Journal of the International Neuropsychological Society (2016), **22**, 851–858.

doi:10.1017/S1355617716000746

BRIEF COMMUNICATION

CVLT-II Forced Choice Recognition Trial as an Embedded Validity Indicator: A Systematic Review of the Evidence

Eben S. Schwartz,[1] Laszlo Erdodi,[2] Nicholas Rodriguez,[2] Jyotsna J. Ghosh,[3] Joshua R. Curtain,[4] Laura A. Flashman,[3] AND Robert M. Roth[3]

[1]Neuroscience Center, Waukesha Memorial Hospital, Waukesha, Wisconsin
[2]Department of Psychology, University of Windsor, Windsor ON, Canada
[3]Neuropsychology Program, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire
[4]Department of Neurology and Neurological Sciences, Stanford Health Care, Stanford, California

(RECEIVED October 15, 2015; FINAL REVISION August 2, 2016; ACCEPTED August 8, 2016)

Abstract

Objectives: The Forced Choice Recognition (FCR) trial of the California Verbal Learning Test, 2nd edition, was designed as an embedded performance validity test (PVT). To our knowledge, this is the first systematic review of classification accuracy against reference PVTs. **Methods:** Results from peer-reviewed studies with FCR data published since 2002 encompassing a variety of clinical, research, and forensic samples were summarized, including 37 studies with FCR failure rates ($N = 7575$) and 17 with concordance rates with established PVTs ($N = 4432$). **Results:** All healthy controls scored >14 on FCR. On average, 16.9% of the entire sample scored ≤14, while 25.9% failed reference PVTs. Presence or absence of external incentives to appear impaired (as identified by researchers) resulted in different failure rates (13.6% *vs*. 3.5%), as did failing or passing reference PVTs (49.0% *vs*. 6.4%). FCR ≤14 produced an overall classification accuracy of 72%, demonstrating higher specificity (.93) than sensitivity (.50) to invalid performance. Failure rates increased with the severity of cognitive impairment. **Conclusions:** In the absence of serious neurocognitive disorder, FCR ≤14 is highly specific, but only moderately sensitive to invalid responding. Passing FCR does not rule *out* a non-credible presentation, but failing FCR rules it *in* with high accuracy. The heterogeneity in sample characteristics and reference PVTs, as well as the quality of the criterion measure across studies, is a major limitation of this review and the basic methodology of PVT research in general. (*JINS*, 2016, *22*, 851–858)

Keywords: Symptom evaluation, Validity of results, Review, Systematic, Memory deficits, Malingering, Neuropsychology

INTRODUCTION

Embedded performance validity tests (embedded PVTs) are a useful complement to stand-alone PVTs. They allow for a shorter evaluation while simultaneously providing information about performance validity and enable ongoing monitoring of test taking effort throughout the course of a testing session (Boone, 2013). By virtue of relying on established neuropsychological tests of core cognitive domains (e.g., attention, memory, processing speed) that also measure performance validity, clinicians can meet the multiple (and sometimes competing) demands of providing a comprehensive assessment of the patient's neurocognitive functioning, performing an objective evaluation of test taking effort, and keeping the test battery length within reason.

PVT cutoffs are optimized for specificity (true negative rate) to protect individuals from being falsely deemed invalid responders. Keeping false positive errors under 10% is a standard guideline for calibrating new instruments (Boone, 2013). Therefore, high specificity (around .90) is a fundamental requirement, and sensitivity (true positive rate) is the test parameter that varies across instruments.

Correspondence and reprint requests to: Eben S. Schwartz, Department of Neuroscience, ProHealth Waukesha Memorial Hospital, 721 American Avenue; Suite 406, Waukesha, WI 53188. E-mail: eben.schwartz@phci.org

California Verbal Learning Test, 2nd Edition, Forced Choice Recognition Trial

The most common ways to develop embedded PVTs include identifying a new indicator (e.g., reliable digit span; Greiffenstein, Baker, & Gola, 1994), a certain cutoff on an existing subtest (e.g., digit span age-corrected scaled score; Spencer et al., 2013), or a logistical regression using a combination of scores (Wolfe et al., 2010). In contrast, the authors of the California Verbal Learning Test – 2nd Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober; 2000) introduced a novel task (Forced Choice Recognition [FCR]), which is administered following a second delay, 10 min after the standard clinical instrument is completed. Adding a delay before FCR may create the perception of increased difficulty and, thus, elicit a lower performance in those prone to poor test taking effort (Tombaugh, 1996; Green, 2003).

