




Flipping the Script: Measuring Both Performance Validity and Cognitive Ability with the Forced Choice Recognition Trial of the RCFT

Kaitlyn Abeare¹,
Kristoffer Romero¹ ,
Laura Cutler¹ ,
Christina D. Sirianni², and
Laszlo A. Erdodi¹ 

Perceptual and Motor Skills
2021, Vol. 128(4) 1373–1408
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00315125211019704
journals.sagepub.com/home/pms



Abstract

In this study we attempted to replicate the classification accuracy of the newly introduced Forced Choice Recognition trial (FCR) of the Rey Complex Figure Test (RCFT) in a clinical sample. We administered the RCFT_{FCR} and the earlier Yes/No Recognition trial from the RCFT to 52 clinically referred patients as part of a comprehensive neuropsychological test battery and incentivized a separate control group of 83 university students to perform well on these measures. We then computed the classification accuracies of both measures against criterion performance validity tests (PVTs) and compared results between the two samples. At previously published validity cutoffs (≤ 16 & ≤ 17), the RCFT_{FCR} remained specific (.84–1.00) to psychometrically defined non-credible responding. Simultaneously, the RCFT_{FCR} was more sensitive to examinees' natural variability in visual-perceptual and

¹Department of Psychology, University of Windsor, Windsor, Ontario, Canada

²School of Social Work, University of Windsor, Windsor, Ontario, Canada

Corresponding Author:

Laszlo A. Erdodi, 168 Chrysler Hall South, 401 Sunset Ave, Windsor, Ontario N9B 3P4, Canada.

Email: lerdodi@gmail.com

verbal memory skills than the Yes/No Recognition trial. Even after being reduced to a seven-point scale (18-24) by the validity cutoffs, both RCFT recognition scores continued to provide clinically useful information on visual memory. This is the first study to validate the RCFT_{FCR} as a PVT in a clinical sample. Our data also support its use for measuring cognitive ability. Replication studies with more diverse samples and different criterion measures are still needed before large-scale clinical application of this scale.

Keywords

Rey Complex Figure Test, forced choice recognition, performance validity, embedded validity indicators

Introduction

The validity of clinical decisions based on neuropsychological test scores hinges on the assumption that examinees gave maximal (or at least typical) cognitive effort during testing (Bigler, 2015; Dandachi-FitzGerald et al., 2016; Merten & Merckelbach, 2013; Roor et al., 2016). Historically, clinicians assumed valid performance by default, and verified it only through behavioral observations (e.g., through observed level of cooperation, apparent ability and willingness to follow instructions). However, the limitations of clinical judgment for detecting non-credible responding have long been demonstrated empirically (Dandachi-FitzGerald et al., 2017; Heaton et al., 1978), and this realization has led to the development and use of objective measures for differentiating valid from invalid performances. Early performance validity tests (PVTs) were free-standing instruments, often based on the forced choice recognition (FCR) paradigm. With the advent of managed care, assessors faced pressure to abbreviate test batteries without compromising test data quality. As expansive, multi-trial free-standing PVTs became harder to justify to third party medical insurers, they were gradually replaced by embedded validity indicators (EVIs) that served the PVT function from within traditional neuropsychological ability tests.

Although EVIs have tended to exhibit inferior signal detection relative to traditional PVTs (Lau et al., 2017) and they have sometimes been criticized for conflating ability and effort (Bigler, 2012; Leighton et al., 2014), their advantages have included (a) cost-effectiveness; (b) reduced mental stamina demands for young or medically/emotionally fragile examinees (Lichtenstein et al., 2017); and (d) an inconspicuousness that made them more difficult for examinees to detect (An et al., 2019; Berger et al., 2019) and, therefore, more resistant to coaching (Brennan et al., 2009; Erdal, 2004; Lippa, 2018; Weinborn et al., 2012). Of equal importance, EVIs protect assessors from the appearance of a

confirmation bias when PVT use seems to have been motivated by clinician expectations of examinee malingering (Boone, 2013).

In recent decades EVI research has proliferated. A popular method of EVI development has been to add a FCR trial to existing memory tests, such as the California Verbal Learning Test (Delis et al., 2000), California Verbal Learning Test – Children’s Version (Lichtenstein et al., 2017; 2018); Rey Auditory Verbal Learning Test (Poreh et al., 2016), and Hopkins Verbal Learning Test (Abeare, Hurtubise et al., 2020; Cutler et al., 2021). Following this trend, Rai et al. (2019) introduced an FCR trial to the Rey Complex Figure Test (RCFT); this RCFT_{FCR} was placed 10 minutes after the end of the standard administration protocol.

The original RCFT (Rey, 1941), in conjunction with its add-on trials and scores (Meyers & Meyers, 1995; Lu et al., 2003; Reedy et al., 2013), offers clinicians a range of tools for evaluating the visual-perceptual and memory skills of adults and children. The paper and pencil copying task (i.e., the reproduction of the complex figure while the stimulus remains exposed to the examinee) can detect impairments in visuospatial construction, fine-motor coordination, and planning/organization skills. Clinically significant variations in performance may signal deficits in encoding, storage, and retrieval processes (Shin et al., 2006).

Meyers and Meyers (1995) first introduced a Yes/No Recognition trial using 24 single line drawings consisting of 12 targets and 12 foils. For this task, examinees are instructed to circle the shapes that they recognize as being part of the original figure. The examinee’s raw score is the number of correct decisions made [i.e., the sum of true positives (targets circled) and true negatives (foils not circled)], ranging from 0 to 24. The Yes/No Recognition trial proved useful as an EVI (Shura et al., 2016; Whiteside et al., 2011). Lu et al. (2003) developed an equation that provided a single-number estimate of the validity of the response set based on multiple RCFT scores, and this equation was subsequently cross-validated (Blaskewitz et al., 2009; Reedy et al., 2013; Shura et al., 2016).

For the newer scale, the RCFT_{FCR}, the initial validation study was, based on the experimental malingering paradigm (Rai et al., 2019). The RCFT_{FCR} had comparable classification accuracy to the Yes/No Recognition trial, despite its different administration format. Instead of the Yes/No Recognition format of presenting examinees with a number of shapes scattered on a page, the RCFT_{FCR} consists of 24 items, each containing a target and a foil. Within each pair, the examinee is asked to *identify the one that was part of the original figure*. As such, the score ranges from 0 (foil chosen every time) to 24 (target chosen every time). (Complimentary digital copies of the Rey Complex Figure and the RCFT_{FCR} trial are available to qualified readers through the senior author.)

The RCFT_{FCR} was specifically developed to appear difficult to the examinee. Unlike most FCR trials that present the examinee with an easy choice between a familiar target and a fairly obvious intrusion error as the alternative option, the RCFT_{FCR} includes a number of items that are genuinely challenging, for two reasons. First, given the growing awareness that a test based on FCR is likely to be a PVT – or a trap to avoid – the authors made the RCFT_{FCR} harder to identify as an EVI and thus, harder for sophisticated malingerers to evade. Second, the authors expressed hope that, following a non-traditional EVI presentation, this FCR might double as an actual measure of cognitive abilities (i.e., perceptual skills and visual memory). To achieve these goals, they engineered the stimulus properties (i.e., discriminability) of the target and foil to increase the cognitive demands of the task (i.e., both items seem comparably plausible at first glance). In the original study with no clinical sample, the only circumstantial evidence supporting this feature was a finding of equivalent mean scores from healthy participants when comparing the Yes/No Recognition and RCFT_{FCR} trials.

As a novel EVI, the RCFT_{FCR} has had limited empirical support. To date, it has only been examined among cognitively intact students within an experimental malingering paradigm, and this has limited the generalizability of the findings to clinical populations (Giromini et al., 2019; Lindstrom et al., 2011; Sullivan & King, 2010; Viglione et al., 2019). In fact, Rai et al. (2019) emphasized the importance of testing their newly introduced instrument in a clinical sample, and they ended their paper with a call for replication among patients with confirmed or suspected genuine memory deficits. The present study, designed to examine the classification accuracy of the RCFT_{FCR} in patients clinically referred for neuropsychological assessment, is an answer to that call.

Given criticisms that it is easier to differentiate healthy controls from those instructed to feign deficits (experimental malingering paradigm) than credible patients with genuine cognitive deficits from patients with co-occurring genuine deficits and invalid performance (Fuermaier et al., 2017; Giromini et al., 2018; Merten & Rogers, 2017; Stevens et al., 2008; van Helvoort et al., 2019), we predicted an attenuation in the RCFT_{FCR}'s classification accuracy when the RCFT_{FCR} was applied to a clinical population. However, based on the results of the original study, we hypothesized that the Yes/No Recognition and the RCFT_{FCR} trials would be comparably sensitive to fluctuations in cognitive ability. Finally, we included a sample of students who were incentivized to perform well in order to address a separate limitation of the experimental malingering paradigm – variable motivation of research volunteers to demonstrate their maximal ability level as a control group (An et al., 2012; Hurtubise et al., 2020; Lace et al., 2020; Roye et al., 2019). Given this added incentivizing, we expected the student controls in our sample to outperform Rai et al.'s (2019) control group.

Method

Participants

Our clinical sample consisted of a consecutive case sequence of 52 patients referred for neuropsychological assessment to the last author's private practice in order to evaluate their cognitive and emotional functioning in the context of determining eligibility for disability benefits. As such, these participants can be considered as positive for having external incentives to appear impaired (Slick et al., 1999). The main inclusion criteria for these participants were the administration of the added RCFT_{FCR} trial and their informed consent for their clinical data to be used for research purposes. Participants' mean age was 37.9 years ($SD = 13.0$; range: 18–63). Their mean education level was 11.1 years ($SD = 2.0$; range: 6–14). Most (88.5%) were Caucasian (5.8% Black, 3.8% mixed, and 1.9% Aboriginal), right-handed (82.7%), and male (57.7%).

Our student sample consisted of 83 undergraduate students enrolled in a third-year course on psychometrics. As part of their grade, they were required to demonstrate credible performance during in-class assignments. This contingency was instituted to ensure that students took their assignments seriously, maximizing the pedagogical value of these experiential learning opportunities. Therefore, they had external incentives to perform well. Inclusion criteria were a valid administration of the RCFT_{FCR} trial, and at least one of the following free-standing PVTs: the first trial of the Test of Memory Malinger (TOMM-1; Tombaugh, 1996) or the Word Choice Test (WCT; Pearson, 2009). Since the RCFT Yes/No Recognition Test and RCFT_{FCR} ($n = 83$), the TOMM-1 ($n = 67$) and the WCT ($n = 75$) were administered on different days, the sample size for these tests differed (as denoted), reflecting the natural fluctuation in student attendance. As noted below in Procedures, students also gave informed consent for their performance data to be used in this research and had the opportunity to opt out from research participation (i.e., withdraw their consent for their anonymized test scores being used for academic research). Only de-identified data were used for this study. The university's Research Ethics Board approved the secondary use of the test scores for research purposes.

Measures

In addition to the RCFT, all patients completed a core battery of neuropsychological tests, including the Matrix Reasoning, Vocabulary, Digit Span and Coding subtests of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008), the Trail Making Test (TMT A & B; Reitan, 1955); the Hopkins Verbal Learning Test – Revised (HVLRT-R; Brandt & Benedict, 2001), letter, category and emotion word fluency (Abeare et al., 2017; Gladsjo et al., 1999), Stroop test of the Delis-Kaplan Executive

Function System (D-KEFS, 2001), Grooved Pegboard Test (GPB; Lafayette Instrument, 2015), Complex Ideational Material (CIM) of the Boston Diagnostic Aphasia Battery (Goodglass et al., 2001); Conners' Continuous Performance Test – Third Edition (CPT-3; Conners, 2015), the Rey Fifteen-Item (Rey-15) and Word Recognition Test (WRT; Rey, 1941); Boston Naming Test – Short Form (Erdodi et al., 2017), the Clock Drawing Test (CDT; Rouleau et al., 1992), the Patient Health Questionnaire (PHQ-9; Spitzer et al., 1999); the Generalized Anxiety Disorder Seven (GAD-7; Spitzer et al., 2006), the Five-Variable Psychiatric Screener (V-5; Erdodi, Jongma, et al., 2020; Sirianni et al., 2021) and the Behavior Rating Inventory of Executive Function (BRIEF; Roth et al., 2005). Demographically adjusted T-scores for the TMT, CIM, GPB and verbal fluency were calculated using the norms by Heaton et al. (2004). The main free-standing PVTs were the TOMM-1 and the WCT enhanced with the time-to-completion (T2C) cutoff.

