8-31-2018

Base rates of failure and classification accuracy of various performance validity tests administered in a medical-legal setting

Jaspreet Kaur Rai
University of Windsor

BASE RATES OF FAILURE AND CLASSIFICATION ACCURACY OF VARIOUS PERFORMANCE VALIDITY TESTS ADMINISTERED IN A MEDICAL-LEGAL SETTING

by

Jaspreet Kaur Rai

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

BASE RATES OF FAILURE AND CLASSIFICATION ACCURACY OF VARIOUS PERFORMANCE VALIDITY TESTS ADMINISTERED IN A MEDICAL-LEGAL SETTING

by

Jaspreet Kaur Rai

APPROVED BY:

_____

K. Boone, External Examiner
The University of California, Los Angeles

_____

D. Kane
Faculty of Nursing

_____

C. Abeare
Department of Psychology

_____

A. Baird
Department of Psychology

_____

L. Erdodi, Advisor
Department of Psychology

August 2, 2018

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this dissertation and that no part of this dissertation has been published or submitted for publication.

I certify that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my dissertation and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

In neuropsychological assessment, performance validity tests (PVTs) are used to assess whether patients' performances on cognitive testing represent their true ability levels. Higher base rates of PVT failure (BR_{Fail}) have been consistently found in the context of external incentives to appear impaired (e.g., medical-legal settings) as compared to settings without such incentives (e.g., clinical referrals). Despite the exponential growth of interest and empirical research on PVTs within the past decades, most studies have had relatively limited sample sizes and focused on a small number of instruments and clinical populations. To address this void in large-scale research, this dissertation aimed to characterize performance on 14 PVTs of interest (3 free-standing PVTs and 11 embedded validity indicators) in a large sample of adults assessed in a predominantly medical-legal setting ($N = 4{,}721$). Specifically, BR_{Fail} were reported as a function of several patient characteristics (i.e., diagnosis, age, education, gender, and English language status) in the overall sample (Study 1) and in a subsample demonstrating valid performance on an external criterion PVT (the Word Memory Test [WMT], Study 2). Classification accuracies for various cutoffs on each PVT of interest were also investigated against the WMT (Study 3). In Studies 1 and 2, free-standing PVTs tended to be more robust to the effects of patient characteristics than embedded validity indicators. In Study 3, the majority of previously published cutoffs for PVTs of interest achieved acceptable specificity, with three isolated exceptions. Free-standing PVTs demonstrated better classification accuracy than embedded validity indicators, although no PVT achieved perfect classification accuracy. Taken together, the results of this dissertation highlight the importance of using multiple PVTs and interpreting individual PVT scores in the context of patient characteristics rather than by rigidly adhering to omnibus cutoffs.

DEDICATION

To my maternal grandfather, Kulwant Singh Dhaliwal (1930 – 2012), for his genuine interest and enthusiastic encouragement of my academic and professional pursuits.

Thank you for believing in my dreams, Papa.

I wish I could have shared this day with you.

ACKNOWLEDGEMENTS

First and foremost, I thank the Lord. All that I am and all that I have are Your blessings, and I am so grateful to have had the opportunity to pursue my dreams in this life.

I extend my deepest appreciation to my doctoral research supervisor, Dr. Laszlo Erdodi, for fostering my interest in performance validity research, helping to develop and refine my research skills, and for always encouraging me to think critically (cynically, even) as both a producer and consumer of research. It is only because of you that I feel confident as a researcher today – and for that, I will always be indebted to you.

I express my sincerest gratitude to Drs. Paul Green (retired Clinical Neuropsychologist) and Roger Gervais (Clinical Psychologist, Edmonton, Alberta, Canada) for sharing their research databases with me for this dissertation and for continually inspiring the scientist-practitioner in me through their lifelong commitments to research and evidence-based practice.

I thank my committee members, Drs. Chris Abeare, Anne Baird, and Debbie Kane, for persevering through the painstaking number of tables in my dissertation(!) and for offering insightful feedback and suggestions throughout the development of this document. To Dr. Kyle Boone, it was an absolute honour to have you as the External Examiner for my defense.

I am eternally grateful to the Department of Psychology at the University of Windsor and the practicum sites at which I trained (in Windsor, Detroit, and London) for providing me with the solid theoretical foundation and clinical skills necessary to become a competent clinical neuropsychologist. I will always look back fondly on my time in Windsor and feel proud to have come from this program.

Finally, to my family: this journey has been demanding, stressful, and exhausting at the best of times, and I truly could not have done it without your support and patience. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

ACS	Advanced Clinical Solutions (Pearson, 2009)
ANX	Anxiety
BR_{Fail}	Base rate of PVT failure/invalid performance
CNS	Consistency between responses on IR and DR (score on WMT, MSVT, and NV-MSVT)
CP/F	Chronic Pain/Fibromyalgia
CT	Category Test
CT-TE	Category Test Total Errors score
DCT	Dot Counting Test
DEP	Depression
DR	Delayed Recognition (subtest on WMT, MSVT, NV-MSVT)
DRA	Delayed Recognition Archetypes (subtest on NV-MSVT)
DRV	Delayed Recognition Variations (subtest on NV-MSVT)
ESL	English as a second language
FIT	Rey 15-Item test
FMS	Failure to Maintain Set (score on WCST)
FR	Free Recall (subtest on WMT, MSVT)
FTT	Finger Tapping Test
FTT-C	Sum of average taps achieved with dominant and non-dominant hands on FTT
FTT-DH	Average taps achieved with dominant hand on FTT
FTT-DIFF	Difference between taps achieved with dominant and non-dominant hands on FTT
FTT-NDH	Average taps achieved with non-dominant hand on FTT

GMIP	Genuine memory impairment profile (on WMT, MSVT, NV-MSVT)
IR	Immediate Recognition (subtest on WMT, MSVT, NV-MSVT)
M-S	Moderate to severe TBI
MC	Multiple Choice (subtest on WMT)
MND	Malingered Neurocognitive Dysfunction
MSVT	Medical Symptom Validity Test
mTBI	Mild traumatic brain injury
NEU	Neurological
NPP	Negative predictive power
NSE	Native speakers of English
NV-MSVT	Non-Verbal Medical Symptom Validity Test
OTH	Other
ORT	Orthopedic
PA	Paired Associates (subtest on WMT, MSVT)
PDRT	Portland Digit Recognition Test
PPP	Positive predictive power
PVT	Performance validity test
RDS	Reliable Digit Span
SENS	Sensitivity
SPEC	Specificity
SMI	Severe mental illness
SVT	Symptom validity test
TBI	Traumatic brain injury

TOMM	Test of Memory Malingering
TMT	Trail Making Test
TMT-A	Time to completion on Part A of TMT
TMT A+B	Sum of times to completion on Parts A and B of TMT
TMT-B	Time to completion on Part B of TMT
TMT-B/A	Ratio of time taken to complete TMT Part B versus TMT Part A
VSVT	Victoria Symptom Validity Test
WAIS	Wechsler Adult Intelligence Scale
WCST	Wisconsin Card Sorting Test
WMT	Word Memory Test

CHAPTER 1

Introduction

Neuropsychological assessment involves the administration of psychometric tests to evaluate cognitive and emotional functioning. The test results, in turn, are integrated with other relevant data (e.g., behavioural observations, clinical history) in order to answer one or more referral questions (Strauss, Sherman, & Spreen, 2006). Neuropsychologists can make valuable contributions in clinical, medical-legal, and forensic settings using these methods. However, the major assumption underlying the use of psychometric measures in neuropsychology is that there is a relationship between an examinee's behaviour, as measured by neuropsychological tests, and the condition of the brain (Lezak, Howieson, & Loring, 2004).

Any non-neurological factor that can affect performance on neuropsychological tests and thereby produce a skewed representation of the underlying brain-behaviour relationship can be considered a threat to the validity of neuropsychological test data (Greiffenstein, 2008). One such threat to validity that has received a great deal of attention in the research literature over the past few decades is invalid responding (Larrabee, 2000; Reynolds, 1998; Sweet, 1999). Invalid responding can manifest as exaggeration or fabrication of reported symptoms and/or the display of reduced abilities on cognitive performance measures (Larrabee, 2003; Heilbronner et al., 2009).

It is imperative to detect invalid responding on neuropsychological testing in order to maximize the validity of the test results and resulting conclusions. However, several studies have shown that neuropsychologists are inaccurate when asked to discriminate valid from invalid performance based on clinical judgment or scores on ability tests alone (e.g., Ekman, O'Sullivan, & Frank, 1999; Faust, 1995; Faust, Hart, Guilmette, & Arkes, 1988). Psychometric measures, on

the other hand, have proven to be quite effective in determining response validity (e.g., Larrabee, 2003): symptom validity tests (SVTs) assess the extent to which the symptoms endorsed by examinees on a self-report measure reflect their true experience, while performance validity tests (PVTs) evaluate the degree to which examinees' performance on cognitive tests reflects their true abilities (Larrabee, 2012). PVTs are further classified as free-standing (i.e., tests that were developed specifically for the purpose of evaluating performance validity and often do not assess any other cognitive functions) or embedded (i.e., indices derived from ability tests, which have been found to be useful in discriminating valid from invalid performance).

This dissertation consisted of three studies. Study 1 aimed to characterize performance on a total of 14 PVTs of interest (three free-standing and 11 embedded) in a large sample of patients assessed in a primarily medical-legal setting. Specifically, base rates of failure (BR_{Fail}) were reported at various cutoffs on each PVT of interest in the overall sample, as a function of performance on a well-validated criterion PVT (Green's Word Memory Test or WMT; Green, 2003), and as a function of five patient variables (i.e., diagnosis, age, education level, gender, English language status). Although Study 1 represented an important first step, considering the estimated 30% base rate of invalid performance for psychological assessment settings involving a mix of clinical and medical-legal referrals (Mittenberg, Patton, Canyock, & Condit, 2002), invalid performance represented a significant confound in Study 1 in terms of determining the true relationships between BR_{Fail} and patient variables. For example, if higher BR_{Fail} were to be observed among less educated examinees as compared to more educated examinees on a given PVT in Study 1, this finding could reflect (a) higher rates of invalid performance among the former group, (b) false positive errors, wherein the patients were performing credibly but failed the published cutoffs due to their low education levels, or (c) a combination of these outcomes.

Considering this, in Study 2, a subsample was created by retaining only those cases that passed the WMT at standard cutoffs, and data on the 14 PVTs of interest were analyzed in a similar manner to Study 1. Thus, Study 2 investigated BR_{Fail} on the 14 PVTs of interest in large sample of patients who had demonstrated clearly valid performance on an external criterion PVT. Finally, in Study 3, classification accuracy statistics were computed for various cutoff scores on each PVT of interest, again using the WMT as the criterion PVT.

The discussion below summarizes the existing literature that is relevant to the present dissertation. Topics discussed include: (1) an introduction to neuropsychological assessment and performance validity, (2) methods for detecting invalid responding, (3) use of PVTs in clinical practice including a brief discussion of malingering and the issue of intent, and (4) the PVT research base, its implications for practice, and its limitations.

Neuropsychological assessment and validity

Neuropsychological assessment involves the integration of scores on standardized measures of cognitive performance with historical, neurological, psychiatric, medical, behavioural, and other relevant information for the purpose of answering one or more referral questions. Cognitive abilities that can be assessed in the course of a neuropsychological assessment include general intellectual functioning, executive functions, attention, memory, language, visual perception, sensory function (including somatosensory, olfactory, and body orientation), and motor function (Strauss et al., 2006). Scores on cognitive tests are often supplemented with self-report questionnaires or rating scales assessing an examinee's mood, personality, and/or adaptive functioning. The results of psychometric testing are combined and interpreted in the context of the examinee's clinical history to answer questions about diagnosis, patient care (management and planning), treatment (identifying treatment needs and/or

evaluating treatment efficacy), or medical-legal/forensic issues (Lezak, Howieson, Bigler, & Tranel, 2012).

A major tenet underlying the practice of neuropsychology is that there is a relationship between an examinee's behaviour, as measured by neuropsychological tests, and the condition of the brain (Lezak et al., 2004). In order to arrive at accurate diagnoses and make appropriate treatment recommendations, therefore, the clinician must first establish that the neuropsychological test data are valid—that they accurately represent the underlying brain-behaviour relationships (Lezak et al., 2012; Suhr & Gunstad, 2002).

Threats to the validity of neuropsychological test data. Non-neurological factors that affect an examinee's performance on one or more neuropsychological tests, and thereby result in a misleading picture of the brain-behaviour relationship, represent threats to the validity of neuropsychological test data (Greiffenstein, 2008). While some threats to validity concern the assessor (e.g., professional competence, selection and use of assessment measures; Larrabee, 2012), most threats relate to the examinee. These may include physical problems (e.g., vision or hearing deficits, difficulties with motor output), inattentiveness secondary to various causes, psychiatric conditions (e.g., depression, anxiety, schizophrenia), limited cooperation, and invalid responding (Heaton, Smith, Lehman, & Vogt, 1978; Larrabee, 1992; Reitan, 2001).

Although numerous factors may jeopardize the validity of neuropsychological test data, the discussion here focuses exclusively on *invalid responding*.

Detection of invalid responding

In the context of neuropsychological evaluation, invalid responding may manifest as reduced capabilities on cognitive performance measures, exaggeration of symptom complaints on self-report measures, or both (Larrabee, 2003; Heilbronner et al., 2009). Not surprisingly,

invalid responding can alter substantially the emerging picture of the brain-behaviour relationship, resulting in an underestimate of the examinee's true abilities and, ultimately, diagnostic error. Considering the potential consequences, it is imperative that neuropsychologists be able to detect invalid responding.

Clinical judgment. Although clinical judgment was once thought to be sufficient to detect invalid responding, several empirical studies have shown that experienced experts are inaccurate when asked to identify valid versus invalid performances using behavioural observations or scores on cognitive ability tests alone (e.g., Ekman et al., 1999; Faust, 1995; Faust et al., 1988). According to Millis and Putnam (1996), clinical judgment is ineffective for detecting invalid responding because (1) response distortion is not evident when examining ability test data alone, (2) clinicians may under- or over-diagnose noncredible performance as a result of confirmatory bias or attribution error, and (3) upon establishing rapport with examinees, clinicians tend to overestimate their capacity to identify individuals who are motivated to perform poorly.