Although Delis et al. refrained from endorsing a specific cutoff, they report a study by Connor, Drake, Bondi, and Delis (1997) that used an early version of the FCR task and produced an impressive combination of sensitivity (.80) and specificity (.97) using a cutoff score of ≤13. They also use the base rate of failure (BR_{Fail}) argument to calibrate FCR: >90% of healthy participants in the normative sample ($N = 1087$) obtained a perfect score, most of the remaining 5 to 8% scored 15, and none scored ≤13. Delis et al. suggested that since only a small percentage of the normative sample scored ≤14, the profiles of such individuals should be viewed with caution as they may be invalid. Of note, participants in the normative sample were not administered standalone PVTs, so the clinical meaning of lower than expected FCR scores is unclear. Although those with very poor overall performances on the CVLT-II were excluded because of presumed dementia or other impairment, the specific exclusion criteria are not provided in the manual. In addition, the percentage of healthy controls that score ≤14 is of less relevance for clinical practice than is the percentage of clinical groups that score ≤14.

Given the limited rationale for choosing ≤14 as the default FCR cutoff, and the variability of reference PVTs and the clinical populations used in cross-validation studies, there is a clear need for a systematic review of the BR_{Fail} across diagnostic categories and concordance rates with established PVTs. This study was designed to review the clinical literature on FCR and critically examine the cumulative evidence of its classification accuracy. A summary of the evidence base on FCR's signal detection profile relative to other PVTs would allow for an empirically based evaluation and clinical interpretation of FCR scores in clinical settings.

METHODS

Search Strategy

Three electronic databases (PubMed, PsycINFO and Scopus) were searched for peer-reviewed original empirical papers on FCR. A combination of six search terms ("CVLT forced choice", "California Verbal Learning Test forced choice", "California Verbal Learning Test effort", "California Verbal Learning Test traumatic brain injury assessment", "California Verbal Learning Test neuropsychology disease treatment" and "CVLT recognition long-term memory performance") was used to achieve a balance between breadth (i.e., include the largest number of papers possible) and depth (i.e., keep the search focused on papers most likely to contain the relevant information). The search was restricted to articles published since 2001, to automatically exclude studies conducted before the FCR trial became publicly available. A total of 329 articles were initially identified and reviewed by the first three authors. Only studies with available FCR raw score data (k = 37; $N = 7575$) were included in the final analyses. Those that included comparison to established PVTs administered to the entire sample were included in comparison analyses.

Data Analyses

BR_{Fail} was reported for FCR and reference PVTs. Sensitivity (true positive rate) and specificity (true negative rate) were computed for FCR against reference PVTs, along with likelihood ratios (LR). A positive LR (+LR) is an index of how much more likely an individual with a given condition (i.e., invalid performance) is to produce a positive test result (i.e., fail a PVT) compared to an individual without the condition (i.e., valid performance). Conversely, negative LR (–LR) is an index of how much less likely an individual with a given condition (i.e., invalid performance) is to produce a negative test result (i.e., pass a PVT) compared to an individual with the condition (i.e., invalid performance). Naturally, the higher the +LR and the lower the –LR, the more informative the test is.

RESULTS

Table 1 presents BR_{Fail} for FCRs ≤14. All 245 healthy controls across four studies scored above this cutoff (Clark et al., 2012; Eikeland, Ljøstad, Mygland, Herlofson, & Løhaugen, 2012; Macher & Earleywine, 2012; Silk-Eglit et al., 2014). Within studies, the presence and/or severity of neurological disease process had a dose-response relationship to BR_{Fail} on FCR, but 322 clinical patients who passed reference PVTs and another 104 who were unexamined for performance validity also had a zero BR_{Fail}. An additional 170 patients with a variety of neuropsychiatric conditions, such as focal frontal lesions (Baldo, Delis, Kramer, & Shimamura, 2002), amnestic mild cognitive impairment (Clark et al., 2012), lyme neuroborreliosis (Eikeland et al., 2012), and mixed clinical with memory impairment (Root, Robbins, Chang, & Van Gorp, 2006), also had zero BR_{Fail}.

BR_{Fail} on FCR covaried with the severity of neurocognitive impairment. Patients in the semi-acute stage of TBI with moderate injury severity were almost five times more likely to fail FCR than those with mild injury (Schiehser et al., 2011).

https://doi.org/10.1017/S1355617716000746 Published online by Cambridge University Press

Table 1. Base rates of failure (BR_{Fail}) for FCR ≤ 14 in the published literature ($N = 7575$)