Validity Composites (EI-5s). To complement the free-standing PVTs and to monitor the modality specificity effect (Lace et al., 2020; Rai & Erdodi, 2019; Schroeder et al., 2019), we developed two validity composites by aggregating individual EVIs, using the methodology developed by Erdodi (2019). The first one was based on tests that appeared to measure memory (EI-5_{MEM}), representing the modality-congruent criterion; the other was based on tests of processing speed (EI-5_{PSP}), representing the modality in-congruent criterion. The presence of an engineered method variance in criterion PVTs allowed for a more rigorous test of the classification accuracy of both RCFT recognition trials by minimizing the risk of spurious findings and improving ecological validity.

First, we recoded each of the five constituent PVTs onto a four-point ordinal scale such that a score that passed the most liberal cutoff was coded as zero, a score that failed the most conservative cutoff was coded as three, failing the next most liberal cutoff was coded as one, and failing the next most conservative cutoff was coded as two (see Table 1). We computed the value of the EI-5s by summing the recoded constituents, yielding a range from 0 (patient passed all five components at the most liberal cutoff) to 15 (patient failed all five components at the most conservative cutoff). An EI-5 value ≤ 1 was considered an overall *Pass*, as it signaled, at most, one marginal failure. EI-5 values 2 and 3 were difficult to interpret, as they might have represented either a couple of marginal failures or a single failure at a conservative cutoff. Neither of these combinations provided sufficient evidence to deem the entire profile invalid; therefore, this range was labeled *Borderline* and was excluded from analyses requiring a dichotomous outcome. However, an EI-5 ≥ 4 indicated either multiple failures at the liberal cutoff, or at least two at the conservative cutoff, crossing the line into the non-credible range (Pearson, 2009).

The majority of the participant samples (55-60%) scored in the *Passing* range on both versions of the EI-5. A quarter of the patients (26-28%) scored in the

Table 1. Components of the EI-5s and Base Rates of Failure at Given Cutoffs (Clinical Sample).

EI-5 _{MEM} Components	EI-5 Values				EI-5 _{PSP} Components	EI-5 Values			
	0	1	2	3		0	1	2	3
	Pass	Fail	FAIL	FAIL		Pass	Fail	FAIL	FAIL
CIM _{BDAE} Base rate	>9 71.7	8–9 18.9	7 3.9	≤6 5.7	Animals Base rate	>31 73.6	26–31 15.3	24–25 5.7	≤23 5.7
DS _{WAIS} Base rate	>6 68.6	5–6 18.9	3–4 5.9	≤2 3.9	CD _{WAIS} Base rate	>5 64.2	5 13.2	4 13.2	≤3 9.4
FCR _{HVLT-R} Base rate	12 83.0	11 5.7	9–10 9.4	≤8 1.9	TMT-A Base rate	>33 67.9	23–33 20.8	20–22 7.5	≤19 3.8
Rey-15 FR Base rate	>11 75.5	7–11 13.2	5–6 5.7	≤4 5.7	VAR _{CPT-3} Base rate	<65 71.7	65–74 13.2	75–79 9.4	≥80 5.7
Rey WRT Base Rate	>6 71.7	5–6 17.0	4 5.7	≤3 5.7	Word _{D-KEFS} Base rate	>5 69.8	4–5 9.5	2–3 9.5	1 11.3

Note. Shading represents the change in confidence in correctly classifying a given score as *invalid* (darker means more likely to be invalid); EI-5_{MEM}: Erdodi Index Five - Memory; EI-5_{PSP}: Erdodi Index Five – Processing Speed; CIM_{BDAE}: Complex Ideational Material subtest of the Boston Diagnostic Aphasia Battery raw score (An et al., 2019; Erdodi, 2019; Erdodi et al., 2016; Erdodi & Roth, 2017); DS_{WAIS}: Digit Span subtest of the Wechsler Adult Intelligence Scale (age-corrected scaled score; Erdodi & Abeare, 2020; Erdodi & Lichtenstein, 2017; Hurtubise et al., 2020; Reese et al., 2012; Shura et al., 2019; Spencer et al., 2013; Webber & Soble, 2018; Whitney et al., 2009); FCR_{HVLT-R}: Forced choice recognition trial of the Hopkins Verbal Learning Test – Revised (Abeare, Hurtubise, et al., 2020; Cutler et al., 2021); Rey-15 FR: Rey Fifteen-Item Test free recall (Boone et al., 2002; Lezak, 1995; Merten et al., 2005; O’Bryant et al., 2003; Poynter et al., 2019; Russeler et al., 2008); Rey WRT: Rey Word Recognition Test (Bell-Sprinkel et al., 2013; Goworowski et al., 2020; Love et al., 2014; Nitch et al., 2006; Smith et al., 2014); Animals: Animal fluency T-score using norms by Heaton et al., 2004 (Hurtubise et al., 2020; Sugarman & Axelrod, 2015); CD_{WAIS}: Coding subtest of the Wechsler Adult Intelligence Scale (age-corrected scaled score; Erdodi & Abeare, 2020; Erdodi, Abeare, et al., 2017; Erdodi & Lichtenstein, 2017; Etherton et al., 2006; Inman & Berry, 2002; Kim et al., 2010; Trueblood, 1994); TMT-A: Trail Making Test – Part A T-score using norms by Heaton et al., 2004 (Abeare, Sabelli, et al., 2019; Ashendorf et al., 2017; Erdodi & Lichtenstein, 2020); VAR_{CPT-3}: Variability score of the Conners’ Continuous Performance Test – Third Edition (Erdodi, Pelletier, et al., 2018; Erdodi, Roth, et al., 2014; Ord et al., 2020); Word_{D-KEFS}: Word Reading trial of the Delis-Kaplan Executive Function System (Arentsen et al., 2013; Boskovic et al., 2018; Donders & Hayden, 2020; Egeland & Langfjaern, 2007; Eglit et al., 2019; Erdodi, Sagar, et al., 2018; Guise et al., 2014).

Failing range. Consistent with previous research (An et al., 2019; Erdodi & Abeare, 2020; Erdodi, Green, et al., 2019; Erdodi & Rai, 2017; Erdodi, Taylor, et al., 2019; Rai & Erdodi, 2019), the EI-5s were significant predictors of the two free-standing PVTs. An EI-5_{MEM} score ≥4 was specific to failing the

TOMM-1 or the WCT (.90–.96), at .65 sensitivity. The EI-5_{PSP} produced a similar combination of sensitivity (.69–.70) and specificity (.86–.96).

Visual-Perceptual Ability Composite (VPA-3). The VPA-3 was designed to serve as the ability measure counterpart to the EI-5s. Thus, the VPA-3 was conceived as a composite of visual-perceptual ability, created as a criterion measure for evaluating the RCFT_{FCR}'s sensitivity to the examinees' natural fluctuations in perceptual skills. As its name suggests, the VPA-3 consisted of three tests designed to assess perceptual reasoning, visual scanning or visuomotor speed: the Matrix Reasoning and Coding subtests of WAIS-IV and TMT-A. Similar to the EI-5s, these components of the VPA-3 were recoded onto a five-point scale in which zero corresponded to two *SDs* below the mean (i.e., Impaired range), whereas four corresponded to two *SDs* above the normative mean (i.e., a Very Superior range). A score of two represented the Average range (see Table 2).

The VPA-3 is analogous to the WAIS-IV index scores (Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed) as it combines information from multiple tests into a single-number summary of the broader construct. Multivariate measurement models based on aggregating different sources of data have been shown to be superior to single test scores (Abeare, Erdodi, et al., 2020; Pearson, 2009; Tyson et al., 2018). For the specific purpose of this study, the VPA-3 served as the criterion for calibrating the RCFT_{FCR} as a measure of visual recognition memory (i.e., *ability test*).

Procedure

Patients completed a clinical interview and a comprehensive neuropsychological test battery. Tests were administered and scored by trained psychometrists under the supervision of a licensed clinical neuropsychologist. Students were administered the RCFT with the FCR trial, the TOMM-1 and WCT as a

Table 2. Components of the Visual-Perceptual Ability Composite (VPA-3) in the Clinical Sample

Test	Scale	VPA-3 value				
		0	1	2	3	4
CD _{WAIS}	ACSS	≤4	5–7	8–12	13–15	≥16
MR _{WAIS}	ACSS	≤4	5–7	8–12	13–15	≥16
TMT-A	T	≤30	31–43	44–56	57–69	≥70

Note. CD_{WAIS}: Coding subtest of the Wechsler Adult Intelligence Scale (age-corrected scaled score); MR_{WAIS}: Matrix Reasoning subtest of the Wechsler Adult Intelligence Scale (age-corrected scaled score); TMT-A: Trail Making Test – Part A T-score using norms by Heaton et al. (2004).

group in a classroom. Failing the validity cutoff embedded within a given assignment resulted in a 0.2–0.6% penalty applied to the final grade, depending on the actual score. In addition, the instructor continuously emphasized the educational value of the in-class assignments to encourage full engagement. As noted earlier, students had the opportunity to opt out from research participation (i.e., withdraw their consent for their anonymized test scores being used for academic research). Only de-identified student data were used for the purposes of this research, and the protocols for testing both participant groups were approved by the Research Ethics Board of the university.

Data Analysis

When relevant, we computed the base rate of failure (BR_{Fail} ; i.e., the percent of the sample that failed a given cutoff). The prevalence of the condition of interest (in this context, BR_{Fail}) is a descriptive statistic that is important for understanding classification accuracy in general (Wald & Bestwick, 2014) and in the context of performance validity assessment specifically (Abeare, Messa, et al., 2019). Although area under the curve (AUC) is useful for comparing overall classification accuracy across models (Altman & Bland, 1994; Fawcett, 2006; Marzban, 2004), its clinical relevance has been called into question (Hand, 2009; Lobo et al., 2008; Wald & Bestwick, 2014). Therefore, sensitivity and specificity values were also computed around relevant cutoffs. In the context of performance validity assessment, specificity is the most important parameter that determines the clinically recommended cutoffs, whereas sensitivity is sacrificed in the interest of minimizing the false positive rate. The lowest acceptable specificity value is .84 (Larrabee, 2003), although values $\geq .90$ are desirable (Roberson et al., 2013). The main inferential statistic was the *t*-test, two-proportions *z*-test, Levene's test of homogeneity of variance, AUCs with 95% CI, and Pearson's product-moment correlations (r_{xy}). All tests were two-tailed; alpha-level was set at .05. Effect size estimates were expressed in Cohen's *d* and squared correlation coefficients (r_{xy}^2).

Results

Neuropsychological Functioning of the Clinical Sample

The patient sample's performance on the Vocabulary ($M=8.1$), Matrix Reasoning ($M=7.6$) and Coding ($M=6.9$) subtests of the WAIS-IV was in the Low Average range. The mean raw score on their RCFT Copy trial was 28.0. Performance on the acquisition ($M_{T-score}=32.6$) and delayed free recall ($M_{T-score}=33.0$) trials of the HVLt-R fell in the Borderline range. The mean raw score on the CDT was 8.5 out of 10. Performance on the TMT-A was in the Low Average range ($M_{T-score}=40.8$). Dominant hand GPB performance, letter

(FAS) and category (animal) fluency were in the Low Average range ($M_{T\text{-score}}$: 40.6–41.2). Self-reported depression on the PHQ-9 was in the Severe range ($M = 16.2$); self-reported anxiety on the GAD-7 was in the Moderate range ($M = 13.7$). The General Executive Composite on the BRIEF was in the clinical range ($M_{T\text{-score}} = 73.8$). Self-reported depression, anxiety and pain on the V-5 fell within the Moderate range (42.0–63.7).

RCFT Yes/No Recognition Versus RCFT_{FCR} Scores for Clinical and Student Samples

Effects of Demographic Variables and Intra-Individual Differences. In the clinical sample, scores on the Yes/No Recognition and RCFT_{FCR} trials were independent of examinees' age ($p = .434$ and $.728$), education ($p = .059$ and $.144$) or sex ($p = .237$ and $.977$). Also, a repeated measures t -test revealed no significant difference between scores on the Yes/No Recognition trial ($M = 19.3$, $SD = 2.2$) and the RCFT_{FCR} trial ($M = 19.3$, $SD = 2.9$): $t(51) = 0.18$, $p = .859$. These two recognition trials were positively correlated [$r(52) = .37$, $p = .001$]. One patient scored below chance level (<12) on the RCFT_{FCR}.