As the limitations of clinical judgment have become apparent, researchers and clinicians have turned to psychometric measures to accomplish the task of identifying invalid responding in neuropsychological assessment. A large number of psychometric measures have shown promise in this regard.

Psychometric measures. For many years, all psychometric measures used for the purpose of detecting invalid responding, including those designed to detect reduced capabilities on cognitive performance measures and those designed to identify exaggeration of symptom complaints on self-report measures, were called symptom validity tests (SVTs). Recently, however, Larrabee (2012a) argued for a shift in terminology. He defined *symptom validity* as the

extent to which examinees' endorsements of symptoms on self-report measures reflect their true experience of those symptoms, and proposed that the term SVT be limited to psychometric indicators that detect invalid responding on self-report measures. The term *performance validity*, on the other hand, was defined as the extent to which examinees' performance on cognitive tests represents their true abilities; accordingly, it was suggested that psychometric measures used to detect invalid performance on cognitive tests should be referred to as performance validity tests (PVTs).

Research using factor analytic techniques has provided support for the SVT/PVT distinction proposed by Larrabee (2012a): indeed, PVTs and SVTs have been found to load on different statistical factors in multiple studies (Nelson, Sweet, Berry, & Bryant, 2007; Ruocco et al., 2008; Van Dyke, Millis, Axelrod, & Hanks, 2008). These results imply that performance validity and symptom validity should be evaluated independently, and that failure in one domain does not automatically invalidate the other domain (Van Dyke et al., 2008). Not surprisingly, modest correlations between performance validity and symptom validity have been reported in some studies (Haggerty, Frazier, Busch, & Naugle, 2007; Whiteside, Dunbar-Mayer, & Waters, 2009).

While it is recognized that both performance validity and symptom validity must be assessed in the course of a neuropsychological assessment, the discussion here is limited to the assessment of performance validity.

Performance validity tests. PVTs are psychometric measures or indices that are designed to detect invalid performance on cognitive tasks. Broadly, there are two types of PVTs at the disposal of the neuropsychologist: (1) free-standing PVTs, which are developed specifically to evaluate performance validity and therefore, in most cases, do not assess any other cognitive

function, and (2) embedded validity indicators, which represent scores derived from standard tests of cognitive ability that were later calibrated as PVTs.

Like cognitive ability tests, PVTs use continuous scales, where scores on one end of the scale indicate clearly valid performance and scores on the other end provide strong evidence of invalid performance. While scores on cognitive ability tests tend to be normally distributed, however, distributions of PVT scores are often skewed such that the majority of examinees perform well on these tests. Research has shown that potentially relevant nuances in performance validity can be identified by examining continuous scores on PVTs (Erdodi, 2017). However, the interpretation of PVT scores in practice is often based on dichotomous outcomes. A cutoff score is identified on the test, and scores representing better performance relative to the cutoff are interpreted as a *Pass*, while those representing worse performance are interpreted as a *Fail.*

In order to be effective, PVTs must assess the credibility of an examinee's performance while remaining unaffected by true cognitive impairment (Hartman, 2002) so that individuals with genuine cognitive dysfunction are not incorrectly identified as exhibiting invalid performance. Research has shown that PVTs are generally able to accomplish this task. Indeed, conditions like brain injury (e.g., Boone, Lu, & Herzberg, 2002a, 2002b; Green, 2003, 2004, 2008; Macciocchi, Seel, Alderson, & Godsall, 2006; Slick, Hopp, Strauss, Hunter, & Pinch, 1994; Tombaugh, 1997), learning disability (Boone, Lu, Herzberg, 2002a, 2002b; Hurtubise, Scavone, Sagar, & Erdodi, 2017), depression (Goldberg, Back-Madruga, & Boone, 2007), and aphasia (Tombaugh, 1997) are not typically associated with failure on PVTs. Failures have, however, been reported among more severely impaired populations (e.g., individuals with dementia or moderate-to-severe intellectual disability), suggesting that PVT failures provide weak evidence of invalid performance in such cases (Milanovich, Axelrod, & Millis, 1996).

Free-standing performance validity tests. Most free-standing PVTs employ a forced-choice recognition paradigm: the presentation of target stimuli is followed by a recognition task in which each target is paired with one or more foils (non-targets) and the examinee is instructed to identify the target. Most commonly, each target is paired with one foil, resulting in a two-alternative forced-choice task. The nature of stimuli varies between tasks and can include words (e.g., Green, 2003), digits (e.g., Binder & Willis, 1991; Binder, 1993a, 1993b), photographs (e.g., Warrington, 1984), and line drawings (e.g., Tombaugh, 1996).

Forced-choice paradigms, particularly those that involve two-alternative recognition tasks, were considered especially effective in the detection of invalid performance in the early days of PVT research because (1) chance levels of performance can be computed on such tasks, and (2) scores that are worse-than-chance can be argued to represent convincing evidence of noncredible performance (e.g., Millis, 1992). However, as research evidence accumulated, it became apparent that individuals with significant cognitive dysfunction (e.g., Warrington, 1984), those asked to feign cognitive deficits (e.g., Guilmette, Hart, & Giuliano, 1993; Martin, Bolter, Todd, Gouvier, & Niccolls, 1993), and those suspected of feigning cognitive deficits (e.g., Hiscock, Branham, & Hiscock, 1994) typically perform above chance levels on these tasks. In light of these findings, investigators have identified more liberal cutoffs, which are based on the lowest scores observed in bona-fide patients (Bianchini et al., 2001). Examples of such measures include the Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997), Portland Digit Recognition Test (PDRT; Binder, 1993a), Test of Memory Malingering, (TOMM; Tombaugh, 1996), and Word Memory Test (WMT; Green, 2003).

While the majority of free-standing PVTs employ a forced-choice paradigm, some free-standing PVTs do not. Non-forced-choice PVTs are designed in accordance with the floor effect

principle (Rogers, Harrell, & Liff, 1993) such that even individuals with true cognitive impairment are able to pass them. They can vary considerably in the types of stimuli used (e.g., digits, letter strings, words, pictures of unfamiliar faces, line drawings) and in the responses required of the examinee. For example, they may evaluate random responding, non-credibly poor performance, atypical errors, or inconsistencies in patterns of responding as compared to individuals with true brain dysfunction (Heilbronner et al., 2009).

The Rey 15-Item test (FIT; Rey, 1964 as cited in Nitch & Glassmire, 2007) has been identified as the most widely used non-forced-choice PVT (Sharland & Gfeller, 2007; Slick et al., 2004). In this task, a page with 15 stimuli is presented to examinees for 10 seconds. After 10 seconds, the page is removed and examinees are asked to draw as many of the 15 items as they can remember on a blank sheet of paper. With limited exposure time and 15 items to remember, the task appears to be difficult. However, due to the simplicity and redundancy of the stimuli, only five conceptual units need to be memorized to correctly reproduce all fifteen items, and even individuals with cognitive dysfunction can pass the task (e.g., Millis & Kler, 1995).

In another non-forced-choice PVT called the Dot Counting Test (DCT; Boone, Lu, & Herzberg, 2002b), examinees are presented with twelve 5”x7” cards with varying numbers of dots and instructed to count the dots as quickly as possible and then verbalize an answer. On the first six cards, the dots are ungrouped (i.e., in a random arrangement), while on the last six cards they are grouped (i.e., arranged in clear visual patterns to facilitate quick counting). Performance is evaluated by comparing the time taken to count (a) larger versus smaller numbers of dots, and (b) ungrouped versus grouped dots. Examinees are expected to take longer to count larger numbers of dots and to count ungrouped dots. There are several non-forced-choice PVTs in addition to the FIT and DCT (see Nitch & Glassmire, 2007, for a review).

Embedded validity indicators. In addition to administering free-standing PVTs, neuropsychologists can glean information about performance validity from indicators embedded within standard tests of neuropsychological abilities. These include tests of intellectual functioning, memory, attention and executive function, and motor and sensory functioning (for a review, see Boone, 2007).

Embedded validity indicators can include (1) traditional scores from ability tests (i.e., scores that are reported as part of the standard scoring procedures for a particular task), (2) indicators that use information from an ability test but are not reported as part of the standard scoring procedures (i.e., indicators developed specifically to evaluate performance validity), and (3) atypical patterns of performance. Patterns of test performance are evaluated for atypicality using scores that compare performance on one (often more difficult) task to another (easier) task. Based on knowledge about neuropsychological functioning, poorer performance is expected on the more difficult task relative to the easier task. Thus, performance is identified as atypical when an examinee's scores on the easier task are either worse than or comparable to their scores on the more difficult task. Finally, some researchers have developed discriminant functions and logistic regression equations that use multiple scores from one or more standard neuropsychological tests to determine performance validity (e.g., Millis & Putnam, 1994; Suhr & Boyer, 1999; Wolfe, Millis, Hanks, Fichtenberg, Larrabee, & Sweet, 2010).

There are several advantages associated with the use of embedded validity indicators in clinical practice (Strauss et al., 2006). First, they can provide information about performance validity independent of that obtained from free-standing PVTs. Second, considering that testing time is limited in most settings (e.g., due to time constraints, reimbursement issues or examinee stamina), embedded indicators do not require any additional testing time: information about

performance validity is obtained from standard administration procedures of existing neuropsychological tests. Third, they are cost-efficient: the tests serve double-duty by simultaneously assessing the purported cognitive functions and performance validity. Fourth, embedded validity checks allow a clinician to monitor an examinee's task engagement throughout the course of the evaluation rather than just at a single point in time. Fifth, given the availability of the neuropsychology literature to the general public and lawyers, embedded validity indicators may be harder to identify and therefore, more resistant to coaching than more common PVT formats, such as forced-choice procedures (Strauss et al., 2006). Finally, embedded validity indicators are useful for retroactively assessing the validity of past neuropsychological evaluations in which free-standing PVTs were not administered (Mittenberg, Aguila-Puentes, Patton, Canyock, & Heilbronner, 2002).

Use of PVTs in Clinical Practice

Performance validity is best conceptualized as falling on a continuum (Heilbronner et al., 2009). On one end of the continuum lie individuals who are fully engaged and demonstrating their true abilities during testing. The other extreme denotes individuals whose level of performance suggests markedly reduced cognitive capabilities but is thought to be non-credible (i.e., their scores likely underestimate their true ability level).

Various factors may underlie the demonstration of markedly reduced cognitive capabilities, and examinees may demonstrate this level of performance in the presence or absence of external incentives (e.g., receiving compensation for an injury, being released from work obligations). Examinees who demonstrate markedly reduced cognitive capabilities in the context of external incentives are described as displaying Malingered Neurocognitive Dysfunction (MND; Slick, Sherman, & Iverson, 1999) or *malingering*.

Because invalid performance may occur for a variety of reasons, of which malingering is only one, the main purpose of assessing performance validity is not to identify malingering but rather to rule out invalid performance as an explanation for neuropsychological test data in the larger context of differential diagnosis (Millis, 2008). Furthermore, inconsistent scores across tasks may be observed when test-taking motivation fluctuates during the course of a neuropsychological evaluation. In these instances, assessment of performance validity, which includes multiple measures administered throughout the evaluation, can help to make sense of seemingly discordant data.

Although the goal of performance validity assessment has expanded over the years from identifying malingering in forensic or medical-legal settings to determining the veracity of neuropsychological test data in clinical contexts, a more elaborate discussion of the construct of malingering and some of the associated issues is warranted from both clinical and research standpoints. From a clinical perspective, it is important to review the definition of malingering; the criteria used to identify it; and the implications of such a label. On the other hand, an understanding of the construct of malingering is necessary to appreciate the research literature in the area of performance validity: when developing and validating new PVTs, researchers frequently assign participants to valid and invalid groups using the proposed criteria for MND or some variant thereof. In light of these arguments, the following discussion covers (1) the definition and diagnostic criteria for malingering, (2) the controversy around neuropsychologists' ability to infer the intent of an examinee and how the issue may be addressed in practice, and (3) the estimated prevalence or base rates of malingering and/or invalid performance in various assessment contexts.

Malingering. Slick, Sherman, and Iverson (1999) coined the term *Malingered Neurocognitive Dysfunction (MND)* to refer to "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding formal duty or responsibility" (pp. 552). They also specified criteria for the identification of MND in clinical, medical-legal, or forensic practice, which are based on (A) information about the presence of substantial external incentive, (B) evidence from neuropsychological testing, and (C) evidence from self-report. Notably, the criteria do not limit the diagnosis of malingering to individuals without psychiatric, neurological, or developmental conditions: the only instance when MND *cannot* be diagnosed is when the behavioural or psychometric evidence can be *fully* accounted for by existing psychiatric, neurological, or developmental factors (Criterion D). The Slick et al. (1999) MND criteria are summarized in Table 1.

Slick et al.'s (1999) criteria have three modifiers reflecting various levels of certainty: definite MND, probable MND, and possible MND. A classification of *definite MND* requires external incentive to malinger and definite negative response bias as indicated by below-chance performance on one or more forced-choice PVTs. *Probable MND* is concluded when, in addition to external incentive to malinger, there is either (a) failure on two or more PVTs, excluding below-chance performance on forced-choice tests (i.e., two of Criteria B2-B6), or (b) one non-below-chance PVT failure and one instance of non-credible self-reported symptoms or self-report discrepancy (i.e., one of Criteria B2-B6 and one of Criteria C1-C5). Finally, *possible MND* may be indicated (a) in the presence of external incentive and discrepant evidence from self-report (one or more of Criteria C1-C5), or (b) when criteria for definite or probable MND are met but the behaviours may be accounted for by psychiatric, neurological, or developmental factors (i.e., Criterion D not met).