Author	Year	Sample	*N*	BR_{Fail}	RR
Baldo et al.	2002	Focal frontal lesions	11	0.0%	-
Moore and Donders	2004	Moderate-severe TBI; EXC: litigation, pre-morbid PSY	112	1.8%	-
Root et al.	2006	Mixed clinical with memory impairment	25	0.0%	
		Forensic; passed TOMM or VIP (both not always administered)	27	7.4%	NA
		Forensic; failed TOMM or VIP (both not always administered)	25	52.0%	7.0
Jacobs and Donders	2007	All severity TBI; EXC: litigating or pre-morbid PSY	104	3.8%	-
Marshall and Happe	2007	Pre-existing intellectual disability (mild-moderate)	100	11.0%	-
Axelrod and Schutte	2010	Mixed clinical VA - passed MSVT	152	11.0%	
		Mixed clinical VA - failed MSVT (not dementia profile)	70	48.0%	4.4
		Mixed clinical VA - failed MSVT (dementia profile)	64	46.0%	4.2
Nelson et al.	2010	Veterans in a research context	75	2.7%	-
		Veterans in a forensic context	44	15.9%	-
Axelrod and Schutte	2011	Mixed clinical VA – passed TOMM	121	8.3%	9.4
		Mixed clinical VA – failed TOMM	32	78.1%	
Donders and Strong	2011	All severity TBI - passed WMT	66	0.0%	
		All severity TBI - failed WMT	24	37.5%	NA
Miller et al.	2011	Moderate-severe TBI; failed ≤1 of TOMM, MSVT, NV-MSVT	42	2.4%	25.6
		Unsuccessful Simulators; failed ≥2 of TOMM, MSVT, NV-MSVT	39	61.5%	
Schiehser et al.	2011	Noncombat military, semi-acute mild TBI	44	6.8%	-
		Noncombat military, semi-acute mild-moderate (or unclear) TBI	5	20.0%	-
		Noncombat military, semi-acute moderate TBI	22	31.8%	-
Schroeder and Marshall	2011	Nonpsychotic PSY; EXC: litigation, noncompliance, IQ < 80	178	2.0%	-
		Psychotic disorders; same EXC	108	8.0%	-
Schutte[a] et al.	2011	Mixed clinical VA - passed MSVT	110	13.6%	3.3
		Mixed clinical VA - failed MSVT	98	44.9%	
Clark et al.	2012	Healthy Controls	35	0.0%	-
		Amnestic mild cognitive impairment	18	0.0%	-
		Nonamnestic mild cognitive impairment	19	10.5%	-
		Mild Alzheimer's disease	16	12.5%	-
		Moderate Alzheimer's disease	21	71.4%	-
Denning	2012	Mixed clinical VA - passed MSVT	315	5.1%	7.8
		Mixed clinical VA - failed MSVT	148	39.9%	
Eikeland[b] et al.	2012	Lyme neuroborreliosis	50	0.0%	-
		Matched Healthy Controls	50	0.0%	-
Macher and Earleywine	2012	Undergraduate students	110	0.0%	-
Peleikis et al.	2012	Schizophrenia spectrum disorder	297	2.4%	-
Morse et al.	2013	Mixed clinical; EXC: dementia, age >60, IQ <70	52	5.8%	-
		Mixed forensic (litigation/disability claim); EXC: same	91	14.3%	-
Tarescavage et al.	2013	Non-TBI disability claimants	251	12.4%	-
Clark et al.	2014	Veterans in a research context	198	6.7%	-
Davis and Millis	2014	Mixed clinical (age 18–65)	154	9.1%	-
Erdodi, Kirsch, et al.	2014	Mixed clinical - passed WMT	146	4.8%	5.7
		Mixed clinical - failed WMT	85	27.1%	
Erdodi, Roth, et al.	2014	All severity TBI - passed WMT	60	4.7%	8.5
		All severity TBI - failed WMT	40	40.0%	
King et al.	2014	Mixed clinical/Research VA without TBI	279	3.2%	-
		Mixed clinical/Research VA with TBI	210	11.4%	-
Kulas et al.	2014	Mixed clinical VA	103	7.8%	-
Maksimovskiy et al.	2014	Mixed clinical VA - passed MSVT	339	2.4%	16.3
		Mixed clinical VA - failed MSVT	23	39.1%	
Proto et al.	2014	VA mild TBI	178	18.0%	-
Silk-Eglit et al.	2014	Undergraduate students	50	0.0%	-
Egeland[b] et al.	2015	Mixed clinical - passed TOMM	110	5.5%	7.3
		Mixed clinical - failed TOMM	20	40.0%	
Heyanka et al.	2015	VA mild TBI - passed TOMM/WMT	134/64	6.7/0.0%	10.3/NA
		VA mild TBI - failed TOMM/WMT	26/111	69.2/35.6%	
Jak et al.	2015	Veterans with mTBI – passed TOMM	305	6.9%	9.2
		Veterans with mTBI – failed TOMM	106	63.2%	

https://doi.org/10.1017/S1355617716000746 Published online by Cambridge University Press

Table 1. (*Continued*)

Author	Year	Sample	*N*	BR_{Fail}	RR
Orff et al.	2015	VA mild to moderate TBI - passed TOMM	140	10.0%	4.7
		VA mild to moderate TBI - failed TOMM	45	46.7%	
Shura et al.	2015	Mixed post-deployment VA - passed WMT	91	1.1%	10.1
		Mixed post-deployment VA - failed WMT	27	11.1%	
Sugarman[c] and Axelrod	2015	Mixed clinical VA - passed TOMM	505	8.3%	9.3
		Mixed clinical VA - failed TOMM	99	76.8%	
Sugarman[c] et al.	2015	Mixed clinical VA - passed MSVT	519	9.1%	5.3
		Mixed clinical VA - failed MSVT	217	47.9%	
Erdodi et al.	2016	Mixed clinical – passed FIT/RMT	181/138	7.2/0.0%	5.0/NA
		Mixed clinical – failed FIT/RMT	14/44	35.7/40.9	
		AVERAGE	-	-	8.9