In the student sample, the mean performance on the RCFT_{FCR} trial ($M = 22.1$, $SD = 1.8$) was significantly higher than on the Yes/No Recognition trial ($M = 21.3$, $SD = 1.6$): $t(82) = 3.90$, $p < .001$, $d = 0.43$ (medium effect). As with the clinical sample, the two recognition trials were positively correlated [$r(83) = .62$, $p < .001$], but no participant scored below chance level on either RCFT recognition trial.

Sensitivity to PVT Failure. Interestingly, in the patient sample, there was no significant difference between patients who passed and those who failed the TOMM-1 on either of the RCFT recognition trials (Table 3). However, among patients, failing the WCT was associated with significantly lower performances on both RCFT recognition trials ($d: 0.82$ – 1.30 , large effects). Similarly, patients who failed the EI-5s produced lower scores on both the Yes/No Recognition and the RCFT_{FCR} trial. However, the effect was more pronounced on the EI-5_{MEM} ($d: 1.13$ – 1.63 , large) than on the EI-5_{PSP} ($d: 0.89$ – 1.09 , large). The effect size associated with failing the EI-5s was greater for the RCFT_{FCR} ($d: 1.09$ – 1.63 , large) than for the Yes/No Recognition trial ($d: 0.89$ – 1.13 , large).

RCFT Variables as EVIs

Clinical Sample. The RCFT Copy trial was a significant predictor of all four criterion PVTs (Table 4). The first cutoff to reach the .90 specificity standard was ≤ 25.0 , with .33–.43 sensitivity. At ≤ 23 , specificity improved (.93–.97) at a reasonable cost to sensitivity (.29–.36). At ≤ 20 , the Copy trial reached perfect specificity.

Table 3. Independent t-Tests Comparing Performance on the RCFT Yes/No and FCR Trials as a Function of Sample and Passing or Failing the Criterion PVTs.

Sample	Criterion		RCFT Y/N recognition						RCFT _{FCR}				
	PVT		<i>n</i>	%	<i>M</i>	<i>SD</i>	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>d</i>	
Clinical	TOMM-1	Pass	31	58.5	19.7	2.1	.087	–	19.6	2.9	.330	–	
		Fail	22	41.5	18.7	2.3			18.8	3.9			
	WCT	Pass	29	54.7	20.1	1.8	.005	0.82	20.7	1.8	<.001	1.30	
		Fail	24	45.3	18.4	2.3			17.4	3.1 ^a			
	EI-5 _{MEM}	Pass	32	69.6	19.9	1.9	<.001	1.13	20.2	2.4	<.001	1.63	
		Fail	14	30.4	17.6	2.0			16.2	2.5			
	EI-5 _{PSP}	Pass	29	65.9	19.9	1.7	.005	0.89	20.4	2.2	.001	1.09	
		Fail	15	34.1	17.9	2.7 ^a			17.2	3.5 ^a			
	Student	TOMM-1	Pass	61	91.0	21.3	1.6	.413	–	22.4	1.6	.099	–
			Fail	6	9.0	20.7	2.9 ^a			21.2	2.6 ^a		
		WCT	Pass	53	70.7	21.4	1.4	.747	–	22.4	1.7	.085	–
			Fail	22	29.3	21.3	1.7			21.7	1.4		

Note. ^aLevene’s test of homogeneity of variance $p < .05$; PVT: Performance Validity Test; RCFT: Rey Complex Figure Test; Y/N: Yes/No; FCR: Forced Choice Recognition raw score; TOMM-1: Trial 1 on the Test of Memory Malingering (Denning, 2012; Fazio et al., 2017; Greve et al., 2006, 2009; Jones, 2013; Kulas et al., 2014; Martin et al., 2020; Powell et al., 2004; Rai & Erdodi, 2019; Webber et al., 2018); WCT: Word Choice Test [Fail defined as accuracy score ≤ 47 (Barhon et al., 2015; Davis, 2014; Erdodi, Kirsch, et al., 2014; Pearson, 2009) or time-to-completion ≥ 156 seconds (Erdodi & Lichtenstein, 2021; Erdodi, Tyson, et al., 2017; Zuccato et al., 2018)]; EI-5_{MEM}: Erdodi Index Five – Memory (Fail defined as ≥ 4); EI-5_{PSP}: Erdodi Index Five – Processing Speed (Fail defined as ≥ 4).

The RCFT Yes/No Recognition trial was a significant predictor of three criterion PVTs, narrowly missing the TOMM-1. The ≤ 18 cutoff failed to reach minimum specificity against any of the criterion PVTs. Lowering the cutoff to ≤ 17 notably improved specificity (.87–.93), at .27–.47 sensitivity. Making the cutoff even more conservative (≤ 16) produced marginal improvements in specificity (.90–.97) at negligible cost to sensitivity (.27–.47). Further lowering the cutoff (≤ 15) reached the point of diminishing return: small gains in specificity (.97–1.00) and a notable decline in sensitivity (.09–.20).

The RCFT_{FCR} was a significant predictor of three of the criterion PVTs. The ≤ 18 cutoff achieved minimum specificity standards (.92–.97) against all criterion PVTs but the TOMM-1 (.81), at a wide range of sensitivity (.52–.92). Lowering the cutoff to ≤ 17 resulted in trivial gains in specificity (.84–.97) but a precipitous drop in sensitivity (.29–.54). Making the cutoff even more conservative (≤ 16)

Table 4. Classification Accuracy of the EVIs Within the RCFT Against Criterion PVTs in the Clinical Sample.

RCFT	Variable	Statistic	AUC (95 CI)	Cutoff	BR _{Fail}	Criterion PVT														
						TOMM-1			WCT			EI-5 _{MEM}			EI-5 _{PSF}					
						SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC					
Copy						41.5	45.3	30.4	34.1											
				≤25	26.9	.68 (.53–.83)	.72 (.58–.86)	.83	.81	.68 (.51–.85)										
				≤24	20.8	.41	.33	.38	.40	.40	.90	.94	.94	.94	.94	.94	.94	.94	.94	.94
				≤23	15.1	.32	.29	.36	.33	.33	.97	.97	.97	.97	.97	.97	.97	.97	.97	.97
Y/N Rec				≤20	9.4	.23	.21	.21	.27	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
				≤18	30.2	.64 (.48–.80)	.71 (.56–.85)	.83	.78	.73 (.54–.91)										
				≤17	19.2	.27	.33	.43	.47	.53	.93	.93	.91	.91	.91	.91	.91	.91	.91	.91
				≤16	17.0	.27	.29	.43	.47	.47	.97	.93	.94	.94	.94	.94	.94	.94	.94	.94
FCR				≤15	5.7	.09	.13	.14	.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
				≤18	32.7	.60 (.44–.76)	.83 (.71–.95)	.97	.88	.79 (.63–.96)										
				≤17	21.2	.29	.43	.54	.47	.73	.93	.97	.91	.91	.91	.91	.91	.91	.91	.91
				≤16	15.4	.19	.35	.38	.40	.40	.96	1.00	.94	.94	.94	.94	.94	.94	.94	.94
Both Y/N Rec & FCR				≤15	11.5	.14	.26	.31	.27	.96	1.00	1.00	.94	.94	.94	.94	.94	.94	.94	
				≤18	19.2	.33	.39	.54	.53	.96	.97	.97	.94	.94	.94	.94	.94	.94	.94	.94
				≤17	13.5	.19	.26	.31	.33	.96	.97	.97	.94	.94	.94	.94	.94	.94	.94	.94
				≤16	7.7	.10	.17	.15	.27	.27	1.00	1.00	.97	.97	.97	.97	.97	.97	.97	.97

Note. EVI: Embedded validity indicators; PVT: Performance Validity Test; RCFT: Rey Complex Figure Test; Y/N Rec: Yes/No recognition raw score; FCR: Forced Choice Recognition raw score; TOMM-1: Trial 1 on the Test of Memory Malinger (Denning, 2012; Fazio et al., 2017; Greve et al., 2006, 2009; Jones, 2013; Kulas et al., 2014; Martin et al., 2019; Powell et al., 2004; Rai & Erdodi, 2019; Webber et al., 2018); WCT: Word Choice Test [Fail defined as accuracy score ≤47 (Barhon et al., 2015; Davis, 2014; Erdodi, Kirsch, et al., 2014; Pearson, 2009) or time-to-completion ≥156 seconds (Erdodi & Lichtenstein, 2020; Erdodi, Tyson, et al., 2017; Zuccato et al., 2018)]; EI-5_{MEM}: Erdodi Index Five – Memory (Fail defined as ≥4); EI-5_{PSF}: Erdodi Index Five – Processing Speed (Fail defined as ≥4); BR_{Fail}: Base rate of failure (% of the sample that failed a given cutoff); SENS: Sensitivity; SPEC: Specificity.

produced the predictable trade-off: further improvement in specificity (.87–1.00) and decline in sensitivity (.19–.40).

As a last step, the Yes/No Recognition trial and the RCFT_{FCR} were combined. *Pass* on the combined measure was defined as a score above the cutoff on either of the two; *Fail* was defined as a score below the cutoff on both. The combined cutoff of ≤ 18 achieved the minimum specificity standard against all criterion PVTs (.90–.97), at .33–.54 sensitivity. Lowering the cutoff to ≤ 17 was the point of diminishing returns: no change in specificity, but a notable drop in sensitivity (.19–.33). Making the cutoff more conservative (≤ 16) consolidated specificity (.94–1.00) at a proportional cost to sensitivity (.10–.27).

Although there was a high rate of agreement (83–87%) between the Yes/No Recognition and the RCFT_{FCR} trial at ≤ 16 and ≤ 17 , 8% of the sample that passed the former failed the latter. Conversely, 6–10% of patients passed RCFT_{FCR} but failed Yes/No Recognition at the same cutoff. As such, the two RCFT recognition trials were non-redundant EVIs.

Student Sample. Neither the Yes/No Recognition nor the RCFT_{FCR} trial was a significant predictor of passing or failing the TOMM-1. The Yes/No Recognition trial also produced a non-significant AUC (.64, 95% CI: .39–.90) against the WCT. However, the RCFT_{FCR} trial was a significant predictor of passing or failing the WCT (AUC = .68, 95% CI .55–.81). The only conceivable cutoff (≤ 18) was highly specific (.93–.97) but insensitive (.05–.33) on both trials, against both criterion PVTs. Failing the ≤ 18 cutoff on both the Yes/No Recognition and the RCFT_{FCR} trials was associated with perfect specificity.

The Effect of External Incentive vs. Genuine Impairment

To dissociate the relative contributions of external incentive status and genuine cognitive deficits of the two RCFT recognition scores, we compared our two samples to Rai et al.'s (2019) control group. Our clinical sample had a significantly lower performance on both RCFT recognition trials ($d = 0.62$, medium effect). Our student sample produced a higher mean on both the Yes/No Recognition ($d = 0.30$, small effect) and the RCFT_{FCR} trial ($d = 0.62$, medium effect) than Rai et al.'s (2019) control group. Our clinical sample had a significantly lower mean on the Yes/No Recognition trial compared to our student sample: $t(130) = 5.68$, $p < .001$, $d = 1.04$ (large effect), and we observed similar results on the RCFT_{FCR}: $t(130) = 6.25$, $p < .001$, $d = 1.39$ (large effect). Figure 1 provides a visual display of the RCFT_{FCR}'s differential sensitivity to incentive status.

Results diverged on free-standing PVTs. BR_{Fail} on the TOMM-1 was significantly higher (40.4%) in our clinical sample compared to our student sample (9.0%): $z(133) = 4.06$, $p < .001$. However, on the WCT the two groups were not statistically different on BR_{Fail} (29.3–30.8%): $z(125) = 0.17$, $p = .862$.