Table 1

Summary of Criteria for Malingered Neurocognitive Dysfunction (MND) as Proposed by Slick, Sherman, & Iverson (1999)

A. Presence of a substantial external incentive
    1. At least one clearly identifiable and substantial external incentive for exaggeration or fabrication of symptoms is present at the time of examination (e.g., personal injury settlement, disability pension, evasion of criminal prosecution, or release from military service).
B. Evidence from neuropsychological tests
    1. *Definite response bias.* Below chance performance (p <.05) on one or more forced-choice PVTs.
    2. *Probable response bias*. Performance on one or more well-validated PVTs is consistent with feigning.
    3. *Discrepancy between test data and known patterns of brain functioning*.
    4. *Discrepancy between test data and observed behavior*.
    5. *Discrepancy between test data and reliable collateral reports*.
    6. *Discrepancy between test data and documented background history*.
C. Evidence from self-report
    1. *Self-reported history is discrepant with documented history.*
    2. *Self-reported symptoms are discrepant with known patterns of brain functioning.*
    3. *Self-reported symptoms are discrepant with behavioral observations.*
    4. *Self-reported symptoms are discrepant with information obtained from collateral informants.*
    5. *Evidence of exaggerated or fabricated psychological dysfunction* (based on well-validated validity scales or indices).
D. Behaviors meeting necessary criteria from groups B or C are not fully accounted for by psychiatric, neurological, or developmental factors.

Slick et al.'s (1999) requirement of below-chance performance on one or more forced-choice PVTs for the diagnosis of definite MND has been criticized (Boone, 2007; Larrabee, Greiffenstein, Greve, & Bianchini, 2007). Among the strongest criticisms of this criterion is the idea that it limits the diagnosis of definite MND to "unsophisticated" (i.e., blatant) malingerers. Indeed, more sophisticated malingerers would likely not exhibit such extremely poor performance and, therefore, would not be identified as displaying definite MND. In light of this, Boone (2007) argued that failure on at least three validated PVTs with minimal shared variance and behavioural evidence of noncredible symptoms should be sufficient for a diagnosis of

definite MND. Similarly, Larrabee (2008, 2012b) has shown that failure on three independent PVTs is diagnostically equivalent to below-chance performance on a forced-choice PVT, when it occurs in a context with external incentive to malinger and cannot be explained by developmental, psychiatric, or neurological factors.

The issue of intent. Given Slick and colleagues' (1999) definition of MND (i.e., "the *volitional* exaggeration or fabrication of cognitive dysfunction"), the diagnosis of MND requires a clinician to make an inference about the intent of the examinee. Not surprisingly, there has been considerable controversy around the idea of clinicians' ability to infer an examinee's intent. Some (e.g., Boone, 2007; Pankratz & Erickson, 1990) have argued that, despite the development and implementation of PVTs and SVTs, neuropsychologists remain unable to establish the degree of conscious intent to appear impaired. Therefore, MND cannot be diagnosed with absolute certainty. Others (e.g., Heilbronner et al., 2009; Larrabee et al., 2007), however, have stated that malingering *can* be diagnosed in some examinees. According to Larrabee and colleagues (2007), intent is established by the presence of multiple improbable findings that occur in the context of external incentive and without any reasonable alternative explanation.

Although there is disagreement about whether or not neuropsychologists are *able* to make inferences about intent, the determination of intent is often unnecessary (Boone, 2007; Tombaugh, 2002). Indeed, the role of the neuropsychologist is to determine whether an examinee exhibits credible cognitive impairment and, used in concert, ability tests and PVTs allow the neuropsychologist to fulfill this role quite well. In cases where results indicate invalid performance, it has been recommended that the *behaviour* be described using terms such as "noncredible", "implausible", and/or "inconsistent with injury", without any inference(s) about the *reason* for the lack of credibility (Boone, 2007; Heilbronner et al., 2009).

Prevalence of invalid performance by setting. Rates of noncredible performance are highest in forensic or medical-legal contexts where there is potential for secondary gain. Considering data from eleven different studies conducted in settings with potential for secondary gain, Larrabee (2003) found an overall BR_{Fail} of 40%. Estimates ranged from 15% (Trueblood & Schmidt, 1993) to 64% (Heaton et al., 1978), which were interpreted by Larrabee (2003) as under- and over-estimates, respectively.

In clinical settings where the purpose of neuropsychological assessment is typically to assist with diagnosis or treatment, there is less apparent incentive on the part of the examinee to appear impaired. Despite this, however, such settings are not immune to invalid performance. In fact, a figure of 15% has been suggested as a *lower* bound for BR_{Fail} in general clinical settings (Boone et al., 2002b as cited in Strauss et al., 2006). In settings where a mixture of clinical and forensic referrals is seen, BR_{Fail} is approximately 30% (Mittenberg et al., 2002).

These estimates indicate that although there are relative differences in the prevalence of invalid performance across assessment contexts, the issue of invalid performance is ubiquitous. Thus, it is important to consider invalid performance in the larger context of differential diagnosis in all settings, as suggested by Millis (2008), and assess performance validity using the appropriate tools.

The PVT Research Base and Its Implications for Practice

In order to fully appreciate the assessment of performance validity, it is necessary to understand the research base underlying the use of PVTs. The following sections provide a review of (1) the research designs used to validate PVTs, (2) the computational aspects of classification accuracy, and (3) methodological limitations of the PVT research base. Implications for practice are also discussed where relevant.

Research designs. The three major research designs used in the validation of PVTs include (1) analogue-simulation designs, (2) criterion groups or "known-groups" designs, and (3) differential prevalence designs (Heilbronner et al., 2009; Larrabee, 2012c).

Analogue-simulation designs. Analogue-simulation designs represent a practical and cost-effective method for investigating new PVTs. Research studies falling in this category typically involve non-clinical participants, such as college students or community volunteers, who are randomly assigned to different experimental conditions. One group may be asked to malinger (i.e., simulate cognitive impairment consistent with a brain injury on one or more tests, for example), while another group is asked to perform to the best of their ability. Test scores from these two groups are then compared to each other and sometimes also to the scores of a clinical group of interest (e.g., individuals with moderate to severe brain injury).

Although they have the advantage of experimental control, analogue-simulation studies have been criticized for their low generalizability (Babikian & Boone, 2007; Bianchini, Mathias, & Greve, 2001; Rogers, 2008). They tend to overestimate classification accuracy (Babikian & Boone, 2007; Sollman & Berry, 2011), and cutoffs that are found to perform optimally in such designs often do not generalize to real-world samples (e.g., DiCarlo, Gfeller, & Oliveri, 2000; Greve, Bianchini, & Roberson, 2007; Tenhula & Sweet, 1996).

Other issues that have been raised with analogue-simulation designs include differences between simulators and real-world malingerers in demographic variables (e.g., age, education level), presence or absence of brain insult, and the smaller monetary incentives to feign cognitive impairment in analogue-simulation studies as compared to real-world settings (Babikian & Boone, 2007; Bianchini et al., 2001; Demakis, 2004). Furthermore, instructions for malingering groups vary between studies. Even within studies, researchers often fail to provide clear

descriptions of the instructions that were given to different experimental groups, particularly the simulating malingerers. Finally, in many studies, researchers do not assess the degree to which simulators comply with instructions. According to Bianchini and colleagues (2001), in studies that evaluated simulator compliance (Frederick, Sarfaty, Johnston, & Powel, 1994; Goebel, 1983), between 6 and 10% of simulators reported making little or no attempt to perform poorly, which is likely an underestimate. These criticisms indicate the need for greater caution in both executing and reporting findings from studies based on analogue-simulation designs.

To increase generalizability, Rogers (1997) has suggested using analogue-simulation designs that include four groups: (a) simulating nonclinical subjects, (b) honestly responding nonclinical subjects, (c) honestly responding clinical subjects, and (d) clinical subjects simulating greater impairment than they actually experience. Such designs are more useful because they address the more clinically relevant question of whether a pattern of impaired performance is due to brain injury or to invalid performance, rather than whether noncredible performance can be differentiated from the performance of normal controls (Millis, 2008).

Criterion groups (known groups) designs. In a criterion-groups design, two groups are established: bona fide patients and noncredible patients (Larrabee, 2012b). Individuals are assigned to the noncredible group based on *a priori* criteria, which may include failure on a well-validated PVT or the Slick et al. (1999) criteria for definite or probable MND. Once established, the two groups are examined by way of a systematic analysis of similarities and differences. When validating a new PVT, this analysis involves comparing performance of the two groups on the test of interest. Conceptually, the selection of the bona fide patient group aids in establishing an "empirical floor" on the test of interest: a norm-referenced, clinically-derived cutoff below which individuals with true impairment rarely perform.

The strength of criterion-groups designs lies in their use of real-world patients with real incentives to perform noncredibly, which ultimately results in increased generalizability to clinical and forensic settings. Furthermore, criterion-groups designs directly assess the critical diagnostic question of whether a pattern of impaired performance is due to brain injury or noncredible performance (Bianchini et al., 2001).

Criterion-groups designs are not without limitations. For example, it may be difficult, if not impossible, to ensure that the group of bona-fide patients does not include any individuals who performed non-credibly and that the noncredible group does not include any bona-fide patients. The purity of the two groups may further be compromised when individuals in the clinical bona-fide patients group are in litigation (Larrabee, 2012b).

According to Heilbronner and colleagues (2009), both analogue-simulation designs and criterion-groups designs represent rigorous and clinically relevant research designs that are useful in the validation of PVTs.

Differential prevalence designs. Differential prevalence designs are based on the assumption that the prevalence of invalid performance varies as a function of the assessment context and the likelihood of incentives to malinger (Larrabee, 2012b). As such, groups are comprised on the basis of *assumed* incentives: individuals who are known to be in litigation (and therefore assumed to have a higher rate of invalid responding) are compared, for example, to those who are not in litigation (and thus are thought to have a lower rate of invalid responding).

Differential prevalence designs are problematic because it is difficult, if not impossible, to know the accuracy of the assumptions that are inherent to the design (Heilbronner et al., 2009). In other words, while rates of invalid performance tend to be higher among compensation-seeking individuals as compared to those not seeking compensation, this may not

be the case in a given sample. In fact, it is highly unlikely that all individuals in a litigating group would demonstrate invalid responding and all of those in a non-litigating group would perform optimally. Thus, one of the major criticisms of differential prevalence designs has been the difficulty in determining which participants in each group, and how many of them, are exhibiting invalid performance (Rogers, 1997).

Differential prevalence designs are considered less rigorous as compared to analogue-simulation and criterion-groups designs. As a result, although such designs can be used to supplement evidence of test validation from other designs, they should not be used as the sole or primary research design to validate a PVT (Heilbronner et al., 2009).

Classification accuracy statistics. In the research designs described above, participants are assigned to either the *valid* group or the *invalid* group based on a criterion of the researcher's choice, which, in many cases, constitutes performance on a given PVT (criterion PVT). When validating a new PVT (PVT of interest), data analysis begins with null hypothesis significance testing, which yields information about whether or not mean scores on the PVT of interest differ significantly between the valid and invalid groups. While this represents an important first step, such analyses may not reflect clinically significant effects and do not provide information that is useful to clinicians (Woods, Weinborn, & Lovejoy, 2003; Larrabee, 2012c). Therefore, null hypothesis significance testing is usually followed by analyses of classification accuracy, which, unlike null hypothesis significance testing, *do* provide diagnostically meaningful information to the clinician (Baldessarini, Finklestein, & Arana, 1983; Glaros & Kline, 1988; Meehl & Rosen, 1955).

Given that no PVT has perfect classification accuracy, the criterion PVT is likely to be imperfect. However, classification accuracy analyses rest on the working assumption that the

researcher's criterion represents the "truth": those cases identified as valid by the criterion PVT are considered to be valid for the purpose of the given analysis, and those identified as invalid are assumed to be invalid. Participants are also designated as *valid* or *invalid* based on scores on the PVT of interest (independently of the criterion PVT). Then, the classification accuracy of PVT of interest, relative to the criterion PVT, is evaluated by computing the concordance rate. Results are reported in the form of traditional classification accuracy statistics, which may include area under the curve, sensitivity, specificity, hit rates, predictive values, likelihood ratios, and odds ratios (Baldessarini et al., 1983; Glaros & Kline, 1988; Ivnik et al., 2001; Woods et al., 2003). Of these, sensitivity and specificity are most relevant for establishing and evaluating cut-off scores on PVTs.

When evaluating the concordance rate between the predictor and criterion PVTs, there are four possible outcomes. Of all cases that were originally identified as invalid by the criterion PVT, some are correctly classified as invalid by the PVT of interest (*true positives*), while others are incorrectly classified as valid (*false negatives*). Similarly, of all cases originally identified as valid by the criterion PVT, a subset is correctly classified as valid by the PVT of interest (*true negatives*), while the rest are incorrectly classified as invalid (*false positives*). Sensitivity refers to the proportion of invalid participants who are correctly identified as invalid (true positives/[sum of true positives and false negatives]), while specificity is the proportion of valid participants who are correctly identified as valid (true negatives/[sum of true negatives and false positives]). The hit rate represents the total proportion of accurately classified cases ([sum of true negatives and true positives]/total number of cases; e.g., Larrabee, 2012c). Table 2 illustrates the concepts of true positives, true negatives, false negatives, and false positives using a 2x2 matrix.

Table 2

Summary of Major Classification Accuracy Parameters

	Criterion PVT	
PVT of interest	Valid (Pass)	Invalid (Fail)
Valid (Pass)	True negatives	False negatives
Invalid (Fail)	False positives	True positives

Implications for practice. Cut-off scores on PVTs, which are used to discriminate valid from invalid performance, are established after careful consideration of the relative cost of false-positive and false-negative errors. There is a consensus in the literature that false positives (i.e., incorrectly identifying an examinee's performance as invalid) carry more significant consequences compared to false negatives (i.e., incorrectly classifying an examinee as valid). In a forensic or medical-legal context, a false positive could result in substantial financial, social, and/or personal ramifications (Greve & Bianchini, 2004). In a clinical context, suspected noncredible performance may impact the clinician's formulation, the diagnosis, and/or the development of a care plan (Seusse, Wong, Stamper, Carpenter, & Scott, 2015). Considering these potential consequences, specificity is usually set at ≥.90 when selecting cutoffs on PVTs, which keeps the false positive rate at ≤10% (Boone, 2007; Larrabee, 2012b; Vickery, Berry, Inman, Harris, & Orey, 2001), although specificity values as low as ≥.84 have been suggested (Larrabee, 2003).