Note. FCR = Forced Choice Recognition trial of the CVLT-II (cutoff, ≤14); RR = Relative risk associated with failing the reference PVT vs. FCR; TOMM = Test of Memory Malingering (standard cutoffs); VIP = Validity Indicator Profile (standard cutoffs); MSVT = Medical Symptom Validity Test (standard cutoffs); WMT = Green's Word Memory Test (standard cutoffs); NV-MSVT = Non-Verbal Medical Symptom Validity Test (standard cutoffs); FIT = Rey 15-item test (standard cutoffs); RMT = Recognition Memory Test – Words (standard cutoffs); EXC = exclusion criteria; PSY = psychiatric disorders; VA = Veteran's Affairs.
[a]Same pool of VA patients with a larger *N*.
[b]Norwegian version of the CVLT-II.
[c]Same pool of VA patients as in Sugarman and Axelrod (2015) and Sugarman et al. (2015).

BR_{Fail} was four times higher in patients with psychotic disorders compared to those with other psychiatric disorders (Schroeder & Marshall, 2011). Patients with moderate Alzheimer's disease were six times more likely to fail FCR than those in earlier stages of the disorder (Clark et al., 2012). Presence of TBI more than tripled the likelihood of FCR failure in a mixed clinical sample of veterans assessed in a research context (King et al., 2014). However, in the absence of data on reference PVTs, the clinical meaning of these findings is difficult to determine.

A combined sample of 1526 patients (excluding dementia and intellectual disability) across 14 studies with no external incentives to appear impaired (as identified by researchers) and no data on reference PVTs produced a weighted mean BR_{Fail} of 3.5%. Conversely, a combined sample of 386 patients (excluding dementia and intellectual disability) across three studies (Morse, Douglas-Newman, Mandel, & Swirsky-Sacchetti, 2013; Nelson et al., 2010; Tarescavage, Wygant, Gervais, & Ben-Porath, 2013) with incentive to appear impaired, but no data on reference PVTs, produced a weighted mean BR_{Fail} of 13.2%.

A combined clinical sample of 3417 patients across 19 studies who *passed* a reference PVT produced a weighted mean BR_{Fail} of 6.8%. Conversely, a combined clinical sample of 1245 patients across the same 19 studies who *failed* a reference PVT produced a weighted mean BR_{Fail} of 49.9%. The average within-study risk ratio associated with passing or failing a reference PVT was 8.5, suggesting that overall, examinees who failed the reference PVT were 8.5 times more likely to also fail FCR.

A subset of studies ($N = 4432$) reported enough data to compute the classification accuracy of FCR against stand-alone PVTs (Table 2). Although comparing parameters derived from different samples, BR_{Fail}, and reference PVTs inevitably increases method variance, a few trends are apparent. Overall, FCR had much higher specificity (.93) than sensitivity (.50) to invalid responding, and better positive (7.7) than negative (.54) likelihood ratios.

BR_{Fail} was also evaluated as a function of criterion measures. The weighted mean BR_{Fail} on reference PVTs was 26.6%, *versus* 17.3% on FCR. However, of the 328 patients across six studies who failed the Test of Memory Malingering (TOMM) at standard cutoffs, 65.6% failed FCR (Axelrod & Schutte, 2011; Egeland, Andersson, Sundseth, & Schanke, 2015; Heyanka et al., 2015; Jak et al., 2015; Orff et al., 2015; Sugarman & Axelrod, 2015). In contrast, only 30.9% of the 249 patients across four studies who failed the Word Memory Test (WMT) at standard cutoffs scored ≤14 on FCR (Donders & Strong, 2011; Erdodi, Kirsch, et al., 2014; Erdodi, Roth, et al, 2014; Heyanka et al., 2015; Shura, Miskey, Rowland, Yoash-Gantz, & Denning, 2015).

DISCUSSION

The FCR trial was introduced to the CVLT-II to screen for invalid responding. To our knowledge, this is the first systematic review of its ability to differentiate cognitive impairment from invalid responding in clinical, research and forensic samples, and classification accuracy against established PVTs. At FCR ≤14, BR_{Fail} was zero for all healthy controls, consistent with the very low BR_{Fail} (<1%) observed in the normative sample (Delis et al., 2000).