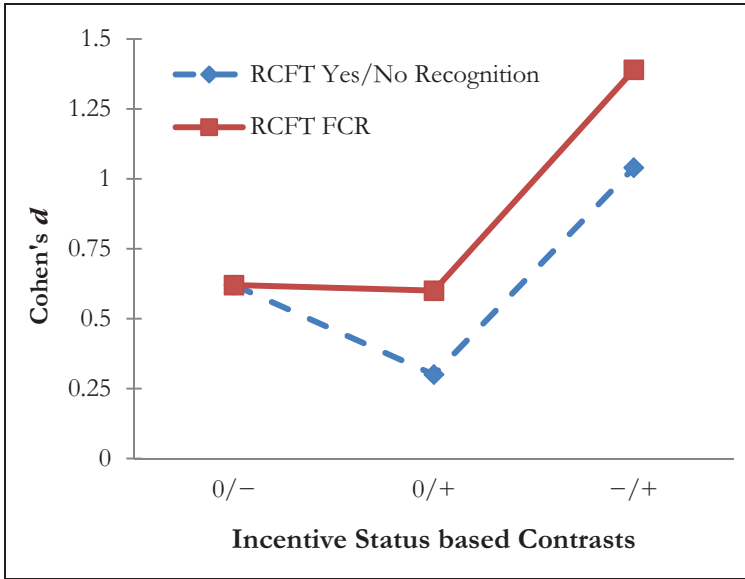


Figure 1. Effect Sizes Associated With Pairwise Contrasts Between Samples as a Function of Incentive Status. 0: Neutral [no incentive to underperform or to perform at maximal ability; represented by the control group ($n = 80$) from the study by Rai et al. (2019)]; -: Negative Incentive [i.e., motivated to underperform; represented by the clinical sample ($n = 52$) from the present study]; +: Positive Incentive [i.e., motivated to perform at maximal ability; represented by the incentivized student sample ($n = 83$) from the present study]; RCFT: Rey Complex Figure Test; FCR: Forced Choice Recognition.

RCFT Recognition Trials as Predictors of Cognitive Ability

A visual inspection of the distribution of raw scores for the two RCFT recognition trials across our two samples revealed several important features of these scales. The RCFT_{FCR} was better at discriminating between the two groups (i.e., students and patients) than was the Yes/No Recognition trial. However, within each sample, the two RCFT recognition trials produced similar cumulative frequency curves.

Correlation coefficients were computed between the Yes/No Recognition trial, the RCFT_{FCR} and select measures of cognitive ability. Both RCFT recognition trials positively correlated (.45–.55, $p < .01$) with the VPA-3. However, the Yes/No Recognition trial was unrelated to performance on the CDT, while the RCFT_{FCR} positively correlated with it [$r(53) = .33$, $p = .018$]. Both trials were correlated with key scores on the HVLt-R. However, only the RCFT_{FCR} produced a significant correlation with the HVLt-R_{FCR} [$r(53) = .37$, $p = .007$]. Overall, the RCFT_{FCR} shared more variance with the

VPA-3, CDT and the HVLTR-R (r_{xy}^2 : .11–.35) than did the Yes/No Recognition trial (r_{xy}^2 : .01–.24).

Finally, the classification accuracy of the Copy trial, the Yes/No Recognition and RCFT_{FCR} trials were computed using the VPA-3, Clock Drawing Test and the Delayed Recall (DR) trial of the HVLTR-R as criterion measures (see Table 5). The score on the Copy trial only produced a significant AUC against the VPA-3 (.70; 95% CI: 62–89). It also produced the most stable overall correct classification (.64–.69). The Yes/No Recognition trial was a significant predictor of the VPA-3 and the HVLTR-R DR, although overall correct classification was lower (.52–.65). The RCFT_{FCR} produced the highest AUC (.73–.87; 95% CI: .57–.97) and overall correct classification (.56–.84).

Clinical Implications

Since a score ≤ 17 on the recognition trials was specific to invalid performance and therefore, an unreliable measure of *cognitive ability*, the Yes/No Recognition and RCFT_{FCR} scores were effectively seven-point scales (18–24). Attempts to provide demographically stratified standard scores that span across the full spectrum of cognitive functioning (from Impaired to Very Superior) within such a restricted range would likely be fraught with scaling artifacts. Therefore, we propose a three-way clinical classification of Inferior (mild deficits), Within Normal Limits (WNL; intact/average range performance) and Superior (above average), with scores ≤ 14 considered Invalid, and scores of 15–17 considered Questionable. Reducing a measurement scale to a small number of clinically meaningful categories is a long-standing practice in neuropsychology (Guilmette et al., 2020; Lezak et al., 2012). Of course, the ultimate interpretation will depend on the clinical context, weighing medically verified neuropsychiatric conditions as mitigating factors against the number and level of PVT failures. Within the clinical sample, there was a strong linear relationship between RCFT recognition scores and VPA-3 and HVLTR-R DR values (see Table 6).

Discussion

This study was the first attempt to validate the RCFT_{FCR} trial as an EVI for a clinical sample. We hypothesized that (a) the RCFT_{FCR}'s classification accuracy would be attenuated by a confluence of genuine deficits and non-credible performance; (b) RCFT_{FCR} classification accuracy would be similarly sensitive to natural variability in cognitive ability as the RCFT's Yes/No Recognition trial; and (c) our student control sample, incentivized to perform well, would produce higher scores on the both the RCFT_{FCR} and the RCFT Yes/No recognition trials than had Rai et al.'s (2019) control group who were not incentivized to perform well. Our results provided mixed support for these hypotheses.

Table 5. RCFT Variables as Predictors of Visuospatial Skills and Verbal Memory in the Clinical Sample.

RCFT Variable	Criterion variable											
	VPA-3				CDT				HVLIT-DR			
	Cutoff	BR	SENS	SPEC	ACC	SENS	SPEC	ACC	SENS	SPEC	ACC	
Copy			.70 (.62-.89)		.64	.71 (.46-.96)		.72	.60 (.45-.76)		.53	
	≥28	63%	.47	.80	.64	.70	.73	.72	.39	.67	.53	
	≥30	40%	.73	.60	.67	.70	.44	.57	.68	.52	.60	
	≥32	31%	.83	.50	.67	.70	.32	.51	.74	.38	.54	
	≥34	17%	.97	.40	.69	.80	.17	.49	.87	.24	.56	
Y/N Rec			.69 (.52-.82)		.59	.65 (.43-.87)		.65	.72 (.58-.86)		.65	
	≥19	69%	.53	.65	.59	.70	.59	.65	.58	.71	.65	
	≥20	54%	.77	.40	.59	.80	.32	.56	.77	.38	.58	
	≥21	29%	.83	.20	.52	.80	.17	.49	.87	.24	.56	
	≥22	17%	1.00	.10	.55	.90	.02	.46	1.00	.10	.55	
FCR			.82 (.71-.94)		.69	.73 (.57-.90)		.70	.87(.77-.97)		.84	
	≥19	67%	.63	.75	.69	.80	.59	.70	.77	.90	.84	
	≥20	50%	.83	.65	.74	.90	.41	.66	.87	.67	.77	
	≥21	35%	.90	.50	.70	.90	.29	.60	.87	.43	.65	
	≥22	25%	.97	.20	.59	1.00	.12	.56	.97	.19	.58	

Note. RCFT: Rey Complex Figure Test; Y/N Rec: Yes/No Recognition raw score; FCR: Forced Choice Recognition raw score; VPA-3: Visual-Perceptual Ability Composite (cutoff; CDT: Clock Drawing Test (Rouleau et al., 1992); HVLIT-R: Hopkins Verbal Learning Test – Revised (Brandt & Benedict, 2001)); DR: Delayed recall raw score; BR: Base rate (%); SENS: Sensitivity; SPEC: Specificity; ACC: Overall accuracy (average of sensitivity and specificity).

Table 6. Cumulative Percentage (%_{CUM}) and Recommended Clinical Classification Ranges for RCFT Recognition Trial Scores within the Student and Clinical Samples.

Raw Score	Yes/No Recognition				RCFT _{FCR}				Classification range
	Student		Clinical		Student		Clinical		
	% _{CUM}	% _{CUM}	M _{VPA-3}	M _{HVLT}	% _{CUM}	% _{CUM}	M _{VPA-3}	M _{HVLT}	
≤12	0	0			0	1.9			Invalid
13	0	0			0	5.8			Invalid
14	0	1.9			0	9.6			Invalid
15	0	5.8			0	11.5			Questionable
16	1.2	17.3			1.2	15.4			Questionable
17	1.2	19.2			2.4	21.2			Questionable
18	6.0	30.8	4.0	25.5	6.0	32.7	1.8	18.8	Inferior
19	12.0	46.2	4.4	36.6	7.2	50.0	5.0	31.4	Inferior
20	30.1	71.2	4.1	37.3	16.9	65.4	3.6	36.6	Within normal limits
21	53.0	82.8	5.3	36.3	26.5	75.0	4.8	41.8	Within normal limits
22	71.1	96.2	4.3	33.0	49.4	90.4	5.3	41.0	Within normal limits
23	95.2	98.1	6.0	40.0	78.3	94.2	5.5	46.5	Superior
24	100.0	100.0	7.0	59.0	100.0	100.0	6.3	46.0	Superior

Note. Shading represents the change in confidence in correctly classifying a given score as *invalid* (darker means more likely to be invalid) and delineates the range of performance that should not be interpreted clinically; RCFT: Rey Complex Figure Test; FCR: Forced Choice Recognition raw score; M_{VPA-3}: Mean score on the Visual-Perceptual Ability Composite; M_{HVLT}: Mean T-score on the Delayed Recall trial of the Hopkins Verbal Learning Test - Revised; **Invalid**: Scores in this range have not been observed in healthy controls, are rare in clinical patients, and when they do occur, they are associated with failure on other performance validity tests; therefore, they should not be interpreted as evidence of impairment;

Questionable: Scores in this range are rare in both healthy controls and clinical patients, and when they do occur, they are associated with failure on other performance validity tests; however, in examinees with otherwise valid neurocognitive profiles, they may be considered evidence of impaired visuoperceptual and memory; **Inferior**: Scores in this range are rare in healthy controls, but observed in a third of clinical patients; therefore, provide evidence of mild cognitive deficits; **Within Normal Limits**: About half of the healthy controls and clinical patients scored in this range, indicating intact performance; **Superior**: A score in this range indicates above average performance in healthy controls, and top 5-10% performance in clinical patients.

AUC values for the two RCFT recognition trials against the TOMM-1 were significantly lower (.60-.64) in our clinical sample compared to the original RCFT_{FCR} study (.78-.82). However, while the AUC for Yes/No Recognition was significantly lower against the WCT compared to the original sample (.71 versus .82), the RCFT_{FCR} achieved essentially the same AUC (.83) as it had in the original study. In fact, the ≤18 cutoff for the RCFT_{FCR} produced comparable specificity values (.81-.97 versus .88-.89) while maintaining similar levels of sensitivity (.52-.92 versus .58-.72). Thus, contrary to our expectations, the RCFT_{FCR} maintained the same classification accuracy in our clinical sample as

had been reported by Rai et al. (2019) in the experimental malingering paradigm.

In terms of its sensitivity to fluctuations in cognitive ability, the RCFT_{FCR} outperformed Yes/No Recognition, explaining a larger proportion of variance (11–35% versus 1–24%) in visuospatial and verbal memory skills and producing superior classification accuracy (AUC: .73–.87 versus .65–.72) as a PVT in our clinical sample. The RCFT_{FCR} distribution was also associated with a stronger, more refined, and clinically meaningful gradient of difficulty when using the VPA-3 and the HVLt-R DR as references (see Table 6).

Our last hypothesis was fully supported in that our incentivized student sample outperformed Rai et al.'s (2019) control group in their original validation study on both RCFT recognition trials (small-medium effect). However, the RCFT_{FCR} was actually more sensitive to changes in performance associated with incentive status than was the Yes/No Recognition trial (d : .60–1.39 versus .30–1.04).

Incidental Findings

Our data were largely consistent with the domain specificity effect, in that similarity in the cognitive domain (attention, memory, processing speed, verbal reasoning) or sensory modality (auditory, visual, tactile) between predictor and criterion PVTs influenced classification accuracy (Abeare, Sabelli, et al., 2019; Erdodi, 2019; Schroeder et al., 2019). AUC values for the RCFT based EVIs were consistently higher against the EI-5_{MEM} (the modality-congruent validity composite) than the EI-5_{PSP} (the modality-incongruent validity composite), suggesting that instrumentation artifacts may exert a subtle but detectable influence on signal detection analyses. Although both the RCFT Yes/No Recognition and the RCFT_{FCR} trials performed well across a strategically engineered variability in criterion PVTs, increasing confidence in our overall findings, domain specificity as a potential confound may warrant further research.