Prioritizing specificity over sensitivity means that fewer instances of invalid performance are detected. Fixing specificity at .90 typically results in sensitivity around .50 – in other words, approximately half of invalid subjects are correctly identified (Vickery et al., 2001). This seemingly inescapable trade-off has been labeled the "Larrabee limit" (Erdodi, Kirsch et al., 2014). With specificity of .90 and sensitivity of .50, PVTs are able to identify the *presence* of invalid performance more effectively than the *absence* of invalid performance. In other words,

failure on a PVT provides strong evidence for invalid performance, while a passing score does not rule out invalid performance (Heilbronner et al., 2009; Seusse et al., 2015).

With regard to their *relative* diagnostic power, PVTs vary considerably in sensitivity and specificity. However, when comparing individual PVTs, free-standing PVTs are not only more sensitive to invalid performance than embedded PVTs, but they also demonstrate higher overall classification accuracy (Iverson & Binder, 2000; Miele, Gunner, Lynch, & McCaffrey, 2012; Sweet & Nelson, 2007).

While there are differences in the classification accuracies of free-standing and embedded PVTs, no single test has perfect sensitivity and specificity. As a result, it has been recommended that clinicians employ multiple PVTs in practice, which are interspersed throughout the test battery and provide relatively independent information regarding performance validity (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2012c; Orey, Cragar, & Berry, 2000; Sweet & Nelson, 2007; Vickery et al., 2004). Support for this recommendation comes from research showing that classification accuracy, and sensitivity to invalid performance in particular, is improved when multiple measures are used (Larrabee, 2003; Nelson, Boone, Dueck, Wagener, Lu, & Grills, 2003; Vickery et al., 2004; Victor, Boone, Serpa, Buchler, & Ziegler, 2009). In the literature, the best overall hit rates have been achieved when invalid performance is defined as failure of ≥2 PVTs, and failing ≥3 or ≥4 PVTs has been associated with zero false-positive errors (Larrabee, 2003; 2014; Vickery et al., 2004; Victor et al., 2009). The improved specificity associated with the use of multiple PVTs is a result of the very low multivariate base rate of failure even among significantly impaired individuals: while a credible examinee may, on rare occasion, obtain one score in the failing range, two or more scores in this range are highly unlikely (Larrabee, 2012b).

Limitations of the PVT research base. The last few decades have seen a rapid expansion of the PVT literature. However, several methodological limitations persist. First, although the methodology described above is a valuable tool in PVT research, the manner in which cutoffs are developed has been criticized for imposing artificial *Pass*/*Fail* dichotomies (Dwyer, 1996). In addition, given the use of inherently imperfect instruments, some amount of error is inevitable (e.g., Bigler, 2014). Furthermore, and perhaps as a result of using an imperfect procedure, optimal cutoffs for many PVTs vary across studies, producing ambiguity around the most appropriate cutoff.

The variability in cutoffs is further complicated by the fact that most PVT studies use small samples consisting of one or few diagnostic groups. In fact, the majority of the research in the area of performance validity has been conducted with mild traumatic brain injury (mTBI) samples, largely to the neglect of other syndromes that may be encountered in clinical and forensic settings (Bianchini et al., 2001). Although the practice of applying cutoffs to samples other than those from which they were derived is not uncommon, the degree to which cutoff scores generalize to different clinical populations remains unknown (Millis, 2008). Considering this, it is important to empirically determine the appropriateness of PVT cutoff scores for new clinical populations.

The need for better characterization of PVT profiles among specific diagnostic groups is also supported by data suggesting that BR_{Fail} may differ depending on the patient population. In a survey of neuropsychologists belonging to the American Academy of Clinical Neuropsychology (AACN), Mittenberg and colleagues (2002) obtained estimates based on a total of 33, 531 cases that had been referred for personal injury (n = 6371), disability (n = 3688), criminal (n = 1341) or medical (n = 22,131) matters. Results revealed a wide range in estimated

BR_{Fail} among different patient populations: mild head injury (39%), fibromyalgia or chronic fatigue (35%), pain or somatoform disorder (31%), neurotoxic disorders (27%), electrical injury (22%), depressive disorders (15%), anxiety (14%), dissociative disorders (11%), seizure disorders (9%), moderate or severe head injury (9%), and vascular dementia (2%). Considering the clinical relevance of such data, further empirical exploration is warranted in this area. Ideally, population-specific cutoffs would be developed for each PVT so that the same measures could be used to evaluate performance validity with confidence across a wide range of diagnostic groups (Pearson, 2009).

In addition to diagnosis, other demographic and/or background variables, such as age, education, and English language status, have been found to impact performance on some PVTs as well. For example, reading level below the 3rd grade was associated with elevated BR_{Fail} on the Word Memory Test (Green & Flaro, 2003). Additionally, examinees who spoke English as a second language demonstrated lower scores on the WAIS Reliable Digit Span as compared to native English speakers (Erdodi, Nussbaum, Sagar, Abeare, & Schwartz, 2017; Salazar, Lu, Wen, & Boone, 2007). Given the clinical implications of such findings, this area also requires further exploration.

Another limitation of the PVT research base is the fact that despite recommendations for the use of multiple PVTs in practice (Boone, 2007; Larrabee, 2012), there is little by way of research evidence to help clinicians determine which and how many PVTs to use, in what order, and in what context (Bigler, 2012). This may be due, at least in part, to the fact that many studies investigate only a few PVTs at a time. More studies in which large numbers of PVTs are evaluated simultaneously would allow clinicians to directly compare PVTs to each other, to select individual PVTs that are most appropriate for specific diagnostic categories, and to explore

the combinations of PVTs that are optimal for detecting invalid performance among different patient groups.

Finally, most studies do not report information in a format that is useful to practitioners. Sensitivity and specificity are the most commonly provided signal detection parameters. Although these statistics are relevant for the development and validation of cutoff scores, they are not helpful when making decisions about performance validity on a case-by-case basis, particularly when an examinee with a known neurological condition fails a PVT with a score that is close to the specified cutoff or within the range of multiple specified cutoffs (i.e., "near-Pass", Bigler, 2014 or "soft Fail", Erdodi & Lichtenstein, 2017). In such cases, a clinician must interpret the score either (1) as representing invalid performance, or (2) as reflecting underlying neuropathology (i.e., that the score belongs to an individual who performed in the lower end of a clinical group demonstrating valid performance). However, sensitivity and specificity offer little information to help a clinician make this determination.

Other classification accuracy statistics, such as positive predictive power (PPP) and negative predictive power (NPP), are more useful when interpreting individual PVT scores obtained by examinees in clinical or forensic settings. PPP is defined as the ratio of true positive scores to total positive scores (Baldessarini et al., 1983) and represents the probability that the profile is invalid given a failure on a PVT (Larrabee, 2012b). NPP, on the other hand, is defined as the ratio of true negatives to total negative scores (Baldessarini et al., 1983) and reflects the probability of valid performance given a passing score on a PVT (Larrabee, 2012b). When sensitivity and specificity are held constant, PPP and NPP depend on BR_{Fail} such that a passing score on a PVT is more likely to be true than a failing score in settings with a low BR_{Fail}, while a failing score on a PVT is more likely to be true than a passing score in settings with high BR_{Fail}

(Baldessarini et al., 1983). Considering this, PPP and NPP represent highly useful indices for clinical practitioners. Unfortunately, however, these statistics remain underused in biomedical and neuropsychological research (Woods et al., 2003) and in clinical practice.

The underuse of PPP and NPP may be due to various factors. First, PPP and NPP require a nuanced understanding of classification accuracy analyses in order to be used accurately and effectively. Furthermore, the interpretation of predictive values is quite abstract and may not be accessible to most clinical practitioners. Finally, such estimates rely on precise knowledge of the setting-specific BR_{Fail}, which many clinicians may not have.

As an alternative to the complex statistical modeling procedures that are prevalent in PVT research, the authors of the *Advanced Clinical Solutions (ACS)* technical manual for the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) and Wechsler Memory Scale—Fourth Edition (WMS-IV) presented a descriptive model for one free-standing PVT and four embedded PVTs, which enables quick and easy interpretation of individual PVT scores in clinical practice (Pearson, 2009). Specifically, scores on each PVT were reported at various BR_{Fail} (i.e., ≤75%, ≤50%, ≤25%, ≤15%, ≤10%, ≤5%, and ≤2%) in an overall mixed clinical sample (N = 371), 10 diagnostic groups (i.e., TBI, temporal lobectomy, schizophrenia, major depressive disorder, mild intellectual disability, autistic disorder, Asperger's disorder, reading disorder, mathematics disorder, and ADHD), and a group of simulators. Given that PVT failures typically represent unusually poor scores, presentation of data in this manner establishes magnitude of failure, where scores observed in ≤2% of the sample, for example, represent more severe failures than scores observed in ≤15% of the sample. In practice, such a methodology helps clinicians to determine the degree to which an examinee's PVT score is unusually poor compared to a general clinical sample and to the specific diagnostic group to which the examinee

belongs. Furthermore, it allows an examinee's score to be compared directly with scores of a group asked to simulate cognitive impairment (i.e., simulator group).

While the clinical utility of the PVT research base can be increased by reporting predictive values (PPP and NPP) or by using descriptive methodologies similar to that used in the ACS manual (Pearson, 2009), the ACS methodology is likely more attractive to the clinical practitioner because it is easily comprehensible and intuitive, yet methodologically sophisticated. It is also less prone to errors in interpretation than are predictive values. Given the simplicity and concreteness of this method, presentation of PVT data in such a manner has the potential to provide a user-friendly, evidence-based template for the clinical interpretation of PVT scores in various assessment contexts.

The Present Dissertation

The present dissertation consisted of three studies and aimed to characterize performance on three free-standing PVTs and 11 embedded validity indicators (collectively referred to as *PVTs of interest*) in a large, real-world medical-legal sample. Performance on each PVT of interest was characterized using descriptive models in Studies 1 and 2 and traditional classification accuracy statistics in Study 3.

Study 1 examined whether BR_{Fail} on each PVT of interest varied as a function of five key patient variables in the overall sample ($N = 4721$). Specifically, relative cumulative frequency distributions were reported for each PVT of interest, showing the BR_{Fail} at various cutoffs for individuals in the overall sample; nine diagnostic groups (mTBI, moderate-severe TBI, neurological, depression, anxiety, severe mental illness, chronic pain/fibromyalgia, orthopedic, and other); five age groups; six education levels; gender (male or female); and as a function of English language status (i.e., native versus non-native speakers of English). Finally, as a first

step to investigating the relationships between the 14 PVTs of interest and the Word Memory Test (WMT; Green, 2003), which was used as a criterion measure in subsequent studies, BR_{Fail} were also reported on each PVT of interest for examinees who passed versus failed the WMT.

While Study 1 was inspired by the methodology used in the ACS manual (Pearson, 2009), it represents a more comprehensive effort than that of Pearson (2009) as it includes more PVTs, a much larger sample of patients (who, in the majority of cases, had real incentives to appear impaired), and several diagnostic groups that are commonly seen in clinical and forensic practice. Furthermore, in addition to clinical diagnosis, Study 1 characterizes PVT BR_{Fail} as a function of age, education level, gender, and English language status.

Study 2 replicated the methodology used in Study 1, while addressing an additional major limitation of the ACS model. Although the authors of the ACS manual reported data separately for clinical patients and a group of simulators, performance validity was not assessed independently of the PVTs on which they reported data. Given estimates indicating a 15% BR_{Fail} in clinical populations without external incentives (Boone et al., 2002b as cited in Strauss et al., 2006), it is likely that the ACS clinical sample included numerous cases demonstrating invalid performance, none of which were identified or reported as having been excluded from the analyses. Although this practice is epistemiologically problematic, it is virtually ubiquitous in normative samples, and has only been recently identified as a potential confound (Erdodi & Lichtenstein, 2017). To address this limitation, the WMT was used as a criterion measure in Study 2, such that only cases passing the WMT at standard cutoffs (Green, 2003) were included in the analyses (valid sample; $N = 3297$). The nature of data reported was similar to that reported in Study 1 (i.e., relative cumulative frequency distributions showing PVT performance across patient variables).

Finally, as an alternative to the descriptive information provided in Studies 1 and 2, Study 3 used traditional classification accuracy analyses to determine the sensitivity and specificity of each of the 14 PVTs of interest against the WMT. In order to enhance the clinical utility of the data presented, PPP and NPP were also reported for each cutoff at five hypothetical BR_{Fail}: 10%, 20%, 30%, 40%, and 50%. Finally, the classification accuracies of the PVTs of interest were compared to each other in order to determine their relative sensitivities to invalid performance.

CHAPTER 2

Method

Participants

For the purpose of this dissertation, two large, independent databases were merged, which contained psychological and neuropsychological test data from adults assessed at private psychological practices in a mid-western Canadian city. Although various studies have been conducted and published with each independent database (by Paul Green and Roger O. Gervais), the databases have not previously been combined as they were for the present dissertation.

The total sample consisted of 4721 independent cases of adults who were referred for psychological or neuropsychological testing. The majority of examinees were evaluated in the context of seeking compensation claims (i.e., they had an external incentive to appear impaired). At the time of assessment, consent was obtained from all examinees for their demographics, diagnostic information, and test data to be used for future archival research. Information unique to each participant was coded and all identifying information was removed before the data were released to the researcher for this dissertation.

The Word Memory Test (WMT; Green, 2003) was used as an independent criterion to classify participants as *valid* or *invalid*: those who passed the WMT at standard cutoffs were identified as valid, while those who failed the WMT at standard cutoffs were identified as invalid. Approximately one third of the overall sample (30.1%) was classified as invalid based on WMT performance. Table 3 shows the mean age, education, and Full Scale IQ for the overall sample, the valid group, and the invalid group. Results of independent samples *t*-tests indicated that participants in the valid group were slightly younger and more educated than those in the invalid group, $t(4719) = -8.23$, $p < .001$ and $t(4715) = 7.94$, $p < .001$ (small effects). Mean Full

Scale IQ was also lower among the invalid group than the valid group, $t(3646) = 16.97$, $p < .001$ (medium effect).