As expected, scores on other PVTs, incentive status and type of criterion measure were associated with BR_{Fail} on FCR. Those who failed reference PVTs were eight times more likely to also fail FCR. Similarly, those with external incentives to appear impaired (as identified by researchers) were four times more likely to fail FCR, suggesting that FCR

https://doi.org/10.1017/S1355617716000746 Published online by Cambridge University Press

Table 2. Sensitivity (SN), specificity (SP), positive (+LR) and negative likelihood ratios (–LR) of FCR ≤14 against reference PVTs (PVT_{Ref})

					BR_{Fail}		Classification accuracy			
Author	Year	*N*	Sample	PVT_{Ref}	PVT_{Ref}	FCR	SENS	SPEC	+LR	–LR
Axelrod[a] and Schutte	2010	286	Mixed clinical VA	MSVT	46.9%	27.9%	.47	.89	4.2	.60
Axelrod and Schutte	2011	153	Mixed clinical VA	TOMM	20.9%	22.9%	.78	.92	9.5	.24
Donders and Strong	2011	90	All severity TBI	WMT	26.7%	9.7%	.38	1.00	NA	.63
Schutte[a] et al.	2011	208	Mixed clinical VA	MSVT	47.1%	28.3%	.45	.86	3.3	.64
Denning	2012	463	Mixed clinical VA	MSVT	32.0%	16.2%	.40	.95	7.8	.63
Erdodi, Kirsch, et al.	2014	231	Mixed clinical	WMT	36.8%	13.0%	.27	.95	5.6	.77
Erdodi, Roth, et al.	2014	100	All severity TBI	WMT	40.0%	18.8%	.40	.95	8.0	.63
Maksimovskiy et al.	2014	362	Mixed clinical VA	MSVT	6.4%	4.7%	.39	.98	16.6	.62
Egeland et al.	2015	130	Mixed clinical	TOMM	15.4%	10.8%	.40	.95	7.3	.63
Heyanka et al.	2015	160	VA mild TBI	TOMM	16.3%	16.9%	.69	.93	10.3	.33
				WMT	53.3%	18.9%	.36	1.00	NA	.64
Jak et al.	2015	411	VA mild TBI	TOMM	25.8%	21.4%	.63	.93	9.2	.40
Orff et al.	2015	185	VA mild-mod TBI	TOMM	24.3%	18.9%	.47	.90	4.7	.59
Shura et al.	2015	118	Mixed VA	WMT	22.9%	3.4%	.11	.99	10.1	.90
Sugarman[b] and Axelrod	2015	604	Mixed clinical VA	TOMM	16.4%	19.5%	.77	.92	9.2	.25
Sugarman, et al.[b]	2015	736	Mixed clinical VA	MSVT	29.5%	20.5%	.48	.91	5.3	.57
Erdodi et al.	2016	195	Mixed clinical	FIT	7.2%	9.2%	.36	.93	5.0	.69
				RMT	24.2%	9.9%	.41	1.00	NA	.59
TOTAL		4432		Weighted average	26.6%	17.3%	.50	.93	7.7	.54

Note. BR_{Fail} = Base rate of failure (% of the sample scored below the cutoff); FCR = Forced Choice Recognition trial of the CVLT-II (cutoff, ≤14); MSVT = Medical Symptom Validity Test (standard cutoffs); TOMM = Test of Memory Malingering (standard cutoffs); WMT = Green's Word Memory Test (standard cutoffs); FIT = Rey 15-item test (standard cutoffs); RMT = Recognition Memory Test – Words (standard cutoffs); VA = Veteran's Affairs; Mod = moderate severity; SENS = sensitivity; SPEC = specificity; +LR = positive likelihood ratio; –LR = negative likelihood ratio.
[a]To be consistent with other studies, MSVT failures with dementia and non-dementia profiles were collapsed.
[b]Same pool of VA patients as in Sugarman and Axelrod (2015) and Sugarman et al. (2015).

is sensitive to poor test taking effort. In addition, those who failed the TOMM were more than twice as likely to fail FCR than were those who failed the WMT.

The heterogeneity of criterion measures is a major limitation of any attempt to summarize PVT research across studies using different instruments. Our review is no exception. Reporting omnibus FCR failure rates in reference to PVTs with different signal detection profiles further inflates method variance. Therefore, we avoided traditional meta-analytic techniques and single-number summaries such as AUC, because they often produce misleading conclusions (Hanczar et al., 2010; Hand, 2009; Lobo, Jiménez-Valverde, & Real, 2008). However, the large number of studies and sample sizes are expected to attenuate these measurement artifacts. Furthermore, the excessive between-studies variability mostly affects the stability of global parameter estimates (i.e., the BR_{Fail} in general, across *all studies and PVTs*), which is beyond the scope of the present study. The most relevant finding is that, overall, FCR produced lower BR_{Fail} (17.3%) than reference PVTs (26.6%). The fact that BR_{Fail} was almost double on WMT (37.4%) compared to TOMM (20.0%) is both a confound in computing the averaging classification accuracy for FCR and an important clinical observation well-supported in the research literature (Green, 2007).

Empirical evidence on FCR's classification accuracy suggests that the ≤14 cutoff is highly specific, but only moderately sensitive to invalid responding. In other words, scoring >14 on FCR does not rule *out* non-credible presentation as it misses half of the invalid response sets, but scoring below that cutoff rules it *in* with high accuracy, keeping false positive errors around 6%. This seemingly inescapable trade-off between sensitivity and specificity has been labeled the "Larrabee limit" (Erdodi, Kirsch, et al., 2014).