Interestingly, the effect of the examinee's incentive status on the outcome of these free-standing PVTs was instrument specific. Namely, BR_{Fail} on the TOMM-1 was 4.5 times higher among clinical patients with an incentive to appear impaired than among students with an incentive to demonstrate their best ability. In contrast, there was no difference between these two groups for BR_{Fail} on the WCT. These findings are consistent with previous reports (Abeare, Erdodi, et al., 2020; Erdodi, Hurtubise, et al., 2018), and they challenge the prominence of the examinees' external incentive status in diagnostic models for malingering (APA, 2013; Erdodi et al., 2018; Slick et al., 1999). While the new RCFT_{FCR} trial was more sensitive to the effect of incentive status than the RCFT Yes/No Recognition trial (Figure 1), it must be noted that incentive to appear impaired and elevated risk of genuine neuropsychological deficits were conflated in our clinical sample.

The unexpectedly high BR_{Fail} on the free-standing PVTs (9.0%-29.2%) in the incentivized control group puts the low BR_{Fail} on the Yes/No Recognition and RCFT_{FCR} trials (1.2%) in perspective. Namely, it neutralizes arguments that EVIs inevitably conflate genuine impairment and non-credible responding (Glassmire et al., 2019; Messa et al., 2020) when compared to free-standing PVTs that are, by design, robust to genuine and severe cognitive impairment (Abeare et al., 2019; Erdodi & Rai, 2017; Whitney et al., 2013). If this finding is replicated by future research, it would further enhance the appeal of the RCFT recognition trials as EVIs.

The improved performance on the RCFT_{FCR} relative to the Yes/No Recognition trial in the student sample may be due to the fact it controls for variability in the subjective threshold of certainty individuals require to endorse a given shape as a target stimulus. Cautious examinees may choose not to circle Yes/No Recognition items that seem familiar (but are not fully confident in their decision) to avoid making an error. In contrast, knowing that one of the drawings within each pair of the RCFT_{FCR} is definitely a target makes it easier to select the more familiar item.

Reflections on Control Group Methodologies

Our a priori, rationally based prediction of relative shrinkage in classification accuracy among patients with genuine cognitive deficits and external incentives to appear impaired was not substantiated. This surprising finding underlines the importance of empirically verifying even intuitive and logically appealing assumptions. Our results supported, instead, previously voiced theoretical concerns (Giromini et al., 2019; McWhirter et al., 2019; van Helvoort et al., 2019) and published data (Abeare et al., 2019; An et al., 2019; Hurtubise et al., 2020; Roye et al., 2019) about the epistemological ambiguity around the incentive status of control groups comprised of undergraduate research volunteers. Researchers long assumed that cognitively healthy university students assigned to the control condition would demonstrate their highest ability level by default. Some newer investigators began to question the validity of this unverified assumption, noting that research participants had been rewarded for their time but not for the quality of the data they produced (An et al., 2017; Powell et al., 2004; Roye et al., 2019; Russeler et al., 2008; Tan et al., 2002). Therefore, the magnitude of their incentive to fully comply with the instructions (i.e., appear impaired without being detected) does not match real-world malingerers who might be incentivized by 7-figure personal injury settlements (Dunn et al., 2003; Grant et al., 2020; Jelcic et al., 2011).

Moreover, emerging evidence suggests that the instructions given to examinees have a weak effect on the credibility of their response sets overall (Abeare, Hurtubise, et al., 2020; Niesten et al., 2017). In other words, reminding real-world patients to provide valid data does not assure that outcome. Likewise,

past studies showed that a variable proportion of student volunteers who were assigned to the control condition and asked to demonstrate their best ability in academic research settings failed PVTs (An et al., 2012; DeRight & Jorgensen, 2015; Ross et al., 2016; Roye et al., 2019; Santos et al., 2014; Silk-Eglit et al., 2014). Ironically, participants assigned to the experimental malingering condition also occasionally demonstrated intact cognitive ability – in other words, they *failed at failing* (Abeare et al., 2020).

Essentially, past performance validity research studies using the experimental malingering paradigm specifically and relying on student volunteer participants generally, were subject to several internal and external validity threats. There seems to be an emerging consensus that there is no guarantee that any given research participants will comply with study instructions. An et al. (2017) went as far as to suggest that an incentive to appear impaired and a lack of incentive to perform well are similar motivational states. Similarly, criterion grouping in studies based on experimental malingering can be considered a pseudo-independent variable (Hurtubise et al., 2020), as the only control investigators have in this circumstance is through instructions given, but does not extend to whether those instructions are executed.

Results from the current study further addressed the credibility of psychometric data produced by cognitively healthy university students. Despite (a) salient demand characteristics (i.e., administering tests in a classroom setting and repeatedly emphasizing the educational value of full engagement) and (b) a performance-based reinforcement contingency (i.e., full points only awarded for valid responses), a surprisingly high proportion of our student controls (29.3%) failed the WCT, a free-standing PVT. This failure rate is twice as high as the rate of non-credible profiles in clinical and even forensic settings (Young, 2015), and it far exceeds findings by previous research on performance validity among undergraduate research volunteers (An et al., 2017; DeRight & Jorgensen, 2015; Ross et al., 2016; Santos et al., 2014; Silk-Eglit et al., 2014). In the context of a 9% failure rate on the TOMM-1 and a 1.2–2.4% failure rate on the RCFT recognition trials, these high WCT failure rates seem to be an isolated anomaly that serve as an important reminder that external incentives fail to explain a significant amount of variance in PVT failures. In fact, since normative data for well-respected tests have not been screened for non-credible responding, instances of invalid performance can even shift normative data toward impairment and inflate error variance in clinical decision making (Erdodi, Hurtubise, et al., 2018).

Despite these general concerns, we found a comparable medium effect size for RCFT recognition trials when comparing classification accuracies of our incentivized control group and controls from the original study (Rai et al., 2019). However, the effect size for the contrast between controls and clinical patients doubled when the comparison was based on students who were motivated to do well. Such discrepancies may have important implications for high-stake

research studies (randomized clinical trials, pharmacological research). More importantly, they suggest that, however imperfect, calculated efforts to “sanitize the sample” (i.e., reduce experimental confounds) can improve data quality. Additionally, our results should sensitize research consumers to the issue of performance validity even among controls who have no apparent reason to underperform.

Clinical Applications

Our results support the use of the RCFT_{FCR} as an EVI for clinically referred patients, especially since the RCFT_{FCR} showed potential to double as a valid measure of actual visuospatial memory. The RCFT_{FCR} demonstrated superior overall psychometric properties and provided unique information about the credibility of the response set, complementing the established use of the Yes/No Recognition trial as an EVI (Blaskewitz et al., 2009; Lu et al., 2003; Sugarman et al., 2016). The RCFT_{FCR}'s low cost (open source, quick and easy to administer and score) and its potential dual-purpose make it a valuable addition to a standard neuropsychological test battery. The distribution of RCFT recognition scores revealed an important scaling artifact. Essentially, when any score ≤ 17 was considered invalid, clinicians were left with just a seven-point performance range (18-24) that was free of the “invalid before impaired paradox” (i.e., a score being deemed *invalid* before a credible deficit can be interpreted; Erdodi & Lichtenstein, 2017). Because such a restricted range does not likely allow for meaningful, demographically adjusted T-scores, we propose that clinicians adopt an interpretive trichotomy (Inferior – WNL – Superior), following differential base rates between an incentivized control group and a clinical sample (Table 6). Fortunately, because performance on RCFT recognition trials was unrelated to age, sex and level of education, raw scores in these analyses seem empirically justified. Naturally, this restricted 7-point range constrains the clinical utility of the RCFT_{FCR} as a measure of visual recognition memory. It also makes the distributional properties and therefore, the clinical interpretation of given scores vulnerable to fluctuations across samples. Although the RCFT_{FCR} may be unfit as a fine-tuned measure of memory functioning, it retains incremental validity over the Yes/No Recognition trial – both as an EVI and as an ability test. In fact, many empirically validated and widely used cognitive screening tests have similarly steep item-characteristic curves with psychometrically inactive left tails (Erdodi, Shahein, et al., 2020; Hilsabeck et al., 2015; Hoops et al., 2009). Of course, replication in larger samples is needed to determine whether these findings generalize to other populations.

Strengths, Limitations, and Directions for Future Research

By extending our investigation to clinical patients and adding an incentivized control group, we addressed several limitations of prior studies that were based on an experimental malingering paradigm (Abeare, Hurtubise, et al., 2020; Erdal, 2004; Niesten et al., 2017). We incorporated two of the same criterion PVTs from the original study (Rai et al., 2019) into the present analyses, permitting a direct comparison between past and present research. In addition, we tested two new validity composites with an engineered method variance to rigorously cross-examine RCFT recognition trials and protect against instrumentation artifacts (Erdodi, Hurtubise, et al., 2018; Erdodi, Tyson, et al., 2018).

The most salient limitation of the study was its small sample size and the composition of the clinical sample. Our predominantly white patients from a single geographic region, referred for disability evaluations may represent a unique set of sample characteristics that limit the generalizability of these findings to other populations (Kura, 2013; Leon & Leon, 2014; Lichtenstein, et al., 2019; Lynn, 2010) with different medical etiologies and incentive structures (Chafetz, 2011; Fuermaier et al., 2019; Giromini et al., 2019; Harrison, 2017; Merten & Rogers, 2017; van Helvoort et al., 2019) and different demographics (race, level of education, socio-economic status). Also, our group administration format for the RCFT within the student sample may have altered the psychometric properties of the instruments. Finally, the validation of the RCFT_{FCR} trial as a measure of visual-perceptual memory was incomplete, as we merely provided a proof of concept for the RCFT_{FCR} as an ability measure. Indeed, the absence of an independent, well-established test of visual memory with a separate recognition trial as a criterion measure is a notable limitation that should be addressed in future research. Before the instrument can be recommended for clinical use, it requires replication using larger, clinically and demographically more diverse samples against established measures of visual memory.

Conclusion

Our results suggest that previously published validity cutoffs on the RCFT_{FCR} trial maintain high specificity to psychometrically defined non-credible responding among clinical patients, and were not fully redundant with decisions made from cutoffs on the Yes/No Recognition trial. As such, the RCFT_{FCR} provides unique and relevant information for performance validity assessment. Moreover, the RCFT_{FCR} had a stronger correlation with visual-perceptual and verbal memory skills than the Yes/No Recognition trial. Even if limited to interpreting valid clinical data from a seven-point scale (i.e., scores of 18–24), both RCFT recognition trials may provide useful clinical information about cognitive functioning. Future research would benefit from further exploring the RCFT_{FCR}'s clinical utility and in its use among children (Lichtenstein

et al., 2017, 2018, 2019). Replication using geographically and demographically diverse samples with a wide range of medically verified neuropsychiatric conditions is needed to determine the generalizability of the current findings.

Ethical Approval

Relevant ethical guidelines were followed throughout the project. All data collection, storage and processing was done with the approval of relevant institutional authorities regulating research involving human participants, in compliance with the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards.

Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Kristoffer Romero  <https://orcid.org/0000-0002-3055-7112>

Laura Cutler  <https://orcid.org/0000-0003-2957-8693>

Laszlo A. Erdodi  <https://orcid.org/0000-0003-0575-9991>

References

- Abeare, C., Erdodi, L., Messa, I., Terry, D. P., Panenka, W. J., Iverson, G. L., & Silverberg, N. D. (2020). Development of embedded performance validity indicators in the NIH Toolbox Cognitive Battery. *Psychological Assessment*. Advance online publication. <https://doi.org/10.1037/pas0000958>
- Abeare, C. A., Freund, S., Kaploun, K., McAuley, T., & Dumitrescu, C. (2017). The Emotion Word Fluency Test (EWFT): Initial psychometric, validation, and physiological evidence in young adults. *Journal of Clinical and Experimental Neuropsychology*, *39*(8), 738–752. <https://doi.org/10.1080/13803395.2016.1259396>.
- Abeare, C. A., Hurtubise, J., Cutler, L., Sirianni, C., Brantuo, M., Makhzoun, N., & Erdodi, L. (2020). Introducing a forced choice recognition trial to the Hopkins Verbal Learning Test –Revised. *The Clinical Neuropsychologist*. Advance online publication. <https://doi.org/10.1080/13854046.2020.1779348>
- Abeare, C. A., Messa, I., Whitfield, C., Zuccato, B., Casey, J. & Erdodi, L. (2019). Performance validity in collegiate football athletes at baseline neurocognitive testing. *Journal of Head Trauma Rehabilitation*, *34*(4), 20–31. <https://doi.org/10.1097/HTR.0000000000000451>
- Abeare, C. A., Sabelli, A., Taylor, B., Holcomb, M., Dumitrescu, C., Kirsch, N., & Erdodi, L. (2019). The importance of demographically adjusted cutoffs: Age and

- education bias in raw score cutoffs within the Trail Making Test. *Psychological Injury and Law*, 12(2), 170–182. <https://doi.org/10.1007/s12207-019-09353>
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 3: Receiver operating characteristic plots. *British Medical Journal*, 309(6948), 188. <https://doi.org/10.1136/bmj.309.6948.188>
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). Author
- An, K. Y., Charles, J., Ali, S., Enache, A., Dhuga, J., & Erdodi, L. A. (2019). Re-examining performance validity cutoffs within the Complex Ideational Material and the Boston Naming Test-Short Form using an experimental malingering paradigm. *Journal of Clinical and Experimental Neuropsychology*, 41(1), 15–25. <https://doi.org/10.1080/13803395.2018.1483488>
- An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist*, 31(1), 193–206. <https://doi.org/10.1080/13854046.2016.1217046>
- An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, 27(8), 849–857. <https://doi.org/10.1093/arclin/acs085>
- Arentsen, T. J., Boone, K. B., Lo, T. T., Goldberg, H. E., Cottingham, M. E., Victor, T. L., Ziegler, E., & Zeller, M. A. (2013). Effectiveness of the Comalli Stroop Test as a measure of negative response bias. *The Clinical Neuropsychologist*, 27(6), 1060–1076. <https://doi.org/10.1080/13854046.2013.803603>
- Ashendorf, L., Clark, E. L., & Sugarman, M. A. (2017). Performance validity and processing speed in a VA polytrauma sample. *The Clinical Neuropsychologist*, 31(5), 857–866. <https://doi.org/10.1080/13854046.2017.1285961>
- Barhon, L. I., Batchelor, J., Meares, S., Chekaluk, E., & Shores, E. A. (2015). A comparison of the degree of effort involved in the TOMM and the ACS Word Choice Test using a dual-task paradigm. *Applied Neuropsychology: Adult*, 22(2), 114–123. <https://doi.org/10.1080/23279095.2013.863775>
- Bell-Sprinkel, T. L., Boone, K. B., Miora, D., Cottingham, M., Victor, T., Ziegler, E., Zeller, M., & Wright, M. (2013). Re-examination of the Rey Word Recognition Test. *The Clinical Neuropsychologist*, 27(3), 516–527. <https://doi.org/10.1080/13854046.2012.744853>
- Berger, C., Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2019). Detection of feigned ADHD using the MOXO-d-CPT. *Journal of Attention Disorders*, 00(0), 1–16. <https://doi.org/10.1177/1087054719864656>
- Bigler, E. D. (2012). Symptom validity testing, effort and neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 632–642. <https://doi.org/10.1017/S1355617712000252>
- Bigler, E. D. (2015). Neuroimaging as a biomarker in symptom validity and performance validity testing. *Brain Imaging and Behavior*, 9(3), 421–444. <https://doi.org/10.1007/s11682-015-9409-1>

- Blaskewitz, N., Merten, T., & Brockhaus, R. (2009). Detection of suboptimal effort with the Rey Complex Figure Test and Recognition Trial. *Applied Neuropsychology, 16*, 54–61. <https://doi.org/10.1080/09084280802644227>
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology*. Guilford.
- Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-item recognition test: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology, 24*(5), 561–573. <https://doi.org/10.1076/jcen.24.5.561.1004>
- Boskovic, I., Biermans, A. J., Merten, T., Jelacic, M., Hope, L., & Merckelbach, H. (2018). The Modified Stroop Task is susceptible to feigning: Stroop performance and symptom over-endorsement in feigned test anxiety. *Frontiers in psychology, 9*, 1195. <https://doi.org/10.3389/fpsyg.2018.01195>
- Brandt, J., & Benedict, R. H. B. (2001). *Hopkins verbal learning test—Revised*. Psychological Assessment Services.
- Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist, 23*(2), 314–328. <https://doi.org/10.1080/13854040802054151>
- Chafetz, M. D. (2011). The psychological consultative examination for Social Security Disability. *Psychological Injury and Law, 4*, 235–244. <https://doi.org/10.1007/s12207-011-9112-5>
- Conners, K. C. (2015). *Conner's Continuous Performance Test 3rd Edition (Conners CPT 3)*. Technical Manual. Multi-Health Systems.
- Cutler, L., Abeare, C., Messa, I., Holcomb, M., & Erdodi, L. A. (2021). This will only take a minute: Time cutoffs are superior to accuracy cutoffs on the Forced Choice Recognition Trial of the Hopkins Verbal Learning Test—Revised. *Applied Neuropsychology: Adult*. Advance online publication.
- Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. H. M. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 257–264. <https://doi.org/10.1080/13803395.2016.1223278>
- Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y., & Ponds, R. W. (2016). Poor symptom and performance validity in regularly referred hospital outpatients: Link with standard clinical measures, and role of incentives. *Psychiatry Research, 239*(30), 47–53. <https://doi.org/10.1016/j.psychres.2016.02.061>
- Davis, J. J. (2014). Further consideration of advanced clinical solutions word choice: Comparison to the recognition memory test—Words and classification accuracy on a clinical sample. *The Clinical Neuropsychologist, 28*(8), 1278–1294. <https://doi.org/10.1080/13854046.2014.975844>
- Delis, D. C., Kaplan, E. F., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. (2000). *The California verbal learning test—second edition*. 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. <https://doi.org/10.1093/arclin/acs044>

- DeRight, J., & Jorgensen, R. S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *The Clinical Neuropsychologist*, 29(1), 101–117.
- Donders, J., & Hayden, A. (2020). Utility of the D-KEFS color word interference test as an embedded measure of performance validity after traumatic brain injury. *The Clinical Neuropsychologist*. Advance online publication.
- Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology*, 18(2), 121–134. <https://doi.org/10.1093/arclin/18.2.121>
- Egeland, J., & Langfjæran, T. (2007). Differentiating malingering from genuine cognitive dysfunction using the Trail Making Test-ratio and Stroop Interference scores. *Applied Neuropsychology*, 14(2), 113–119. <https://doi.org/10.1080/09084280701319953>
- Eglt, G. M. L., Jurick, S. M., Delis, D. C., Filoteo, J. V, Bondi, M. W., & Jak, A. J. (2019). Utility of the D-KEFS Color Word Interference Test as an embedded measure of performance validity. *The Clinical Neuropsychologist*, 34(2), 332–352. <https://doi.org/10.1080/13854046.2019.1643923>
- Erdal, K. (2004). The effects of motivation, coaching, and knowledge of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology*, 19(1), 73–88. [https://doi.org/10.1016/S0887-6177\(02\)00214-7](https://doi.org/10.1016/S0887-6177(02)00214-7)
- Erdodi, L. A. (2019). Aggregating validity indicators: The salience of domain specificity and the indeterminate range in multivariate models of performance validity assessment. *Applied Neuropsychology: Adult*, 26(2), 155–172. <https://doi.org/10.1080/23279095.2017.1384925>
- Erdodi, L. A., & Abeare, C. A. (2020). Stronger together: The Wechsler Adult Intelligence Scale—Fourth Edition as a multivariate performance validity test in patients with traumatic brain injury. *Archives of Clinical Neuropsychology*, 35(2), 188–204. <https://doi.org/10.1093/arclin/acz032/5613200>
- Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B., T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). WAIS-IV processing speed scores as measures of non-credible responding—The third generation of embedded performance validity indicators. *Psychological Assessment*, 29(2), 148–157.
- Erdodi, L. A., Green, P., Sirianni, C., & Abeare, C. A. (2019). The myth of high false positive rates on the Word Memory Test in mild TBI. *Psychological Injury and Law*, 12(2), 155–169.
- Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. (2018). The D-KEFS Trails as performance validity tests. *Psychological Assessment*, 30(8), 1081–1095. <http://doi.org/10.1037/pas0000561>
- Erdodi, L. A., Jongsma, K. A., & Issa, M. (2017). The 15-item version of the Boston Naming test as an index of English proficiency. *The Clinical Neuropsychologist*, 31(1), 168–178. doi: 10.1080/13854046.2016.1224392
- 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. <https://doi.org/10.1007/s12207-014-9197-8>
- Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: An emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist*, 31(6-7), 1029–1046. <http://doi.org/10.1080/13854046.2017.1323119>

- Erdodi, L. A., Pelletier, C. L., & Roth, R. M. (2018). Elevations on select Conners' CPT-II scales indicate noncredible responding in adults with traumatic brain injury. *Applied Neuropsychology: Adult*, 25(1), 19–28. <https://doi.org/10.1080/23279095.2016.1232262>
- Erdodi, L. A., & Rai, J. K. (2017). A single error is one too many: Examining alternative cutoffs on Trial 2 on the TOMM. *Brain Injury*, 31(10), 1362–1368. <https://doi.org/10.1080/02699052.2017.1332386>
- Erdodi, L. A., & Roth, R. M. (2017). Low scores on BDAE Complex Ideational Material are associated with invalid performance in adults without aphasia. *Applied Neuropsychology: Adult*, 24(3), 264–274. <https://doi.org/10.1080/23279095.2017.1298600>
- 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(5), 456–466. <https://doi.org/10.1093/arclin/acu026>
- Erdodi, L. A., Sagar, S., Seke, K., Zuccato, B. G., Schwartz, E. S., & Roth, R. M. (2018). The Stroop Test as a measure of performance validity in adults clinically referred for neuropsychological assessment. *Psychological Assessment*, 30(6), 755–766. <https://doi.org/10.1037/pas0000525>
- Erdódi, L. A., Shahein, A. G., Fareez, F., Rykalski, N., Sabelli, A. G., & Roth, R. M. (2020). Increasing the cutoff on the MMSE and DRS-2 improves clinical classification accuracy in highly educated older adults. *Psychology and Neuroscience*, 13(1), 93–113. <https://doi.org/10.1037/pne0000185>
- Erdodi, L. A., Taylor, B., Sabelli, A., Malleck, M., Kirsch, N. L., & Abeare, C. A. (2019). Demographically adjusted validity cutoffs in the Finger Tapping Test are superior to raw score cutoffs. *Psychological Injury and Law*, 12(2), 113–126. <https://doi.org/10.1007/s12207-019-09352-y>
- Erdodi, L. A., Tyson, B. T., Abeare, C. A., Lichtenstein, J. D., Pelletier, C. L., Rai, J. K., & Roth, R. M. (2016). The BDAE Complex Ideational Material—A measure of receptive language or performance validity? *Psychological Injury and Law*, 9, 112–120. <https://doi.org/10.1007/s12207-016-9254-6>
- Erdodi, L. A., Tyson, B. T., Abeare, C. A., Zuccato, B. G., Rai, J. K., Seke, K. R., Sagar, S., & Roth, R. M. (2018). Utility of critical items within the Recognition Memory Test and Word Choice Test. *Applied Neuropsychology: Adult*, 25(4), 327–339. <https://doi.org/10.1080/23279095.2017.1298600>
- Erdodi, L. A., Tyson, B. T., Shahein, A., Lichtenstein, J. D., Abeare, C. A., Pelletier, C. L., Zuccato, B. G., Kucharski, B., & Roth, R. M. (2017). The power of timing: Adding a time-to-completion cutoff to the Word Choice Test and Recognition Memory Test improves classification accuracy. *Journal of Clinical and Experimental Neuropsychology*, 39(4), 369–383. <https://doi.org/10.1080/13803395.2016.1230181>
- Etherton, J. L., Bianchini, K. J., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and performance on the WAIS-III Processing Speed Index. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1218–1237. <https://doi.org/10.1080/13803390500346595>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861–874.