Table 3

Means and Standard Deviations of Age, Years of Education, and Full Scale IQ, for Participants in the Overall Sample and the Valid and Invalid Groups

	Overall Sample[a]		Valid Group[b]		Invalid Group[c]				
	M	*SD*	*M*	*SD*	*M*	*SD*	*df*	*t*	*d*
Age	42.3	11.2	41.4	11.3	44.3	10.8	4719	-8.23*	-0.25
Education	12.0	2.61	12.2	2.54	11.6	2.70	4715	7.94*	0.23
FSIQ	99.7	14.2	102.1	13.5	93.5	14.0	3646	16.97*	0.63

Note. Valid Group = Participants who passed the Word Memory Test (WMT; Green, 2003); Invalid Group = Participants who failed the WMT; FSIQ = Full Scale IQ.
[a]$n = 4721$. [b]$n = 3297$. [c]$n = 1424$.
*$p < .05$.

Table 4 summarizes additional characteristics of participants in the overall sample, the valid group, and the invalid group, including gender, handedness, primary diagnosis, first language, and oral English fluency. Information about gender, handedness, and first language was collected during the clinical interview. Examinees who identified themselves as non-native English speakers were classified as *English as a second language* (ESL). Oral English fluency was coded subjectively by the licensed clinical psychologist who conducted each evaluation based on examinees' abilities to express themselves and comprehend the examiner while conversing in English. Data from ESL examinees was only included in the present dissertation if the assessments were completed in English and without an interpreter. All ESL cases tested with an interpreter were excluded from the analyses.

In the database, each participant's primary diagnosis was identified by assigning one of several diagnostic classes to the case. The classes included mTBI, moderate-severe TBI, neurological conditions, depression, anxiety, bipolar disorder/psychosis, chronic

pain/fibromyalgia, orthopedic, and *other* diagnoses. Notably, participants with diagnoses of intellectual disability and probable dementia were excluded from this dissertation. In most cases, the diagnostic classes assigned to participants represented the diagnoses under which they were referred for the psychological or neuropsychological assessments. There were, however, some cases where patients were not formally diagnosed prior to the assessment or where the diagnosis rendered at the conclusion of the assessment differed from the diagnosis of referral. In such cases, the diagnostic class coded in the database reflected the diagnosis rendered by the psychologist after the assessment.

When assigning diagnostic classes for research purposes, TBI severity was determined primarily using post-traumatic amnesia (PTA): patients with PTA of 24 hours or longer were identified as having sustained a moderate-severe TBI, while those with PTA of less than 24 hours were identified as having sustained a mild TBI. Information about length of PTA was obtained from medical records where they were available. Patient reports of PTA duration were accepted in cases where PTA was undocumented but other data (e.g., GCS, neuroimaging, other medical records) provided clear evidence of a severe TBI. Predictably, the mTBI group had a mean PTA duration of 1.19 hours (SD = 4.60), and the moderate-severe TBI group had a mean PTA duration of approximately 17 days (M = 416.25 hours; SD = 543.8). The mild and moderate-severe TBI groups had mean GCS scores of 14.6 (SD = 1.1) and 8.7 (SD = 4.0), respectively. Approximately half of mTBI patients (52.4%) did not have any intracranial abnormalities on neuroimaging, while abnormalities were documented in 15.2% of mTBI cases. Among moderate-severe TBI patients, there was evidence of intracranial abnormalities in 77.3% of cases, and no evidence of any abnormalities in 7.7% cases. For the remainder of TBI cases (32.3% mTBI patients, 15.0% moderate-severe TBI), results of neuroimaging were not available

for review. Although information about time since injury (i.e., how much time had elapsed between a patient's TBI and the neuropsychological assessment) was not available in the database, according to the source of the database, TBI cases were invariably assessed more than three months post-injury, and often one or more years post-injury (P. Green, personal communication, May 3, 2018).

The diagnostic class labeled *other* contained a wide range of individuals not captured by the other eight classes, including those assessed in the context of general medical conditions (e.g., leukemia, AIDS), substance abuse, and work-related accidents. Notably, the diagnostic classes identified in the database are not mutually exclusive. Although individuals assessed in clinical practice often have multiple diagnoses, only information about the primary diagnostic class was coded for each case. Information about comorbid conditions was not available.

Results of chi-square tests of independence indicated that English language status and primary diagnosis were related to being classified as valid versus invalid based on WMT performance, $\chi^2\ (1, N = 4501) = 108.97, p < .001$, and $\chi^2\ (8, N = 4721) = 97.66, p < .001$; these variables explained 2.4% and 2.1% of the variance in WMT results (*Pass* or *Fail*) in the overall sample (small effect). Effect sizes for analyses involving gender, oral English fluency, and handedness were close to zero, suggesting that these variables were not meaningfully related to WMT classification. The statistically significant chi-square tests of independence for gender, $\chi^2\ (1, N = 4721) = 23.70, p < .001$, and oral English fluency, $\chi^2\ (1, N = 4716) = 9.71, p = .002$, were likely the result of overpowered analyses given the large sample size. The chi-square test involving handedness did not reach statistical significance, $\chi^2\ (1, N = 4662) = 0.02, p = .89$.

Table 4

Characteristics of Participants in the Overall Sample and in the Valid and Invalid Groups

	Overall Sample[a]		Valid Group[b]		Invalid Group[c]			
	N	%	*n*	%	*n*	%	χ^2	Φ^2
Gender								
Male	2817	59.7	1892	67.2	925	32.8	23.70*	.005
Female	1904	40.3	1405	73.8	499	26.2		
Handedness								
Right	4198	90.0	2934	69.9	1260	30.0	0.02	.000
Left	468	10.0	326	69.7	142	30.3		
Primary diagnosis								
Mild TBI	643	13.6	386	60.0	257	40.0	97.66*	.021
Moderate-severe TBI	220	4.7	173	78.6	47	21.4		
Neurological	247	5.2	187	75.7	60	24.3		
Depression	888	18.8	606	68.2	282	31.8		
Anxiety	917	19.4	655	71.4	262	28.6		
Severe mental illness	65	1.4	46	70.7	19	29.2		
Chronic pain/ Fibromyalgia	899	19.0	574	63.8	325	36.2		
Orthopaedic injury	396	8.4	321	81.0	75	18.9		
Other	446	9.4	349	78.3	97	21.7		
First language								
English	3872	86.0	2801	72.3	1071	27.7	108.97*	.024
Other	629	14.0	325	51.7	304	48.3		
Oral English fluency								
Fluent	4622	98.0	3245	70.2	1377	29.8	9.71*	.002
Not fluent	94	2.0	52	55.3	42	44.6		

Note. Valid Group = Participants who passed the Word Memory Test (WMT; Green, 2003); Invalid Group = Participants who failed the WMT; TBI = traumatic brain injury.
[a]$n = 4721$. [b]$n = 3297$. [c]$n = 1424$.
*$p < .05$.

Table 5 summarizes characteristics of the assessment context for the overall sample, valid group, and invalid group. Although results of chi-square tests of independence were statistically significant for assessment type (χ^2 (2, $N = 4669$) = 43.61, $p < .001$) and referral source (χ^2 (4, $N = 4644$) = 28.22, $p < .001$), effect sizes were very small, indicating the absence of any meaningful relationship between these variables and WMT classification. The significant findings are likely due to the analyses being overpowered by the large sample size.

Table 5

Characteristics of the Assessment Context for the Overall Sample and the Valid and Invalid Groups

	Overall Sample[a]		Valid Group[b]		Invalid Group[c]			
	N	%	*n*	%	*n*	%	χ^2	Φ^2
Assessment type								
Psychological	2392	51.2	1620	67.7	772	32.2	43.61*	.009
Neuropsychological	1743	37.3	1198	68.7	545	31.2		
Vocational	534	11.4	438	82.0	96	17.9		
Referral Source								
WCB/Legal	3414	72.5	2332	68.3	1082	31.7	28.22*	.006
Insurance	843	17.9	608	72.1	235	27.9		
Canada Pension	27	0.6	12	44.4	15	55.6		
Private	64	1.4	49	76.5	15	23.4		
Other	296	6.3	235	79.4	61	20.6		

Note. Valid Group = Participants who passed the Word Memory Test (WMT; Green, 2003); Invalid Group = Participants who failed the WMT; TBI = traumatic brain injury.
[a]$N = 4721$. [b]$n = 3297$. [c]$n = 1424$.
*$p < .05$.

Measures

As part of each assessment, participants were administered a comprehensive test battery containing cognitive tests, self-report symptom and personality inventories, as well as various PVTs and SVTs. Due to a flexible approach to test selection, however, not all participants were administered all tests and some PVTs were administered much more consistently than others.

Of all the measures in the database, only those relevant to this dissertation are described below. Standard cutoffs on free-standing PVTs represent those reported in the respective test manuals. All other cutoffs described below reach the minimal threshold for specificity (.84; Larrabee, 2003) and most exceed .90. Cutoffs producing an unacceptably high proportion of false-positive errors (i.e., unacceptably low specificity) are explicitly noted as such.

Free-standing PVTs. The dissertation includes a total of four free-standing PVTs.

Word Memory Test (WMT; Green, 2003). The WMT was used as the main criterion PVT. The WMT is a verbal memory task in which the examinee is presented twice with a list of 20 semantically-linked word pairs. The presentation of stimuli is followed by a two-alternative forced-choice Immediate Recognition (IR) trial in which each target is paired with a foil and the examinee is asked to select the target. Auditory and visual feedback is provided, which assists with learning for later trials. After a 30-minute delay, a second two-alternative forced-choice recognition trial is administered (Delayed Recognition or DR), which includes new foils. The DR subtest is followed by the Multiple Choice (MC), Paired Associates (PA), and Free Recall (FR) subtests, respectively. The WMT may be administered orally or on the computer.

Performance validity is determined using scores on IR, DR, and the consistency (CNS) of responses between IR and DR, where ≤82.5% accuracy on one or more of these indicators is considered an overall *Fail* (Green, 2003). Children with various neurodevelopmental conditions have been shown to pass the WMT performance validity indicators (Carone, 2014; Green, Lees-Haley, & Allen, 2003), suggesting that the WMT performance validity subtests place minimal demands on attention, language abilities, and memory. The test does, however, require a 3rd grade English reading level (Green, 2003).

Patients with very severe cognitive impairment (e.g., as seen in dementia) exhibit higher BR_{Fail} on the WMT. To protect against false-positive errors in such groups, a genuine memory impairment profile (GMIP) has been developed for the instrument (Green, 2003), which further distinguishes severe cognitive impairment from noncredible performance in cases where standard cutoffs are failed. The GMIP criteria rely on scores from the easy (performance validity) subtests, scores from the difficult (memory) subtests, and information about the patient's clinical history and presentation. The GMIP has been shown to reduce false positives

on the WMT in groups with very severe verbal memory impairment (Green, Montijo, & Brockhaus, 2011).

The WMT is well studied in the literature, with much of the validation research involving individuals seeking compensation rather than simulators (Green et al., 2003; Strauss et al., 2006). The WMT has been found to have high sensitivity and high specificity (Tan, Slick, Strauss, & Hultsch, 2002), and to differentiate valid vs. invalid groups with a large effect size in several studies (d = 1.07-3.07; Batt, Shores, & Chekaluk, 2008; Greve, Ord, Curtis, Bianchini, & Brennan, 2008; Hubbard, 2008; Lindstrom, Lindstrom, Coleman, Nelson, & Gregg, 2009). Failure on the WMT has been associated with lower scores on a large battery of neuropsychological tests in a compensation-seeking sample (Green, Rohling, Lees-Haley, & Allen, 2001; Green, 2007). Finally, in a survey of neuropsychologists, the WMT was identified as one of the top five most accurate tests for the detection of invalid performance (Sharland & Gfeller, 2007).

In the present dissertation, the WMT was used as an independent criterion for performance validity at standard cutoffs. The GMIP was not used due to several reasons: patients with very severe cognitive and functional impairment are rarely, if ever, assessed at the practices from which the data were obtained; the small group of patients who were assigned primary diagnoses of probable dementia were excluded from analyses; and information about patients' clinical histories and/or presentations was not available.

Medical Symptom Validity Test (MSVT; Green, 2004). Modeled after the WMT, the MSVT contains 10 word pairs, with each pair representing a single concept (e.g., ballpoint pen). The word pairs are presented twice at the beginning of the test and followed by an IR trial. The remaining subtests (DR, PA, and FR) are administered 10 minutes after the completion of IR.

The test may be administered orally or on a computer, although there is evidence to suggest that the computerized version is more sensitive to invalid performance than the oral version (Green, 2004). The IR, DR, and CNS scores represent the primary indicators of performance validity on the MSVT (Green, 2004), with scores ≤85% on one or more of these subtests producing a failure on the test. Standard cutoffs were used for the purposes of the present dissertation.

A GMIP has been developed for this instrument as well (Green, 2004) and has been found to reduce false positives in patients referred to a memory disorders clinic (Howe et al., 2007; Howe & Loring, 2009), patients with very severe verbal memory impairment (Green et al., 2011), and patients with dementia (Singhal et al., 2009). However, recent research suggests that the GMIP is observed in a substantial proportion of individuals asked to simulate dementia (Armistead-Jehle & Denney, 2014) and in veterans between the ages of 18 and 64 years, who are highly unlikely to have dementia (Reslan & Axelrod, 2017). These findings highlight the importance of interpreting MSVT scores in the context of clinical history and presentation. Given that such information was unavailable in the data set used for this dissertation, the GMIP was not used.

Non-verbal Medical Symptom Validity Test (NV-MSVT; Green, 2008). The NV-MSVT is a computerized visual-memory based PVT that consists of a total of six subtests, two of which assess memory. The stimuli consist of 10 artist-drawn coloured images, each of which contains two items that tend to be associated with each other (e.g., horse and cart). The set of 10 images is presented twice on a computer screen and then the examinee is asked to select the targets as part of a two-alternative forced-choice IR trial. A 10-minute delay follows the IR subtest. At the beginning of the delay, the examinee is presented with a sheet of paper containing several images, which are used as foils in the DR subtest to be administered after the delay. The images

on this sheet are "degraded" by white lines that are drawn through them. Following the delay, the examinee is presented with a two-alternative forced-choice recognition task containing three types of test items and is instructed to choose the images that were seen earlier *on the computer*. The three types of test items are used to compute three different subtest scores. They include: (1) the original target stimuli paired with the images on the degraded foil sheet (called *Delayed Recognition* or DR for scoring purposes), (2) the foils from the IR trial paired with archetypal images such as a snake or a bat (*Delayed Recognition-Archetypes* or DRA), and (3) the original target stimuli paired with similar images containing slight variations such as a horse and cart versus a horse and cart with a wheel missing (*Delayed Recognition-Variations* or DRV). The delayed recognition trial is followed by the PA and FR subtests, respectively.