One limitation of most studies is that it is unclear how many of these failures represent marginal ("near-pass") versus more extreme failures. That information would be important because the performance of individuals who fail FCR at ≤13 or ≤12 (as in case studies reported by Binder, Spector, & Youngjohn, 2012; Yochim, Kane, Horning, & Pepin, 2010) could be better understood if the BR_{Fail} associated with a score that low in individuals with severe impairment and/or those who pass reference PVTs was known. Conversely, it has also been suggested that a less conservative cutoff (≤15) merits investigation (D. Delis, personal communication, May 2012; Erdodi, Kirsch, et al., 2014; Root et al., 2006). Other limitations included the fact that some studies did not group disease or injury by severity, which limits the meaning that can be drawn from those groups, and many studies administered only one PVT to group participants, which is not reflective of recommended (Heilbronner et al., 2009) or actual (Sweet, Benson, Nelson, & Moberg; 2015) practice.

Investigations of the relationship between FCR and other CVLT-II indices (such as recall, recognition, or executive type

https://doi.org/10.1017/S1355617716000746 Published online by Cambridge University Press

errors) would allow a more nuanced approach to interpreting FCR failures in groups with higher or lower BR_{Fail}. It is also worth noting that, while the CVLT-II normative sample excluded participants who performed poorly enough to suspect dementia or other significant cognitive impairment, they were not systematically screened with PVTs. That stated, nearly 1% of that group failed FCR, as did a variable proportion (0.0–13.6%) of the samples that passed reference PVTs, underscoring the poor negative predictive power of FCR.

Future research on FCR would benefit from reporting the frequency distribution for FCR, rather than group means, standard deviations or BR_{Fail}, as that would provide a better informed clinical interpretation of FCR scores, allow to explore alternative cutoffs, and establish an empirical basis for the stratification of marginal versus extreme failures (Bigler, 2012). First, the highly skewed distribution suggests that a cutoff-based interpretation (rather than treating FCR as a continuous scale) is the appropriate clinical interpretation of the instrument. Second, FCR violates the basic assumption of normality underlying most statistical tests, rendering some of the designs using regression analyses with FCR as a continuous variable of questionable validity. Finally, a systematic review of the classification accuracy of other PVTs embedded within the CVLT-II (e.g., Yes/No recognition hits) would be a valuable addition to the literature.

We found that non-credible examinees are eight times more likely than credible ones to fail FCR. However, non-credible examinees are only half as likely to pass FCR compared to credible examinees. Therefore, failing FCR is a strong predictor of invalid performance, but passing FCR is a weak predictor of valid performance. This signal detection profile is consistent with the test authors' description, who stated that FCR "is best suited for the detection of suboptimal effort in more blatant, unsophisticated exaggerators." (Delis et al., 2000, pp. 54–55). Overall, we interpret this burgeoning literature as providing support for the utility of FCR as a PVT in many applications, while recognizing that judgment about performance validity must be based on multiple validity indicators and incorporate findings from other sources of information such as clinical history, observation, and known patterns of neuropsychiatric dysfunction (Heilbronner et al., 2009).

ACKNOWLEDGMENTS

We are grateful to the authors of studies listed in Table 1 and 2 who provided additional data, which was done in compliance with the Helsinki Declaration. The authors report no conflicts of interest. This project was unfunded.

REFERENCES

Axelrod, B.N., & Schutte, C. (2010). Analysis of the dementia profile on the Medical Symptom Validity Test. *The Clinical Neuropsychologist*, *24*, 873–881.

Axelrod, B.N., & Schutte, C. (2011). Concurrent validity of three forced-choice measures of symptom validity. *Applied Neuropsychology*, *18*(1), 27–33.

Baldo, J.V., Delis, D., Kramer, J., & Shimamura, A. (2002). Memory performance on the California Verbal Learning Test-II: Findings from patients with focal frontal lesions. *Journal of the International Neuropsychological Society*, *8*, 539–546.

Bigler, E.D. (2012). Symptom validity testing, effort and neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 632–642.

Binder, L.M., Spector, J., & Youngjohn, J.R. (2012). Psychogenic stuttering and other acquired nonorganic speech and language abnormalities. *Archives of Clinical Neuropsychology*, *27*, 557–568.

Boone, K.B. (2013). *Clinical practice of forensic neuropsychology*. New York, NY: Guilford.

Clark, A.L., Amick, M.M., Fortier, C., Milberg, W.P., & McGlinchey, R.E. (2014). Poor performance validity predicts clinical characteristics and cognitive test performance of OEF/OIF/OND veterans in a research setting. *The Clinical Neuropsychologist*, *28*(5), 802–825.