- Fazio, R. L., Denning, J. H., & Denney, R. L. (2017). TOMM Trial 1 as a performance validity indicator in a criminal forensic sample. *The Clinical Neuropsychologist*, *31*(1), 251–267. <https://doi.org/10.1080/13854046.2016.1213316>
- Fuermaier, A. B. M., Fricke, J. A., de Vries, S. M., Tucha, L., & Tucha, O. (2019). Neuropsychological assessment of adults with ADHD: A Delphi consensus study. *Applied Neuropsychology: Adult*, *26*(4), 340–354. <https://doi.org/10.1080/23279095.2018.1429441>
- Fuermaier, A. B. M., Tucha, O., Koerts, J., Lange, K. W., Weisbrod, M., Aschenbrenner, S., & Tucha, L. (2017). Noncredible cognitive performance at clinical evaluation of adult ADHD: An embedded validity indicator in a visuospatial working memory test. *Psychological Assessment*, *29*(12), 1466–1479. <https://doi.org/10.1037/pas0000534>
- Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2019). Using the inventory of problems—29 (IOP-29) with the Test of Memory Malingered (TOMM) in symptom validity assessment: A study with a Portuguese sample of experimental feigners. *Applied Neuropsychology: Adult*. Advanced online publication. <https://doi.org/10.1080/23279095.2019.1570929>
- Giromini, L., Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Perfetti, B., di Carlo, D. A., & Zennaro, A. (2019). Beyond rare-symptom endorsement: A clinical comparison study using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law*, *12*(3-4), 212–224. <https://doi.org/10.1007/s12207-019-09357-7>
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity and IOP-29 with an Italian sample. *Psychological Injury and Law*, *12*(3-4), 212–224. <https://doi.org/10.1007/s12207-018-9314-1>
- Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, *6*(2), 147–178.
- Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2019). Examining false-positive rates of Wechsler Adult Intelligence Scale (WAIS-IV) processing speed based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment*, *31*(1), 120–125. [10.1037/pas0000650](https://doi.org/10.1037/pas0000650)
- Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination* (3rd ed.). Lippincott Williams & Wilkins.
- Goworowski, L., Vagt, D., Salazar, C., Mulligan, K., & Webbe, F. (2020). Normative values of the Rey Word Recognition Test in college athletes. *Applied Neuropsychology: Adult*, *27*(1), 94–97. <https://doi.org/10.1080/23279095.2018.1488716>
- Grant, A. F., Lace, J. W., Teague, C. L., Lowell, K. T., Ruppert, P. D., Garner, A. A., & Gfeller, J. D. (2020). Detecting feigned symptoms of depression, anxiety, and ADHD, in college students with the Structured Inventory of Malingered Symptomatology. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1769097>
- Greve, K. W., Bianchini, K. J., Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingered in traumatic brain injury: Results of a known-group analysis. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1176–1190. <https://doi.org/10.1080/13803390500263550>

- Greve, K. W., Etherton, J. L., Ord, J., Bianchini, K. J., & Curtis, K. L. (2009). Detecting malingered pain-related disability: Classification accuracy of the Test of Memory Malingering. *The Clinical Neuropsychologist*, *23*(7), 1250–1271. <https://doi.org/10.080/13854040902828272>
- Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., . . . Conference Participants. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist*, *34*(3), 437–453.
- Guise, B. J., Thompson, M. D., Greve, K. W., Bianchini, K. J., & West, L. (2014). Assessment of performance validity in the Stroop Color and Word Test in mild traumatic brain injury patients: A criterion-groups validation design. *Journal of Neuropsychology*, *8*(1), 20–33. <https://doi.org/10.1111/jnp.12002>
- Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123. <https://doi.org/10.1007/s10994-009-5119-5>
- Harrison, A. G. (2017). Clinical, ethical and forensic implications of a flexible threshold for LD and ADHD in postsecondary settings. *Psychological Injury and Law*, *10*, 138–150.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Psychological Assessment Resources.
- Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*(5), 892–900. <https://doi.org/10.1037/0022-006X.46.5.892>
- Hilsabeck, R. C., Holdnack, J. A., Munro Cullum, C., Drozdick, L. W., Edelstein, B., Fiske, A., & Wahlstrom, D. (2015). The Brief Cognitive Status Examination (BCSE): Comparing diagnostic utility and equating scores to the Mini-Mental State Examination (MMSE). *Archives of Clinical Neuropsychology*, *30*(5), 216–219.
- Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745. <https://doi.org/10.1212/WNL.0b013e3181c34b47>
- Hurtubise, J. L., Baher, T., Messa, I., Cutler, L., Shahein, A., Hastings, M., Carignan-Querqui, M., & Erdodi, L. (2020). Verbal fluency and Digit Span variables as performance validity indicators in experimentally induced malingering and real world patients with TBI. *Applied Neuropsychology: Child*. Advanced online publication. <https://doi.org/10.1080/21622965.2020.1719409>
- Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology*, *17*, 1–23. <https://doi.org/10.1093/arclin/17.1.1>
- Jelicic, M., Ceunen, E., Peters, M. J., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, *67*(9), 850–855. <https://doi.org/10.1002/jclp.20805>

- Jones, A. (2013). Test of Memory Malingering: Cutoff scores for psychometrically defined malingering groups in a military sample. *The Clinical Neuropsychologist*, 27(6), 1043–1059. <https://doi.org/10.1080/13854046.2013.804949>
- Kim, N., Boone, K. B., Victor, T., Lu, P., Keatinge, C., & Mitchell, C. (2010). Sensitivity and specificity of a digit symbol recognition trial in the identification of response bias. *Archives of Clinical Neuropsychology*, 25(5), 420–428. <https://doi.org/10.1903/arclin/acq040>
- 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. <https://doi.org/10.1007/s12207-014-9200-4>
- Kura, K. (2013). Japanese north–south gradient in IQ predicts differences in stature, skin color, income, and homicide rate. *Intelligence*, 41(5), 512–516. <https://doi.org/10.1016/j.intell.2013.07.001>
- Lace, J. W., Grant, A. F., Kosky, K. M., Teague, C. L., Lowell, K. T., & Gfeller, J. D. (2020). Identifying novel embedded performance validity test formulas within the repeatable battery for the assessment of neuropsychological status: A simulation study. *Psychological Injury and Law*, 13, 303–315. <https://doi.org/10.1007/s12207-020-09382-x>
- Lafayette Instrument. (2015). *Grooved Pegboard user manual*. Lafayette, IN.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17(3), 410–425. <https://doi.org/10.1076/clin.17.3.410.18089>
- Lau, L., Basso, M. R., Estevis, E., Miller, A., Whiteside, D. M., Combs, D., & Arentsen, T. J. (2017). Detecting coached neuropsychological dysfunction: A simulation experiment regarding mild traumatic brain injury. *The Clinical Neuropsychologist*, 31(8), 1412–1431. <https://doi.org/10.1080/13854046.2017.1318954>
- Leighton, A., Weinborn, M., & Maybery, M. (2014). Bridging the gap between neurocognitive processing theory and performance validity assessment among the cognitively impaired: A review and methodological approach. *Journal of the International Neuropsychological Society*, 20, 873–886. <https://doi.org/10.1017/S135561771400085X>
- León, F. R., & León, A. B. (2014). Why complex cognitive ability increases with absolute latitude. *Intelligence*, 46, 291–299. <https://doi.org/10.1016/j.intell.2014.07.011>
- Lezak, M. D. (1995). *Neuropsychological assessment*. Oxford University Press.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment*. Oxford University Press.
- Lichtenstein, J. D., Erdodi, L. A., & Linnea, K. S. (2017). Introducing a forced-choice recognition task to the California Verbal Learning Test—Children’s Version. *Child Neuropsychology*, 23(3), 284–299. <https://doi.org/10.1080/09297049.2015.1135422>
- Lichtenstein, J. D., Erdodi, L. A., Rai, J. K., Mazur-Mosiewicz, A. & Flaro, L. (2018). Wisconsin Card Sorting Test embedded validity indicators developed for adults can be extended to children. *Child Neuropsychology*, 24(2), 247–260. <https://doi.org/10.1080/09297049.2016.1259402>
- Lichtenstein, J. D., Flaro, L., Baldwin, F., Rai, J. K., & Erdodi, L. A. (2019). Further evidence for embedded validity tests in children within the Conners’ Continuous Performance Test—Second Edition. *Developmental Neuropsychology*, 44(2), 159–171. <https://doi.org/10.1080/87565641.2019.1565536>

- Lichtenstein, J. D., Greenacre, M. K., Cutler, L., Abeare, K., Baker, S. D., Kent, K., J., Ali, S., & Erdodi, L. A. (2019). Geographic variation and instrumentation artifacts: In search of confounds in performance validity assessment in adults with mild TBI. *Psychological Injury and Law, 12*(2), 127–145. <https://doi.org/10.1007/s12207-019-0935>
- Lichtenstein, J. D., Holcomb, M., & Erdodi, L. A. (2018). One-Minute PVT: Further evidence for the utility of the California Verbal Learning Test—Children’s Version Forced Choice Recognition Trial. *Journal of Pediatric Neuropsychology, 4*(3-4), 94–104. <https://doi.org/10.1007/s40817-018-0057-4>
- Lindstrom, W., Coleman, C., Thomassin, K., Southall, C. M., & Lindstrom, J. H. (2011). Simulated dyslexia in postsecondary students: Description and detection using embedded validity indicators. *The Clinical Neuropsychologist, 25*(2), 302–322. <https://doi.org/10.1080/13854046.2010.537280>
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*(3), 391–421. <https://doi.org/10.1080/13854046.2017.1406146>
- Lobo, J. M., Jimenez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography, 17*(2), 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Love, C. M., Glassmire, D. M., Zanolini, S. J., & Wolf, A. (2014). Specificity and false positive rates of the Test of Memory Malingering, Rey 15-Item Test, and Rey Word Recognition Test among forensic inpatients with intellectual disabilities. *Assessment, 21*(5), 618–627. <https://doi.org/10.1177/1073191114528028>
- Lu, P. H., Boone, K. B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers recognition trial in the detection of suspect effort. *The Clinical Neuropsychologist, 17*(3), 426–440. <https://doi.org/10.1076/clin.17.3.426.18083>
- Lynn, R. (2010). In Italy, north–south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence, 38*(1), 93–100. <https://doi.org/10.1016/j.intell.2009.07.004>
- Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist, 34*(1), 88–119. <https://doi.org/10.1080/13854046.2019.1637027>
- Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting, 19*(6), 1106–1114. <https://doi.org/10.1175/825.1>
- McWhirter, L., Sargent, B., Ritchie, C., Stone, J., & Carson, A. (2019). I think, therefore I forget—Using experimental simulation of dementia to understand functional cognitive disorders. *CNS Spectrums, 1*–8. <https://doi.org/10.1017/S1092852919001329>
- Merten, T., Green, P., Henry, M., Blaskewitz, N., & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology, 20*(6), 719–726. <https://doi.org/10.1016/j.acn.2005.04.004>
- Merten, T., & Merckelbach, H. (2013). Forced-choice tests as single-case experiments in the differential diagnosis of intentional symptom distortion. *Journal of Experimental Psychopathology, 4*(1), 20–37. <https://doi.org/10.5127/jep.023711>

- Merten, T., & Rogers, R. (2017). An international perspective on feigned mental disabilities: Conceptual issues and continuing controversies. *Behavioral Sciences and the Law*, 35(2), 97–112. <https://doi.org/10.1002/bsl.2274>
- Messa, I., Holcomb, M., Lichtenstein, J., Tyson, B., Roth, R., & Erdodi, L. (2020). They are not destined to fail: A systematic examination of scores on embedded performance validity indicators in patients with intellectual disability. *Australian Journal of Forensic Sciences*. Advance online publication. <https://doi.org/10.1080/00450618.2020.1865457>
- Meyers, J. E., & Meyers, K. R. (1995). *Rey complex figure test and recognition trial: Professional manual*. Psychological Assessment Corporation.
- Nielsen, I. J. M., Muller, W., Merckelbach, H., Dandachi-FitzGerald, B., & Jelicic, M. (2017). Moral reminders do not reduce symptom over-reporting tendencies. *Psychological Injury and Law*, 10(4), 368–384. <https://doi.org/10.1007/s12207-0179303-9>
- Nitch, S., Boone, K. B., Wen, J., Arnold, G., & Alfano, K. (2006). The utility of the Rey Word Recognition Test in the detection of suspect effort. *The Clinical Neuropsychologist*, 20(4), 873–887. <https://doi.org/10.1080/13854040590967603>
- O'Bryant, S. E., Hilsabeck, R. C., Fisher, J. D., & McCaffrey, R. J. (2003). Utility of the Trail Making Test in the assessment of malingering in a sample of mild traumatic brain injury litigants. *The Clinical Neuropsychologist*, 17(1), 69–74. <https://doi.org/10.1076/clin.17.1.69.15624>
- Ord, A. S., Miskey, H. M., Lad, S., Richter, B., Nagy, K., & Shura, R. D. (2020). Examining embedded validity indicators in Connors continuous performance test-3 (CPT-3). *The Clinical Neuropsychologist*, 1–16. <https://doi.org/10.1080/13854046.2020.1751301>
- Pearson, N. C. S. (2009). *Advanced clinical solutions for WAIS-IV and WMS-IV: Administration and scoring manual*. The Psychological Corporation.
- Poreh, A., Bezdicek, O., Korobkova, I., Levin, J. B., & Dines, P. (2016). The Rey Auditory Verbal Learning Test forced-choice recognition task: Base-rate data and norms. *Applied Neuropsychology: Adult*, 23(3), 155–161. <https://doi.org/10.1080/23279095.2015.1027343>
- Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached malingerers with the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, 19(5), 693–702. <https://doi.org/10.1016/j.acn.2004.04.001>
- Poynter, K., Boone, K. B., Ermshar, A., Miora, D., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2019). Wait, there's a baby in this bath water! Update on quantitative and qualitative cut-offs for Rey 15-Item Recall and Recognition. *Archives of Clinical Neuropsychology*, 34(8), 1367–1380. <https://doi.org/10.1093/arclin/acy087>
- Rai, J., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choice recognition trial to the Rey Complex Figure Test. *Psychology and Neuroscience*, 12(4), 451–472. <https://doi.org/10.1037/pne0000175>
- Rai, J., & Erdodi, L. (2019). The impact of criterion measures on the classification accuracy of TOMM-1. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2019.161.1613994>
- Reedy, S. D., Boone, K. B., Cottingham, M. E., Glaser, D. F., Lu, P. H., Victor, T. L., Ziegler, E., Zeller, M., & Wright, M. J. (2013). Cross validation of the Lu and