The NV-MSVT yields six subtest scores (IR, DR, DRA, DRV, PA, and FR) and a CNS score. Invalid performance is defined by the following algorithm: (1) mean of IR, DR, CNS, DRA, DRV, and PA ≤90% (Criterion A1), and/or (2) mean of DR, CNS, DRA, and DRV is <88% (Criterion A2; Green, 2008). For the purpose of this dissertation, the standard cutoffs (Criteria A1 and A2) were used on the NV-MSVT in all analyses. Although a GMIP has been identified on this test (Green, 2008) and has been shown to reduce false positives in patients with dementia (Henry et al., 2009; Singhal et al., 2009), it was not used.

Test of Memory Malingering (TOMM; Tombaugh, 1996). The TOMM is a 50-item picture recognition test, which includes two learning trials and a retention trial. Stimuli consist of line drawings of common objects. In each learning trial (Trial 1 and Trial 2), the examinee is presented with 50 stimuli, one after the other, at a rate of 3 seconds per picture. Following the first presentation of the list (Trial 1), a two-alternative forced-choice recognition task is administered where each target is paired with a foil and the examinee is asked to select the target.

Feedback is provided after each item. Immediately after this task, the set of 50 drawings is presented again, in a different order than the first presentation (Trial 2), and followed with a second two-alternative forced-choice recognition task with new foils. Again, feedback is provided to the examinee. The Retention Trial consists of only the forced-choice recognition task and is administered after approximately a 20-minute delay. According to the test manual (Tombaugh, 1996), the Retention Trial of the TOMM is optional except when the performance validity criterion involving Trial 2 is failed. Tombaugh (1996) suggests that a score <45 on Trial 2 indicates possible invalid performance and should be followed with administration of the Retention Trial. A score of <45 on the Retention Trial provides further evidence of invalid performance.

Although standard cutoffs on the TOMM demonstrate adequate specificity in most individuals with cognitive impairment (Duncan, 2005; Teichner & Wagner, 2004), they have been found to be less sensitive to invalid performance as compared to other free-standing PVTs. For example, in one study, a cutoff of Trial 2 <45 correctly identified less than half (44%) of those who failed the WMT (Greiffenstein, Greve, Bianchini, & Baker, 2008). In another study (Armistead-Jehle & Gervais, 2011), this cutoff detected only 21% of claimants failing the WMT, 32% of those failing the NV-MSVT, and 35% of those failing the MSVT. The standard cutoff on the TOMM also demonstrated low sensitivity (.42) against a criterion measure derived from multiple PVTs (Kulas, Axelrod, & Rinaldi, 2014).

Over the years, researchers have proposed more liberal cutoffs on TOMM Trial 2 to improve its sensitivity to invalid performance, while retaining acceptable specificity. For example, Trial 2 ≤46 and ≤47 detected 55% of individuals with moderate-to-severe TBI who were identified as demonstrating probable/definite MND (Greve, Bianchini, & Doane, 2006).

Trial 2 ≤48 correctly identified 73% of individuals feigning low IQ (Smith et al., 2014) and 70% of mTBI litigants exhibiting probably/definite MND (Greve, Bianchini, & Doane, 2006a). Trial 2 ≤49 correctly detected 55% of TBI patients and 39% of psychiatric patients failing the WMT, with good specificity (.91 and .96, respectively; Erdodi & Rai, 2017). The same cutoff also achieved .55 sensitivity in a sample of patients with malingered pain-related disability (Greve, Etherton et al., 2009), and sensitivities of .61, .77, and .77, respectively, in individuals with toxic exposure (Greve et al., 2006b), mTBI litigants (Greve et al., 2006a), and forensic examinees (Schroeder et al., 2013) exhibiting probable/definite MND, while retaining acceptable specificity. In another study (Jones, 2013), Trial 2 ≤49 detected 96% of those obtaining below-chance performance on the Victoria Symptom Validity Test (Slick et al., 1997), 86% of those failing a total of three PVTs and/or SVTs (with no scores below chance), and 56% of those failing a total of two PVTs and/or SVTs (with no scores below chance) at .96 specificity.

In addition to exploring more liberal cutoffs on Trial 2, researchers have investigated the utility of Trial 1 as an indicator of performance validity against a variety of criterion measures including the standard administration of the TOMM (Bauer et al., 2007; Fazio et al., 2016; Hilsabeck et al., 2011; Horner et al., 2006; Wisdom et al., 2012), other free-standing PVTs used alone (Denning, 2012) and in combination (Jones, 2013; Kulas, Axelrod, & Rinaldi, 2014), and criteria for probable or definite MND (Greve et al., 2006a; Greve et al., 2006b; O'Bryant et al., 2007; Smith et al., 2014) and MPRD (Greve, Etherton et al., 2009). Interestingly, there is considerable variability in the cutoffs identified as optimal across studies, with Trial 1 ≤35 representing the most conservative cutoff (Horner et al., 2006) and ≤45 the most liberal cutoff (Greve et al., 2006b). This variability appears to be due, in large part, to the criterion measures used in different studies. In general, studies evaluating Trial 1 cutoffs against the standard

administration of the TOMM appear to suggest more conservative cutoffs (e.g., Trial 1 ≤34 through ≤39; Bauer et al., 2007; Fazio, 2016; Horner et al., 2006; Wisdom et al., 2012), while those evaluating Trial 1 cutoffs against MND or MPRD criteria suggest more liberal cutoffs (e.g., ≤43 and ≤44; Greve et al., 2006a; Greve et al., 2006b; O'Bryant et al., 2007; Schroeder et al., 2013). These findings are not surprising given the more recent research on the TOMM, in which standard cutoffs have been shown to be overly conservative and less sensitive to noncredible performance compared to other free-standing PVTs (Armistead-Jehle & Gervais, 2011; Greiffenstein et al., 2008; Kulas et al., 2014). Nevertheless, in an attempt to review and combine TOMM Trial 1 findings, Denning (2012) reported that a high level of classification accuracy (mean weighted sensitivity of .77 and specificity of .92) can be achieved using a Trial 1 cutoff of ≤40 in most populations, although lower cutoffs are needed for individuals with dementia or severe amnestic disorders (Greve et al., 2006; Hilsabeck, Gordon, Hietpas-Wilson, & Zartman, 2011).

Scores on TOMM Trials 1 and 2 were used for the purpose of this dissertation. Scores from the Retention Trial were excluded as the Retention Trial is administered very infrequently in the practices from which the data were obtained.

Embedded validity indicators. A total of eleven embedded validity indicators, derived from five different neuropsychological tests, were used in this dissertation. Each of these indicators has been examined previously in the literature.

Six of the embedded validity indicators included in this dissertation represent scores that are traditionally reported as part of standard scoring procedures for the respective tasks (TMT-A, TMT-B, FTT-DH, FTT-NDH, CT-TE, WCST FMS); three represent scores that are derived from performance on clinical tests of neuropsychological ability but not routinely reported as part of

standard scoring procedures (RDS, TMT A+B, FTT-C); and two represent comparison scores used to identify atypical patterns of performance (TMT-B/A, FTT-DIFF).

Wechsler Adult Intelligence Scale—Third and Fourth Editions (WAIS-III, Wechsler, 1997; WAIS-IV, Wechsler, 2008). The Digit Span subtest of the WAIS is a measure of auditory attention and working memory in which the examinee is asked to repeat strings of digits (of increasing length) after the examiner in forward and backward order. Based on this test, Greiffenstein, Baker, and Gola (1994) developed an embedded validity indicator called the Reliable Digit Span (RDS), which is calculated by summing the highest number of digits that is repeated reliably (across both of two trials) in the forward and backward conditions.

In Greiffenstein and colleagues' (1994) original study, an RDS cutoff of ≤7 correctly identified 70% of probable malingerers, 89% of those with persistent post-concussive syndrome, and 73% of closed head injury patients. In subsequent studies, a cutoff of ≤7 detected 48.9% of litigating patients with mild closed head injury (Meyers & Volbrecht, 1998); 67% mixed-severity TBI referrals identified as MND (Mathias, Greve, Bianchini, Houston, & Crouch, 2002); 50-52% of moderate-severe closed head injury MND patients (Larrabee, 2003; Heinly, Greve, Bianchini, Love, & Brennan, 2005); and 60% of chronic pain MND patients (Etherton, Bianchini, Greve, & Heinly, 2005). However, the original cutoff of ≤7 was found to produce unacceptably high false-positive error rates in a number of clinical groups including those with psychiatric difficulties (28%; Heinly et al., 2005), cerebrovascular accidents (44%; Heinly et al., 2005), memory disorders (52%; Heinly et al., 2005), and a mixed clinical group referred for neuropsychological testing (23%; Babikian, Boone, Lu, & Arnold, 2006).

A more conservative cutoff of RDS ≤6 has been found to correctly classify 56.6% of malingering pretrial and pre-sentence detainees (Duncan & Ausborn, 2002); 46% of mTBI cases

identified as MND (Heinly et al., 2005); and 66% of patients in a mixed clinical sample displaying noncredible test performance in the context of external incentives (Babikian et al., 2006). Sensitivity was modest in two samples of veterans: .18 in one study (Spencer, Axelrod, Drag, Waldron-Perrine, Pangilinan, & Bieliauskas, 2013) and .24 in another (Young, Sawyer, Roper, & Baughman, 2012). Notably, RDS ≤6 was found to produce unacceptably high false positive error rates in patients with moderate (24%) and severe (83%) dementia (Kiewel, Wisdom, Bradshaw, Pastorek, & Strutt, 2012) and patients with schizophrenia spectrum disorders (Glassmire, Ross, Kinney, & Nitch, 2016).

Consistent with these findings, in a systematic review and cross-validation study (Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012), RDS ≤7 achieved good sensitivity (.48 when using weighted averages) but was found to have specificity rates below 90% for the pooled data and for all clinical subgroups. A cutoff of ≤6 produced lower sensitivity rates (.30 when using weighted averages) but increased specificity to adequate levels (i.e., above 90%) for most clinical subgroups. Notably, a few diagnostic groups consistently produced specificity rates below 90% for both cutoff scores (≤7 and ≤6). These included individuals with cerebrovascular accident, severe memory disorders (e.g., dementia), intellectual disability, borderline IQ and below, and English as a second language.

An RDS cutoff of ≤5 has been found to be associated with perfect and near-perfect (.98) specificity in TBI referrals identified as non-malingering (Mathias et al., 2002) and veterans with various diagnoses who passed the WMT (Young et al., 2012), respectively. This cutoff also demonstrated adequate specificity (.84; Larrabee, 2003) in individuals with schizophrenia spectrum disorders who had Full Scale IQs of 70 and higher (Glassmire et al., 2016). Furthermore, in another study (Zenisek, Millis, Banks, & Miller, 2016), while RDS cutoffs of ≤7

and ≤6 were associated with high BR_{Fail} among individuals suspected for dementia, scores of ≤5 were found to be infrequent (BR_{Fail} = 4.3%). These findings suggest that a score of ≤5 is highly specific to noncredible performance.

Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The WCST is a measure of concept formation and abstract problem solving, which requires examinees to match cards based on changing rules. Standard scoring procedures yield scores on several different indicators of performance, one of which is Failure to Maintain Set (FMS). An FMS error is recorded when an examinee matches a card incorrectly after making five or more consecutive correct responses, but before successfully completing the operating category. The present dissertation investigates the FMS raw score as an indicator of performance validity.

Larrabee (2003) proposed an FMS cutoff of ≥2, which correctly identified 48% of definite MND cases in his study. In another study, a cutoff of ≥3 correctly identified 29% of mTBI cases classified as MND (Greve, Heinly, Bianchini, & Love, 2009). Although specificity rates were adequate in patients with mTBI, moderate-severe TBI, psychiatric disorders, and cerebrovascular accidents, an unacceptably high proportion of false positives were observed in individuals with memory disorders (.17) and severe TBI (.18; Greve, Heinly et al., 2009).

Category Test (CT; DeFlippis & McCampbell, 1979, 1991, 1997). The CT is measure of concept formation, novel problem solving, and the ability to learn from experience (Strauss et al., 2006), which can be administered in booklet or computer form (Choca & Morris, 1992). The test consists of 208 items. Although several CT scores have been examined for their utility as PVTs, the present dissertation only examines the raw number of Total Errors (CT-TE).

In an analogue-simulation design, CT-TE ≥87 correctly detected 51.1% of simulating malingerers (Tenhula & Sweet, 1996). Similarly, in a subsequent study, CT-TE ≥87 correctly

classified 76.7% of uncoached simulators, 34.4% of coached simulators, 100% of optimal controls, and 83.3% of TBI patients (DiCarlo, Gfeller, & Oliveri, 2000).

Greve, Bianchini, and Roberson (2007) also reported classification accuracy data for several CT-TE cutoffs. CT-TE ≥90 achieved .41 sensitivity in individuals with mixed-severity TBI identified as MND, although false-positive error rates in non-TBI general clinical patients were unacceptably high (18%). CT-TE ≥100 produced .25 and .18 sensitivity in TBI and general clinical patients, respectively, with adequate specificity. Finally, a cutoff of ≥120 was associated with zero false positives in TBI and a cutoff of ≥140 was associated with zero false positives in a group of general clinical patients.

Trail Making Test (TMT; Reitan & Wolfson, 1985). The TMT includes two tasks: Part A, which assesses processing speed and attention, and Part B, which assesses cognitive flexibility. The present dissertation utilized four embedded PVTs derived from the standard administration of the TMT including Part A completion time (TMT-A; raw score in seconds), Part B completion time (TMT-B; raw score in seconds), sum of completion times on Parts A and B (TMT A+B; in seconds), and the ratio of Part B completion time (in seconds) to Part A completion time (in seconds; TMT-B/A). The logic of TMT-B/A is based on the notion that examinees should take approximately twice as long to complete TMT-B than TMT-A due to the added cognitive load (Lamberty, Putnam, Chatel, Bieliauskas, & Adams, 1994). Therefore, as a validity indicator, TMT-B/A is designed to detect a violation of the normative difficulty gradient (i.e., performing better on a more difficult task compared to an easier task).