Clark, L.R., Stricker, N.H., Libon, D.J., Delano-Wood, L., Salmon, D.P., Delis, D.C., & Bondi, M.W. (2012). Yes/No forced choice recognition memory in mild cognitive impairment and Alzheimer's Disease: Patterns of impairment and associations with dementia severity. *Clinical Neuropsychology*, *26*, 1201–1216.

Connor, D.J., Drake, A.I., Bondi, M.W., & Delis, D.C. (1997). *Detection of feigned cognitive impairments in patients with a history of mild to severe closed head injury*. Paper presented at the American Academy of Neurology, Boston.

Davis, J.J., & Millis, S.R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist*, *28*(2), 199–214.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B. (2000). *The California Verbal Learning Test-Second Edition*. San Antonio TX: The Psychological Corporation.

Denning, J.H. (2012). The efficiency and accuracy of the Test of Memory Malingering Trial 1, errors on the first 10 items of the Test of Memory Malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, *27*(4), 417–432.

Donders, J., & Strong, C.H. (2011). Embedded effort indicators on the California Verbal Learning Test – second edition (CVLT-II): An attempted cross-validation. *The Clinical Neuropsychologist*, *25*, 173–184.

Egeland, J., Andersson, S., Sundseth, Ø.Ø., & Schanke, A.K. (2015). Types or modes of malingering? A confirmatory factor analysis of performance and symptom validity tests. *Applied Neuropsychology. Adult*, *22*(3), 215–226.

Eikeland, R., Ljøstad, U., Mygland, A., Herlofson, K., & Løhaugen, G.C. (2012). European neuroborreliosis: Neuropsychological findings 30 months post-treatment. *European Journal of Neurology*, *19*, 480–487.

Erdodi, L.A., Abeare, C.A., Lichtenstein, J.D., Tyson, B.T., Kucharski, B., Zuccato, B.G., & Roth, R.M. (2016). WAIS-IV processing speed scores as measures of non-credible responding – The third generation of embedded performance validity indicators. *Psychological Assessment*. (Advance online publication.

Erdodi, L.A., Kirsch, N.L., Lajiness-O'Neill, R., Vingilis, E., & Medoff, B. (2014). Comparing the Recognition Memory Test and the Word Choice Test in a mixed clinical sample: Are they equivalent? *Psychological Injury and Law*, *7*(3), 255–263.

Erdodi, L.A., Roth, R.M., Kirsch, N.L., Lajiness-O'Neill, R., & Medoff, B. (2014). Aggregating validity indicators embedded in Conners' CPT-II outperforms individual cutoffs at separating valid from invalid performance in adults with traumatic brain injury. *Archives of Clinical Neuropsychology*, *29*, 456–466.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K.B., Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 50–77). New York, NY: Guilford Press.

Green, P. (2003). *Word Memory Test for Windows: User's manual and program*. Edmonton, AB: Green's Publishing.

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E.R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, *26*(6), 822–830.

Hand, D.J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123.

Heilbronner, R.L., Sweet, J.J., Morgan, J.E., Larrabee, G.J., 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.

Heyanka, D.J., Thaler, N.S., Linck, J.F., Pastorek, N.J., Miller, B., Romesser, J., … Sim, A.H. (2015). A factor analytic approach to the validation of the Word Memory Test and Test of Memory Malingering as measures of effort and not memory. *Archives of Clinical Neuropsychology*, *30*, 369–376.

Jacobs, M.L., & Donders, J. (2007). Criterion validity of the California Verbal Learning Test-Second Edition (CVLT-II) after traumatic brain injury. *Archives of Clinical Neuropsychology*, *22*, 143–149.

Jak, A.J., Gregory, A., Orff, H.J., Colón, C., Steele, N., Schiehser, D.M., … Twamley, E.W. (2015). Neuropsychological performance in treatment-seeking Operation Enduring Freedom/ Operation Iraqi Freedom Veterans with a history of mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *37*(4), 379–388.

King, P.R., Donnelly, K.T., Wade, M., Donnelly, J.P., Dunnam, M., Warner, G., … Alt, M. (2014). The relationships among premilitary vocational aptitude assessment, traumatic brain injury, and postdeployment cognitive functioning in combat veterans. *Archives of Clinical Neuropsychology*, *29*(4), 391–402.

Kulas, J.F., Axelrod, B.N., & Rinaldi, A.R. (2014). Cross-validation of supplemental Test of Memory Malingering scores as performance validity measures. *Psychological Injury and Law*, *7*(3), 236–244.

Lobo, J.M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151.

Macher, R.B., & Earleywine, M. (2012). Enhancing neuropsychological performance in chronic cannabis users: The role of motivation. *Journal of Clinical and Experimental Neuropsychology*, *34*, 405–415.

Maksimovskiy, A.L., McGlinchey, R.E., Fortier, C.B., Salat, D.H., Milberg, W.P., & Oscar-Berman, M. (2014). White matter and cognitive changes in veterans diagnosed with alcoholism and PTSD. *Journal of Alcoholism and Drug Dependence*, *2*(1), 144.

