences in schizophrenia and bipolar disorder. *The Australian and New Zealand Journal of Psychiatry*, 44(3), 220–229. <https://doi.org/10.3109/00048670903446882>.
- Gold, J. M., Queern, C., Iannone, V. N., & Buchanan, R. W. (1999). Repeatable battery for the assessment of neuropsychological status as a screening test in schizophrenia I: Sensitivity, reliability, and validity. *The American Journal of Psychiatry*, 156(12), 1944–1950. <https://doi.org/10.1176/ajp.156.12.1944>.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23(7), 1093–1129. <https://doi.org/10.1080/13854040903155063>.
- Hook, J. N., Marquine, M. J., & Hoelzle, J. B. (2009). Repeatable battery for the assessment of neuropsychological status effort index performance in a medically ill geriatric sample. *Archives of Clinical Neuropsychology*, 24(3), 231–235.
- Jones, A. (2016). Repeatable battery for the assessment of neuropsychological status: Effort index cutoff scores for psychometrically defined malingering groups in a military sample. *Archives of Clinical Neuropsychology*, 31(3), 273–283. <https://doi.org/10.1093/arclin/acw006>.
- Lippa, S. M., Hawes, S., Jokic, E., & Caroselli, J. S. (2013). Sensitivity of the RBANS to acute traumatic brain injury and length of post-traumatic amnesia. *Brain Injury*, 27(6), 689–695. <https://doi.org/10.3109/02699052.2013.771793>.
- Lippa, S. M., Agbayani, K. A., Hawes, S., Jokic, E., & Caroselli, J. S. (2014). Effort in acute traumatic brain injury: Considering more than pass/fail. *Rehabilitation Psychology*, 59(3), 306–312.
- Lippa, S. M., Lange, R. T., Bhagwat, A., & French, L. M. (2017). Clinical utility of embedded performance validity tests on the repeatable battery for the assessment of neuropsychological status (RBANS)

- following mild traumatic brain injury. *Appl Neuropsychol Adult*, 24(1), 73–80. <https://doi.org/10.1080/23279095.2015.1100617>.
- Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Analysing and Presenting Results. In J. Deeks, P. Bossuyt & C. Gatsonis (Eds.), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0* (pp. 1–61). The Cochrane Collaboration. Retrieved from <http://methods.cochrane.org/>
- McKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology*, 22(1), 91–98. <https://doi.org/10.1016/j.acn.2006.11.003>.
- McKay, C., Wertheimer, J. C., Fichtenberg, N. L., & Casey, J. E. (2008). The repeatable battery for the assessment of neuropsychological status (RBANS): Clinical utility in a traumatic brain injury sample. *The Clinical Neuropsychologist*, 22(2), 228–241. <https://doi.org/10.1080/13854040701260370>.
- Moore, R. C., Davine, T., Harmell, A. L., Cardenas, V., Palmer, B. W., & Mausbach, B. T. (2013). Using the repeatable battery for the assessment of neuropsychological status (RBANS) effort index to predict treatment group attendance in patients with schizophrenia. *Journal of the International Neuropsychological Society*, 19(2), 198–205.
- Morra, L. F., Gold, J. M., Ossenfort, K., & Strauss, G. (2014). C-78 predicting insufficient effort in schizophrenia using the repeated battery for the assessment of neuropsychological status effort index. *Archives of Clinical Neuropsychology*, 29(6), 602–603.
- Morra, L. F., Gold, J. M., Sullivan, S. K., & Strauss, G. P. (2015). Predictors of neuropsychological effort test performance in schizophrenia. *Schizophrenia Research*, 162(1), 205–210.
- Novitski, J., Steele, S., Karantzoulis, S., & Randolph, C. (2012). The repeatable battery for the assessment of neuropsychological status effort scale. *Archives of Clinical Neuropsychology*, 27(2), 190–195. <https://doi.org/10.1093/arclin/acr119>.