TMT-A ≥63" correctly identified 16.7% of individuals with mTBI and 11.1% of those with well-defined TBI who failed at least one free-standing PVT, with perfect and near-perfect (.98) specificity, respectively (Iverson, Lange, Green, & Franzen, 2002).

A TMT-B cutoff of ≥200" produced good specificity in mild and well-defined TBI with sensitivity rates of 7.1% and 18.5%, respectively (Iverson et al., 2002). In consecutive patients assessed in a multi-specialty private practice and identified as biased or unbiased responders based on the TOMM, Busse and Whiteside (2012) found .61 sensitivity using TMT-B ≥120" with acceptable specificity.

TMT A+B ≥170" correctly identified 48% of invalid response sets on the TOMM (Busse & Whiteside, 2012) while maintaining acceptable specificity, but was found to be overly conservative (SENS = .11, SPEC = .99) in a sample of post-deployment veterans (Shura Miskey, Rowland, Yoash-Gantz, & Denning, 2016). In the Shura et al. (2016) study, TMT A+B ≥137" achieved .90 specificity and .21 sensitivity.

Finally, a cutoff of ≤1.49 on TMT-B/A correctly identified 2.4% (mTBI) and 7.4% (well-defined TBI) of those failing one or more free-standing PVTs (Iverson et al., 2002).

Finger Tapping Test (FTT; Reitan, 1969). Originally called the Finger Oscillation Test, the FTT assesses self-directed manual motor speed using a specially adapted tapper and counter (Strauss et al., 2006). Examinees are instructed to tap the index fingers of each hand on a key as quickly as possible over multiple 10-second trials. Although a few different administration and scoring procedures have been identified (Straus et al., 2006), the procedure of Reitan and Wolfson (1985) was used in the data collected for this dissertation. This procedure requires five consecutive 10-second trials within a range of five taps for each hand; scoring involves calculation of the mean over five trials. If the five-trials-within-five-taps criterion is not met, ten trials are administered for each hand and the score for each hand is the mean number of taps achieved over the 10 trials (Reitan & Wolfson, 1985).

Four indices based on the FTT have been investigated as embedded PVTs including the average number of taps achieved with the dominant hand (FTT-DH), the average number of taps achieved with the non-dominant hand (FTT-NDH), the sum of the average number of taps achieved with the dominant and non-dominant hands (FTT-Combined or FTT-C), and the difference between the average number of taps achieved with the dominant and non-dominant hands (FTT-DIFF). Like TMT-B/A, FTT-DIFF is designed to detect a violation of the normative difficulty gradient (i.e., performing better on a more difficult task compared to an easier task).

FTT-DH ≤43.6 produced .48 sensitivity in an mTBI group identified as invalid, while a more conservative cutoff of ≤37.7 was associated with zero false positives (Backhaus, Fichtenberg, & Hanks, 2004). Based on evidence for gender differences in finger tapping speed, Arnold, Boone, Lu, Dean, Wen, Nitch, and McPherson (2005) proposed cutoffs of ≤35 for men and ≤28 for women, which produced .50 and .61 sensitivity, respectively, in mixed clinical examinees identified as noncredible based on psychometric and behavioural criteria. Arnold and colleagues' (2005) FTT-DH cutoffs correctly identified 50% of forensic examinees and 36% of veterans failing two or more PVTs (Axelrod, Meyers, & Davis, 2014). However, FTT-DH ≤35 only achieved .12 sensitivity in another, mostly male sample of postdeployment veterans (Shura et al., 2016). Increasing the FTT-DH cutoff from ≤35 to ≤37 produced modest gains in sensitivity (.19) in the latter study, while maintaining adequate specificity. Finally, Arnold and colleagues' (2005) FTT-DH cutoff of ≤28 for females produced unacceptably high false-positive rates in credible female patients with low IQ; a cutoff of ≤23, however, achieved acceptable specificity (.91) and correctly detected 37% of women feigning low IQ (Smith et al., 2014). In this dissertation, FTT-DH was examined separately in males ($FTT\text{-}DH_M$) and females ($FTT\text{-}DH_F$).

FTT-NDH ≤38.9 was associated with .56 sensitivity in mTBI patients identified as invalid, while a cutoff of ≤32.9 was associated with .28 sensitivity and no false-positive errors (Backhaus et al., 2004). In Arnold and colleagues' (2005) study, cutoffs of ≤30 and ≤25 produced .36 sensitivity in males and females, respectively, in the noncredible sample. The cutoff of ≤25 for females, however, resulted in unacceptably low specificity rates (.81) in credible female patients with low IQ in another study (Smith et al., 2014); lowering the cutoff to ≤19 resulted in correct detection of 31% of women feigning low IQ with adequate specificity. In this dissertation, FTT-NDH was examined separately in males ($FTT\text{-}NDH_M$) and females ($FTT\text{-}NDH_F$).

FTT-C ≤62 correctly identified 40% of malingerers (Larrabee, 2003) and 55% of individuals who failed two or more PVTs in a forensic context (Axelrod et al., 2014). Cutoffs of ≤66 for men and ≤58 for women achieved .43 to .55 sensitivity in the noncredible sample (Arnold et al., 2005). In this dissertation, FTT-C was examined separately in males ($FTT\text{-}C_M$) and females ($FTT\text{-}C_F$).

Finally, FTT-DIFF cutoffs of ≤-2 in men and ≤-5 in women achieved .21 sensitivity in the noncredible sample with acceptable specificity rates (Arnold et al., 2005). In this dissertation, FTT-DIFF was examined separately in males ($FTT\text{-}DIFF_M$) and females ($FTT\text{-}DIFF_F$).

Data Analysis

Studies 1 and 2. In Studies 1 and 2, performance was examined on a total of 14 PVTs of interest, which included three free-standing PVTs and 11 embedded validity indicators. Among the free-standing PVTs were the MSVT, NV-MSVT, and TOMM (Trials 1 and 2). Embedded PVTs included the RDS, WCST FMS, CT-TE, TMT-A, TMT-B, TMT A+B, TMT-B/A, FTT-DH, FTT-NDH, FTT-C, and FTT-DIFF.

In Study 1, performance on each PVT of interest was characterized in the overall sample (N = 4721) and across a number of key patient variables using relative cumulative frequency distributions. Specifically, BR_{Fail} were reported at various cutoffs on each PVT of interest in the overall sample and across nine diagnostic groups, five age groups, six levels of education, gender (i.e., males versus females), and English language status (i.e., native speakers of English [NSE] versus non-native English speakers [ESL]). Diagnostic groups included mTBI, moderate-severe TBI, neurological conditions, depression, anxiety, bipolar disorder/psychosis, chronic pain/fibromyalgia, orthopedic, and *other* diagnosis. Age groups included participants aged 16-29 years, 30-39 years, 40-49 years, 50-59 years, and 60-69 years. Education groups included those with ≤8 years, 9-11 years, 12 years, 13-15 years, 16 years, and ≥17 years of formal education.

In addition to the aforementioned variables, in Study 1, BR_{Fail} on each PVT of interest were reported as a function of outcome on a well-validated free-standing PVT: the WMT (Green, 2003). In previous research, the prevalence of PVT failure/invalid responding in settings with potential for secondary gain has been estimated to fall between 30% and 40% (e.g., Larrabee, 2003; Mittenberg et al., 2002). Given that the majority of examinees in this data set had external incentives to appear impaired, it was expected that neuropsychological test data from a significant proportion of cases in the overall sample would be invalid. Thus, to examine the impact of invalid performance on PVTs of interest, in Study 1, BR_{Fail} on each PVT of interest were also reported separately for those who passed versus failed the WMT.

Because a substantial proportion of the overall sample was expected to demonstrate invalid performance, invalid performance represented a significant confound with respect to determining the true relationships between BR_{Fail} on PVTs of interest and the patient variables examined in Study 1. As a result of this, in Study 2, a valid subsample was created, which

consisted of only those participants in the overall sample who passed the WMT at standard cutoffs (n = 3297; referred to as the valid sample). Data were analyzed in a similar manner to Study 1. Specifically, BR_{Fail} were reported at various cutoffs on each PVT across the total valid sample, nine diagnostic groups, five age groups, six levels of education, as well as gender (i.e., male or female) and English language status (i.e., NSE and ESL).

Throughout Studies 1 and 2, chi-square tests of independence and Fisher's exact tests were used to determine whether scores on each PVT of interest were related to the demographic/background variables examined (diagnosis, age, education, gender, English language status) and to performance on the WMT (in Study 1 only) at various cutoffs. Chi-square tests were conducted if (1) the value of expected cell counts was 5 or more in at least 80% of cells, and (2) no cells had an expected cell count of less than 1 (McHugh, 2013), whereas Fisher's exact tests were conducted in cases where cell sizes were small and the aforementioned assumptions were not met. Neither test was performed if one or more groups had a BR_{Fail} of zero.

For the chi-square and Fisher's exact test analyses, Phi squared (Φ^2) was used as an estimate of effect size for analyses that involved variables with multiple levels (i.e., diagnosis, age, education). Likelihood ratios were used as an estimate of effect size for analyses that involved variables with only two levels (i.e., gender [male or female], English language status [NSE or ESL], and WMT [pass or fail]). Likelihood ratios were calculated by dividing the cumulative percentage of one group (e.g., males) failing a given cutoff by the cumulative percentage of the corresponding group (e.g., females) failing the same cutoff. The calculations were performed from left to right in the relevant tables, such that cumulative percentages for males, NSEs, and those who passed the WMT served as numerators in each computation, and

cumulative percentages for females, ESL patients, and those who failed the WMT served as denominators. As a result of this, the reported ratios reflect the likelihood of males, NSEs, and individuals who passed the WMT failing a given PVT of interest cutoff relative to their counterparts. With regard to interpretation, likelihood ratios at or around 1.00 reflect comparable BR_{Fail} among the two groups being compared. As likelihood ratios become smaller (i.e., closer to zero), they indicate large differences in BR_{Fail} among the two groups, with lower rates in the left-column group (i.e., males, NSEs, those who passed the WMT). As likelihood ratios become larger (further away from 1.00 in the increasing direction; e.g., 2.00), they represent large differences in BR_{Fail} among the two groups, with higher rates in the left-column group.

Studies 1 and 2 contained a very large number of null-hypothesis significance tests, which inflated the Type I error rate. This problem was further compounded by the large sample size, as a result of which many of the analyses were likely overpowered. In light of these facts, effect sizes were emphasized in the interpretation of Studies 1 and 2 results rather than the statistical significance of chi-square tests or Fisher's exact tests.

Given evidence for gender differences on the FTT (e.g., Arnold et al., 2005), it was decided that data on FTT-based PVTs would be reported separately for males and females in Studies 1 and 2. This precluded chi-square analyses involving gender for these PVTs. Further examination of the dataset revealed that the FTT was administered only about half as often to females as it was administered to males. As a result of the small sample size, data were not reported for FTT-based indicators in females as part of Studies 1 and 2.

Selection of cutoffs for Studies 1 and 2. In order to adequately characterize performance on the PVTs of interest, multiple scores on each PVT of interest were selected at which cumulative percentages (i.e., BR_{Fail}) would be reported across various groups. To select these

scores, the PVTs of interest were conceptually separated into those on which scores fell into a restricted or relatively-restricted range and those on which scores varied widely. Three PVTs of interest were considered to fall in the former category (TOMM Trials 1 and 2, RDS, WCST FMS), while nine PVTs of interest were thought to fall in the latter category (CT-TE, TMT-A, TMT-B, TMT A+B, TMT-B/A, FTT-DH, FTT-NDH, FTT-C, FTT-DIFF). For PVTs of interest with a restricted range of possible scores, cumulative percentages were reported at multiple scores, including values below, above, and corresponding to previously-identified cutoffs. For PVTs of interest on which scores could vary widely, a methodology similar to that of Pearson (2009) was used: scores at or near the 2^{nd}, 5^{th}, 10^{th}, 15^{th}, 25^{th} percentiles in the overall sample were identified, and cumulative percentages were reported at these scores in both Studies 1 and 2.

While BR_{Fail} were reported at multiple cutoffs on most PVTs of interest, on the MSVT and NV-MSVT, BR_{Fail} were only reported at standard cutoffs specified in the test manuals (Green, 2004, 2008). This is because these tests use fixed, published algorithms based on multiple subtests, rather than single scores, to make determinations about performance validity.

It is important to note that the sample sizes for data reported in Studies 1 and 2 varied depending on a number of factors. First, sample sizes were affected by how widely a given PVT was administered. Second, because the valid sample represents a subset of the overall sample, for each PVT of interest, analyses involving the overall sample contained more cases than analyses involving the valid sample. Finally, considering the skewed distributions of most PVT scores, it is not only rare for examinees to demonstrate poor performance on these tests, but also, the *worse* the performance, the *less* likely it is to be observed. Thus, for each PVT of interest, sample sizes were larger at more liberal cutoffs as compared to conservative cutoffs. In light of these

issues, readers are encouraged to note the sample size associated with each PVT and each cutoff when examining data for Studies 1 and 2.

Study 3. In Study 3, the classification accuracy of each PVT of interest was evaluated against the WMT (Green, 2003), which served as the criterion PVT. Invalid performance was defined as failure on the WMT at standard cutoffs specified in the test manual (Green, 2003). Scores on the all PVTs of interest were dichotomized such that a score of zero represented a *Pass* on that PVT and a score of one represented a *Fail*. For the MSVT and NV-MSVT, scores were dichotomized according to standard cutoffs (Green 2004, 2008). For each of the other PVTs of interest, multiple potential cutoffs were examined for their sensitivity and specificity against the WMT, including scores identified as cutoffs in previous literature as well as alternative scores that produced acceptable levels of specificity (.84-1.0; Larrabee, 2003). Positive and negative predictive values (PPP and NPP) were reported at five hypothetical base rates of invalid performance (10%, 20%, 30%, 40%, and 50%) for each cutoff.

Although the total sample used for Study 3 was very large ($N = 4721$), the number of cases included in specific analyses depended on how widely the relevant PVT of interest was administered. All cases containing data on a PVT of interest were included in the analyses for that PVT.