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *The Clinical Neuropsychologist*, *21*, 826–840.

Miller, J.B., Millis, S.R., Rapport, L.J., Bashem, J.R., Hanks, R.A., & Axelrod, B.N. (2011). Detection of insufficient effort using advanced clinical solutions for the Wechsler Memory Scale, fourth edition. *The Clinical Neuropsychologist*, *25*, 160–172.

Moore, B.A., & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *18*, 975–984.

Morse, C.L., Douglas-Newman, K., Mandel, S., & Swirsky-Sacchetti, T. (2013). Utility of the Rey-15 recognition trial to detect invalid performance in a forensic neuropsychological sample. *The Clinical Neuropsychologist*, *27*, 1395–1407.

Nelson, N.W., Hoelzle, J.B., McGuire, K.A., Ferrier-Auerbach, A. G., Charlesworth, M.J., & Sponheim, S.R. (2010). Evaluation context impacts neuropsychological performance of OEF/OIF veterans with reported combat-related concussion. *Archives of Clinical Neuropsychology*, *25*, 713–723.

Orff, H.J., Jak, A.J., Gregory, A.M., Colón, C.C., Schiehser, D.M., Drummond, S.P., … Twamley, E.W. (2015). Sleep disturbance, psychiatric, and cognitive functioning in veterans with mild to moderate traumatic brain injury. *Journal of Sleep Disorders: Treatment and Care*, *4*(2), 1–6.

Peleikis, D.E., Varga, M., Sundet, K., Lorentzen, S., Agartz, I., & Andreassen, O.A. (2012). Schizophrenia patients with and without Post-traumatic Stress Disorder (PTSD) have different mood symptom levels but same cognitive functioning. *Acta Psychiatrica Scandinavica*, *127*(6), 455–463.

Proto, D.A., Pastorek, N.J., Miller, B.I., Romesser, J.M., Sim, A.H., & Linck, J.F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury-related postconcussive symptoms. *Archives of Clinical Neuropsychology*, *29*, 614–624.

Root, J.C., Robbins, R.N., Chang, L., & Van Gorp, W.G. (2006). Detection of inadequate effort on the California Verbal Learning Test- Second Edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society*, *12*, 688–696.

Schiehser, D.M., Delis, D.C., Filoteo, J.V., Delano-Wood, L., Han, S.D., Jak, A.J., … Bondi, M.W. (2011). Are self-reported symptoms of executive dysfunction associated with objective executive function performance following mild to moderate traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology*, *33*(6), 704–714.

Schroeder, R.W., & Marshall, P.S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, *25*, 437–453.

Schutte, D., Millis, S., Axelrod, B., & VanDyke, S. (2011). Derivation of a composite measure of embedded symptom validity indices. *The Clinical Neuropsychologist*, *25*, 454–462.

Shura, R.D., Miskey, H.M., Rowland, J.A., Yoash-Gantz, R.E., & Denning, J.H. (2015). Embedded performance validity measures with postdeployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology. Adult*, *23*, 94–104.

Silk-Eglit, G.M., Stenclik, J.H., Gavett, B.E., Adams, J.W., Lynch, J.K., & Mccaffrey, R.J. (2014). Base rate of performance invalidity among non-clinical undergraduate research participants. *Archives of Clinical Neuropsychology*, *29*(5), 415–421.

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

Sugarman, M.A., & Axelrod, B.N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology. Adult*, *22*(2), 141–146.

Sugarman, M.A., Holcomb, E.M., Axelrod, B.N., Meyers, J.E., & Liethen, P.C. (2015). Embedded measures of performance validity in the Rey complex figure test in a clinical sample of veterans. *Applied Neuropsychology. Adult*, *23*, 105–114.

Sweet, J.J., Benson, L.M., Nelson, N.W., & Moberg, P.J. (2015). The American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA Division 40) 2015 TCN professional practice and 'salary survey': Professional practices, beliefs, and incomes of US neuropsychologists. *The Clinical Neuropsychologist*, *29*(8), 1069–1162.

Tarescavage, A.M., Wygant, D.B., Gervais, R.O., & Ben-Porath, Y. S. (2013). Association between the MMPI-2 Restructured Form (MMPI-2-RF) and malingered neurocognitive dysfunction among non-head injury disability claimants. *The Clinical Neuropsychologist*, *27*(2), 313–335.

Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM)*. New York: Multi-Health Systems, Inc.

Wolfe, P.L., Millis, S.R., Hanks, R., Fichtenberg, N., Larrabee, G.J., & Sweet, J.J. (2010). Effort indicators within the California verbal learning test-II (CVLT-II). *The Clinical Neuropsychologist*, *24*(1), 153–168.

Yochim, B.P., Kane, K.D., Horning, S., & Pepin, R. (2010). Malingering or expected deficits? A case of herpes simplex encephalitis. *Neurocase*, *16*, 451–460.