- O'mahar, K. M., Duff, K., Scott, J. G., Linck, J. F., Adams, R. L., & Mold, J. W. (2011). Brief report: The temporal stability of the repeatable battery for the assessment of neuropsychological status effort index in geriatric samples. *Archives of Clinical Neuropsychology*, 27(1), 114–118.
- Patton, D. E., Duff, K., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2003). Performance of cognitively normal African Americans on the RBANS in community dwelling older adults. *The Clinical Neuropsychologist*, 17(4), 515–530. <https://doi.org/10.1076/clin.17.4.515.27948>.
- Paulson, D., Horner, M. D., & Bachman, D. (2015). A comparison of four embedded validity indices for the RBANS in a memory disorders clinic. *Archives of Clinical Neuropsychology*, 30(3), 207–216. <https://doi.org/10.1093/arclin/acv009>.
- Peck, C., Masse, A., & Schroeder, R. (2014). C-86 the utility of the RBANS effort index in an inpatient psychiatric setting. *Archives of Clinical Neuropsychology*, 29(6), 605.
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Neuropsychology*, 20(1), 33–65. <https://doi.org/10.1016/j.acn.2004.02.005>.
- Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The repeatable battery for the assessment of neuropsychological status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology*, 20(3), 310–319. <https://doi.org/10.1076/jcen.20.3.310.823>.
- Review Manager (RevMan) [Computer program]. (2014). Version 5.3. Copenhagen: The Nordic Centre, The Cochrane Collaboration.
- Schoenberg, M. R., Rinehardt, E., Duff, K., Mattingly, M., Bharucha, K. J., & Scott, J. G. (2012). Assessing reliable change using the repeatable battery for the assessment of neuropsychological status (RBANS) for patients with Parkinson's disease undergoing deep brain stimulation (DBS) surgery. *The Clinical Neuropsychologist*, 26(2), 255–270. <https://doi.org/10.1080/13854046.2011.653587>.
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, 19(1), 21–30. <https://doi.org/10.1177/1073191111428764>.
- Sieck, B. C., Smith, M. M., Duff, K., Paulsen, J. S., & Beglinger, L. J. (2012). Symptom validity test performance in the Huntington disease clinic. *Archives of Clinical Neuropsychology*, 28(2), 135–143.
- Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). An effort index for the repeatable battery for the assessment of neuropsychological status (RBANS). *The Clinical Neuropsychologist*, 21(5), 841–854. <https://doi.org/10.1080/13854040600850958>.
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-y;ft545](https://doi.org/10.1076/1385-4046(199911)13:04;1-y;ft545).
- Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, 26(8), 774–789. <https://doi.org/10.1093/arclin/acr066>.
- Umaki, T. M., & Denney, R. L. (2013). Neurocognitive deficits associated with the hepatitis C virus among incarcerated men. *The Clinical Neuropsychologist*, 27(3), 426–436. <https://doi.org/10.1080/13854046.2012.758315>.
- Vickers, A. J., & Smith, C. (2000). Incorporating data from dissertations in systematic reviews. *International Journal of Technology Assessment in Health Care*, 16(2), 711–713.
- Wilk, C. M., Gold, J. M., Humber, K., Dickerson, F., Fenton, W. S., & Buchanan, R. W. (2004). Brief cognitive assessment in schizophrenia: Normative data for the repeatable battery for the assessment of neuropsychological status. *Schizophrenia Research*, 70(2–3), 175–186. <https://doi.org/10.1016/j.schres.2003.10.009>.
- Young, J. C., Baughman, B. C., & Roper, B. L. (2012). Validation of the repeatable battery for the assessment of neuropsychological status–effort index in a veteran sample. *The Clinical Neuropsychologist*, 26(4), 688–699. <https://doi.org/10.1080/13854046.2012.679624>.
- Zimmer, A., Heyanka, D., & Proctor-Weber, Z. (2017). Concordance validity of PVTs in a sample of veterans referred for mild TBI. *Applied Neuropsychology: Adult*, 1–10.