- colleagues (2003) Rey-Osterrieth Complex Figure Test effort equation in a large known-group sample. *Archives of Clinical Neuropsychology*, 28, 30–37. <https://doi.org/10.1093/arclin/acs106>
- Reese, C. S., Suhr, J. A., & Riddle, T. L. (2012). Exploration of malingering indices in the Wechsler Adult Intelligence Scale—Fourth Edition Digit Span subtest. *Archives of Clinical Neuropsychology*, 27(2), 176–181. <https://doi.org/10.1093/arclin/arc117>
- Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*, 19(5), 393–394. <https://doi.org/10.1037/h0044509>
- Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique [Psychological examination in cases of traumatic encephalopathy]. *Archives de Psychologie*, 28, 286–340.
- Roberson, C. J., Boone, K. B., Goldberg, H., Miora, D., Cottingham, M., Victor, T., Ziegler, E., Zeller, M., & Wright, M. (2013). Cross validation of the b Test in a large known groups sample. *The Clinical Neuropsychologist*, 27(3), 495–508. <https://doi.org/10.1080/13854046.2012.737027>
- Roor, J. J., Dandachi-FitzGerald, B., & Ponds, R. W. M. (2016). A case of misdiagnosis of mild cognitive impairment: The utility of symptom validity testing in the out-patient memory clinic. *Applied Neuropsychology: Adult*, 23(3), 172–178. <https://doi.org/10.1080/23279095.2015.1030018>
- Ross, T. P., Poston, A. M., Rein, P. A., Salvatore, A. N., Wills, N. L., & York, T. M. (2016). Performance invalidity base rates among healthy undergraduate research participants. *Archives of Clinical Neuropsychology*, 31(1), 97–104.
- Roth, R. M., Gioia, G. A., & Isquith, P. K. (2005). *BRIEF-A: Behavior Rating Inventory of Executive Function—Adult Version*.
- Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C., & McGuire, K. (1992). Quantitative and qualitative analyses of clock face drawings in Alzheimer's and Huntington's diseases. *Brain and Cognition*, 18(1), 70–87. [https://doi.org/10.1016/0278-2626\(92\)90112-Y](https://doi.org/10.1016/0278-2626(92)90112-Y)
- Roye, S., Calamia, M., Bernstein, J. P., De Vito, A. N., & Hill, B. D. (2019). A multi-study examination of performance validity in undergraduate research participants. *The Clinical Neuropsychologist*, 33(6), 1138–1155. <https://doi.org/10.1080/13854046.2018.1520303>
- Rüsseler, J., Brett, A., Klaue, U., Sailer, M., & Münte, T. F. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology*, 8(1). <https://doi.org/10.1186/1471-2377-8-37>
- Santos, O. A., Kazakov, D., Reamer, M. K., Park, S. E., & Osmon, D. C. (2014). Effort in college undergraduate is sufficient on the Word Memory Test. *Archives of Clinical Neuropsychology*, 29, 609–613.
- Schroeder, R. W., Martin, P. K., Heindrichs, R. J., & Baade, L. E. (2019). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist*, 33(3), 466–477. <https://doi.org/10.1080/13854046.2018.1484517>
- Shin, M. S., Park, S. Y., Park, S. R., Seol, S. H., & Kwon, J. S. (2006). Clinical and empirical applications of the Rey-Osterrieth Complex Figure Test. *Nature Protocols*, 1(2), 892–899. <https://doi.org/10.1038/nprot.2006.115>

- Shura, R. D., Martindale, S. L., Taber, K. H., Higgins, A. M., & Rowland, J. A. (2019). Digit Span embedded validity indicators in neurologically-intact veterans. *The Clinical Neuropsychologist*, 34(5), 1025–1037. <https://doi.org/10.1080/13854046.2019.1635209>
- Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gatz, R. E., Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, 23(2), 94–104. <https://doi.org/10.1080/23279095.2015.1014556>
- 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.
- Sirianni, C. D., Abeare, C. A., Ali, S., Razvi, P., Kennedy, A., Pyne, S. R., & Erdodi, L. A. (2021). The V-5 provides quick and accurate measures of psychiatric symptoms for single and repeated in-person and online assessments and is valid in culturally diverse examinees. *Psychiatry Research*. Advance online publication. <https://doi.org/10.1016/j.psychres.2020.113651>
- Slick, D. J., Sherman, E. M. S., & 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\(99911\)13:04;1-YFT545](https://doi.org/10.1076/1385-4046(99911)13:04;1-YFT545)
- Smith, K., Boone, K., Victor, T., Miora, D., Cottingham, M., Ziegler, E., Zeller, M., & Wright, M. (2014). Comparison of credible patients of very low intelligence and non-credible patients on neurocognitive performance validity indicators. *The Clinical Neuropsychologist*, 28(6), 1048–1070. <https://doi.org/10.1080/13854046.2014.931465>
- 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. <https://doi.org/10.1080/13854046.2013.845248>
- Spitzer, R. L., Kroenke, K., Williams, J.B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Achieves of Internal Medicine*, 166(10), 1092–1097.
- Spitzer, R. L., Kroenke, K., & Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*, 282(18), 1737–1744.
- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, 157(1-3), 191–200. <https://doi.org/10.1016/j.psychres.2007.01.003>
- 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. <https://doi.org/10.1080/23279095.2013.873439>
- Sugarman, M. A., Holcomb, E. M., Axelrod, B. N., Meyers, J. E., & Liethen, P. C. (2016). Embedded measures of performance validity in the Rey complex figure test in a clinical sample of veterans. *Applied Neuropsychology: Adult*, 23(2), 105–114. <https://doi.org/10.1080/23279095.2015.1014557>

- Sullivan, K., & King, J. (2010). Detecting faked psychopathology: A comparison of two tests to detect malingered psychopathology using a simulation design. *Psychiatry Research* 176, 75–81.
- Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, 16 (4), 495–5050. <https://doi.org/10.1076/clin.16.4.495.13909>
- Tombaugh, T. N. (1996). *Test of Memory Malingering*. Multi-Health Systems.
- Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology*, 14(4), 697–607. <https://doi.org/10.1080/01688639408402671>
- Tyson, B. T., Baker, S., Greenacre, M., Kent, K., J., Lichtenstein, J. D., Sabelli, A., & Erdodi, L.A. (2018). Differentiating epilepsy from psychogenic nonepileptic seizures using neuropsychological test data. *Epilepsy & Behavior*, 87, 39–45.
- van Helvoort, D., Merckelbach, H., & Merten, T. (2019). The Self-Report Symptom Inventory (SRSI) is sensitive to instructed feigning, but not genuine psychopathology in male forensic inpatients: An initial study. *The Clinical Neuropsychologist*, 33(6), 1069–1082. <https://doi.org/10.1080/13854046.2018.1559359>
- Viglione, D. J., Giromini, L., Landis, P., McCullaugh, J. M., Pizitz, T. D., O'Brien, S., Wood, S., Connell, K., & Abramsky, A. (2019). Development and validation of the False Disorder Score: The Focal Scale of the Inventory of Problems–29. *Journal of Personality Assessment*, 101(6), 653–661. <https://doi.org/10.1080/00223891.2018.1492413>
- Wald, N. J., & Bestwick, J. P. (2014). Is the area under an ROC curve a valid measure of the performance of a screening or diagnostic test? *Journal of Medical Screening*, 21(1), 51–56. <https://doi.org/10.1177/0969141313517497>
- Webber, T. A., Bailey, K. C., Alverson, W. A., Critchfield, E. A., Bain, K. M., Messerly, J. M., O'Rourke, J., J., F., Kirton, J., W., Fullen, C., Marceaux, J., C., & Soble, J. R. (2018). Further validation of the Test of Memory Malingering (TOMM) Trial 1 performance validity index: Examination of false positives and convergent validity. *Psychological Injury and Law*, 11(4), 325–335. <https://doi.org/10.1007/s12207-018-9335-9>
- Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance among performance validity among cognitively impaired and unimpaired examinees. *The Clinical Neuropsychologist*, 32 (4), 657–670. <https://doi.org/10.1080/13854046.2017.1415374>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Test — Fourth Edition (WAIS-IV)*. Pearson.
- Weinborn, M., Woods, S. P., Nulsen, C., & Leighton, A. (2012). The effects of coaching on the verbal and nonverbal medical symptom validity tests. *The Clinical Neuropsychologist*, 26(5), 832–849. <https://doi.org/10.1080/13854046.2012.686630>
- Whitney, K. A., Davis, J. J., Shepard, P. H., Bertram, D. M., & Adams, K. M. (2009). Digit span age scaled score in middle-aged military veterans: Is it more closely associated with TOMM failure than reliable digit span?. *Archives of Clinical Neuropsychology*, 24(3), 263–272. <https://doi.org/10.1093/arclin/acp034>
- Whitney, K. A., Shepard, P. H., & Davis, J. J. (2013). WAIS-IV Digit Span variables: Are they valuable for use in predicting TOMM and MSVT failure? *Applied Neuropsychology: Adult*, 20(2), 83–94. <https://doi.org/10.1080/09084282.2012.670167>

- Whiteside, D., Wald, D., & Busse, M. (2011). Classification accuracy of multiple visual spatial measures in the detection of suspect effort. *The Clinical Neuropsychologist*, 25, 287–301. <https://doi.org/10.1080/13854046.2010.538436>
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15± 15%. *Psychological Injury and Law*, 8(3), 188–199. <https://doi.org/10.1007/s12207-015-9232-4>
- Zuccato, B. G., Tyson, B. T., & Erdodi, L. A. (2018). Early bird fails the PVT? The effects of timing artifacts on performance validity tests. *Psychological Assessment*, 30(11), 1491–1498. <https://doi.org/10.1037/pas0000596>

Author Biographies

Kaitlyn Abeare is a student of psychology at the University of Windsor. Her research interests include performance validity, ecological validity in assessment, and executive functioning in individuals with neurodegenerative diseases.

Kristoffer Romero is a neuropsychologist and an Assistant Professor in the Department of Psychology at the University of Windsor. His research interests include metamemory, episodic memory, technology use in neuropsychology, and racial/ethnic diversity in neuropsychology.

Laura Cutler is a graduate student in the Clinical Psychology Program at the University of Windsor. Her research interests include performance validity assessment, the short-term cognitive effects of cannabis consumption, and the efficacy of remote onscreen neuropsychological assessments.

Christina D. Sirianni is a recent graduate from the Master of Social Work (MSW) program at the University of Windsor. Her research and clinical interests include child psychopathology, bio-psycho-social assessment, as well as the relationship between trauma history and various executive functions.

Laszlo A. Erdodi is a clinical neuropsychologist and Associate Professor in the Department of Psychology at the University of Windsor. His research focuses on performance validity assessment, the interaction between emotional and cognitive functioning and cross-cultural neuropsychology.