Once the classification accuracies of each PVT were computed, PVTs of interest were compared in terms of their relative sensitivities to invalid performance. The most liberal cutoff to achieve at least 90% specificity was identified on each PVT of interest, and sensitivity values at these cutoffs were compared across PVTs. Because multiple cutoffs were not examined on the MSVT and NV-MSVT, sensitivity values associated with standard cutoffs were used for these instruments.

CHAPTER 3

Results

Studies 1 and 2

Results from Studies 1 and 2 are presented together in this dissertation to facilitate comparison between the performances of the overall and valid samples on each PVT of interest. Data on the WMT (criterion PVT) is presented first and followed by the PVTs of interest, beginning with free-standing PVTs. For each PVT of interest, performance is initially characterized as a function of WMT result (*Pass/Fail*). Then, corresponding tables from Studies 1 and 2 (i.e., those examining BR_{Fail} on a particular PVT of interest across the same demographic/background variable) are presented one after the other for diagnosis, age, education, gender and English language status, in that order.

As mentioned above, given (1) the large sizes of the overall and valid samples, and (2) the inflated Type I error rate resulting from multiple comparisons, chi-square tests of independence and Fisher's exact tests were likely to reach statistical significance in the absence of meaningful effects. In light of this, when interpreting results from Studies 1 and 2, effect sizes are emphasized rather than the results of null-hypothesis significance tests.

WMT. Approximately one-third of the overall sample (30.2%) failed the WMT at standard cutoffs. Table 6 characterizes WMT BR_{Fail} as a function of diagnosis. The chi-square test of independence was statistically significant $\chi^2\ (8, N = 4721) = 97.66, p < .001$. However, diagnosis accounted for only 2.1% of the variance in BR_{Fail} (small effect). Patients with mild TBI and those with chronic pain/fibromyalgia demonstrated the highest BR_{Fail} (40.0% and 36.2%, respectively), while patients with *other* diagnoses (21.7%), moderate-to-severe TBI (21.4%) and orthopedic injuries (18.9%) had the lowest BR_{Fail}.

Table 6

Percentages of Participants Failing the Standard Cutoff on the Word Memory Test in the Valid Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
WMT	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
STN	1424	30.2	40.0	21.4	24.3	31.8	28.6	29.2	36.2	18.9	21.7	97.66*	.02

Note. Word Memory Test (Green, 2003); STN = standard cutoffs specified in the test manual; TBI = traumatic brain injury, Mild = mild TBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other.
[a]$n = 4721$. [b]$n = 643$. [c]$n = 220$. [d]$n = 247$. [e]$n = 888$. [f]$n = 917$. [g]$n = 65$. [h]$n = 899$. [i]$n = 396$. [j]$n = 446$.
*$p < .05$.

Table 7 characterizes WMT BR_{Fail} as a function of age. The chi-square test of independence was statistically significant, $\chi^2\ (4, N = 4127) = 60.67, p < .001$, although age only accounted for 1.3% of the variance in BR_{Fail} (small effect). The youngest group (16-29 years) had the lowest BR_{Fail}, and BR_{Fail} increased with older age.

Table 7

Percentages of Participants Failing the Standard Cutoff on the Word Memory Test in the Valid Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
WMT	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
STN	1424	30.2	19.8	28.7	31.2	34.5	40.8	60.67*	.013

Note. Word Memory Test (Green, 2003); STN = standard cutoffs specified in the test manual.
[a]$n = 4721$. [b]$n = 718$. [c]$n = 1159$. [d]$n = 1480$. [e]$n = 1133$. [f]$n = 218$.
*$p < .05$.

Table 8 characterizes WMT BR_{Fail} as a function of education level. The chi-square test of independence was statistically significant, $\chi^2\ (5, N = 4127) = 82.46, p < .001$. Education level accounted for only 1.9% of the variance in BR_{Fail} (small effect). Patients with ≤8 years of formal education demonstrated the highest BR_{Fail} (46.9%), followed by those with 9-11 years of

education (34.7%), and BR_{Fail} were generally comparable (at about 25.0%) among those with 12 and more years of education.

Table 8

Percentages of Participants Failing the Standard Cutoff on the Word Memory Test in the Valid Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
WMT	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
STN	1424	30.2	46.9	34.7	27.9	23.8	25.6	26.0	82.46*	.017

Note. Word Memory Test (Green, 2003); STN = standard cutoffs specified in the test manual.
[a]$n = 4721$. [b]$n = 343$. [c]$n = 1265$. [d]$n = 1769$. [e]$n = 773$. [f]$n = 363$. [g]$n = 204$.
*$p < .05$.

Table 9 characterizes WMT BR_{Fail} as a function of gender and English language status. Males demonstrated a somewhat higher BR_{Fail} than females, $\chi^2\ (1, N = 4127) = 23.70, p < .001$ (LR = 1.25), and the BR_{Fail} among native English speakers (NSEs) was approximately half that of patients who spoke English as a second language (ESL), $\chi^2\ (1, N = 4127) = 108.97, p < .001$ (LR = 0.54).

Table 9

Percentages of Participants Failing the Standard Cutoff on the Word Memory Test in the Valid Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
WMT	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
STN	1424	30.2	32.8	26.2	23.70*	1.25	26.2	48.3	108.97*	0.54

Note. Word Memory Test (Green, 2003); STN = standard cutoffs specified in the test manual; LR = likelihood ratio; NSE = Native speakers of English; ESL = English as a second language.
[a]$n = 4721$. [b]$n = 2817$. [c]$n = 1904$. [d]$n = 3872$. [e]$n = 629$.
*$p < .05$.

MSVT. In the overall sample, 21.3% of cases failed the MSVT at standard cutoffs. Table 10 shows that there was considerable overlap between performance on the MSVT and performance on the WMT, χ^2 (1, N = 2165) = 580.66, p <.001. Participants who passed the WMT (i.e., the valid sample) demonstrated a lower MSVT BR_{Fail} (6.8%) than those who failed the WMT (52.4%). Compared to participants who failed the WMT, those who passed the WMT were only 0.13 as likely to fail the MSVT.

Table 10

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
MSVT	*f*	%	Pass[b]	Fail[c]	χ^2	LR
STN	462	21.3	6.8	52.4	580.66*	0.13

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the Word Memory Test at standard cutoffs; LR = likelihood ratio.
[a]n = 2165. [b]n = 1476. [c]n = 689.
*p < .05.

Tables 11 and 12 characterize MSVT BR_{Fail} as a function of clinical diagnosis in the overall and valid samples, respectively. Effect sizes indicated that diagnosis accounted for only 0.3% of the variance in MSVT BR_{Fail} in the overall sample and only 0.5% of the variance in MSVT BR_{Fail} in the valid sample (very small effects). Although the contrast was statistically significant in the overall sample, this was likely an artifact of the large sample size.

Visual inspection of BR_{Fail} indicates that patients with mTBI (38.6%) and orthopedic injury (32.7%) demonstrated the highest BR_{Fail} in the overall sample, while those with anxiety (15.7%) and moderate-severe TBI (14.1%) had the lowest BR_{Fail}. In the valid sample, BR_{Fail} were comparable across most groups at about 9%, with lower rates among participants with anxiety (5.0%), depression (5.6%), and other diagnoses (6.9%).

Table 11

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Overall Sample and Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
MSVT	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
STN	462	21.3	38.6	14.1	21.0	20.4	15.7	22.5	24.9	32.7	17.2	55.05*	.00

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other.
[a]*n* = 2165. [b]*n* = 184. [c]*n* = 85. [d]*n* = 124. [e]*n* = 592. [f]*n* = 562. [g]*n* = 40. [h]*n* = 349. [i]*n* = 341. [j]*n* = 174.
*$p < .05$.

Table 12

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
MSVT	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
STN	101	6.8	9.7	8.7	9.5	5.6	5.0	8.3	9.0	9.4	6.9	7.89	.005

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other.
[a]*n* = 1476. [b]*n* = 103. [c]*n* = 69. [d]*n* = 95. [e]*n* = 396. [f]*n* = 403. [g]*n* = 24. [h]*n* = 223. [i]*n* = 32. [j]*n* = 131.
*$p < .05$.

Tables 13 and 14 characterize MSVT BR_{Fail} by age in the overall and valid samples, respectively. Although MSVT BR_{Fail} varied significantly across age groups in both samples, $\chi^2 (4, N = 2163) = 34.47, p < .001$ (overall sample) and $\chi^2 (4, N = 1475) = 15.67, p = .003$ (valid sample), age accounted for only 1.1 to 1.6% of the variance in MSVT BR_{Fail} (small effects).

In the overall sample, BR_{Fail} increased steadily with age and ranged from 9.6% (in individuals aged 16-29 years) to 27.9% (in those aged 60-69 years). In the valid sample, BR_{Fail} were comparable among individuals aged 16-49 years, and almost twice as high in older groups (≥50 years).

Table 13

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
MSVT	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
STN	462	21.3	9.6	20.8	21.0	25.9	27.9	34.47*	.016

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual.
[a]$n = 2165$. [b]$n = 281$. [c]$n = 433$. [d]$n = 696$. [e]$n = 613$. [f]$n = 140$.
*$p < .05$.

Table 14

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
MSVT	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
STN	101	6.8	4.8	4.1	5.9	10.7	9.1	15.67*	.011

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual.
[a]$n = 1476$. [b]$n = 231$. [c]$n = 293$. [d]$n = 460$. [e]$n = 403$. [f]$n = 88$.
*$p < .05$.

Tables 15 and 16 characterize MSVT BR_{Fail} as a function of education in the overall and valid samples, respectively. Effect sizes indicate that education accounted for only 1.5% of the variance in MSVT BR_{Fail} in the overall sample (small effect) and 0.4% of the variance in MSVT BR_{Fail} in the valid sample (very small effect). Although the chi-square test of independence reached statistical significance in the overall sample, $\chi^2\ (5, N = 2165) = 32.57, p < .001$, given the effect size, this finding was likely driven by the large sample size.

The highest BR_{Fail} were observed in individuals with the fewest years of formal education, and BR_{Fail} generally decreased with higher levels of education. Notably, in both samples, individuals with ≥17 years of formal education demonstrated higher BR_{Fail} than would be expected based on the trends.

Table 15

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
MSVT	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
STN	462	21.3	36.7	24.8	18.8	18.6	15.3	24.7	32.57*	.015

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual.
[a]$n = 2165$. [b]$n = 139$. [c]$n = 536$. [d]$n = 921$. [e]$n = 322$. [f]$n = 170$. [g]$n = 77$.
*$p < .05$.

Table 16

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
MSVT	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
STN	101	6.8	11.3	7.0	7.4	5.8	3.1	7.1	5.71	.004

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual.
[a]$n = 1476$. [b]$n = 71$. [c]$n = 496$. [d]$n = 649$. [e]$n = 242$. [f]$n = 128$. [g]$n = 56$.
*$p < .05$.

Tables 17 and 18 characterize MSVT performance as a function of gender and English language status in the overall and valid samples, respectively. Results of analyses involving gender showed that males were slightly more likely to fail the MSVT than females in both samples (LR = 1.22 in the overall sample; LR = 1.06 in the valid sample), although differences only reached statistical significance in the overall sample, $\chi^2\ (1, N = 2165) = 5.19, p = .023$.

With regard to English language status, in the overall sample, MSVT BR_{Fail} were significantly lower among NSEs than ESL patients, $\chi^2\ (1, N = 2157) = 52.36, p < .001$. Likelihood ratios indicated that NSEs were only 0.49 times as likely to fail the MSVT as ESL patients. In the valid sample, BR_{Fail} were roughly comparable between the two groups (6.8% for

NSEs and 8.0% in ESL patients), and the chi-square test failed to reach statistical significance, $\chi^2 (1, N = 1470) = 0.29, p = .591$.

Table 17

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
MSVT	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
STN	462	21.3	23.0	18.9	5.19*	1.22	19.0	38.5	52.36*	0.49

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; LR = likelihood ratio; NSE = Native speakers of English; ESL = English as a second language.
[a]$n = 2165$. [b]$n = 1292$. [c]$n = 873$. [d]$n = 1895$. [e]$n = 262$.
*$p < .05$.

Table 18

Percentages of Participants Failing the Standard Cutoff on the Medical Symptom Validity Test in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
MSVT	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
STN	101	6.8	7.0	6.6	0.09	1.06	6.8	8.0	0.29	0.85

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; LR = likelihood ratio; NSE = Native speakers of English; ESL = English as a second language.
[a]$n = 1476$. [b]$n = 841$. [c]$n = 635$. [d]$n = 1332$. [e]$n = 138$.
*$p < .05$.

NV-MSVT. In the overall sample, 18.7% of participants failed the NV-MSVT at standard cutoffs. As shown in Table 19, there was considerable overlap between NV-MSVT performance and WMT performance, $\chi^2 (1, N = 1645) = 320.78, p < .001$. There was a 7.3% NV-MSVT BR_{Fail} among those who passed the WMT, and a 44.6% NV-MSVT BR_{Fail} among those who failed the WMT. Compared to examinees who failed the WMT, examinees who passed the WMT were only 0.16 times as likely to fail the NV-MSVT.

Table 19

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
NV-MSVT	*f*	%	Pass	Fail	χ^2	LR
STN	308	18.7	7.3	44.6	320.78*	0.16

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio.
[a]$n = 1645$. [b]$n = 1141$. [c]$n = 504$.
*$p < .05$.

Tables 20 and 21 characterize NV-MSVT BR_{Fail} across clinical groups in the overall and valid samples, respectively. Although NV-MSVT BR_{Fail} varied significantly across clinical groups in the overall sample, $\chi^2\,(8, N = 1645) = 23.91, p = .002$, the effect size was small ($\Phi^2 = .015$). The highest BR_{Fail} was observed in the mTBI group (28.0%) and was closely followed by orthopedic injury patients (26.5%) and those with chronic pain/fibromyalgia (23.9%). Individuals with moderate-severe TBI had the lowest BR_{Fail} (4.9%).

In the valid sample, neither a chi-square test nor a Fisher's exact test was conducted. Patients with *other* diagnoses (9.7%), those with mTBI (9.5%), and those with chronic pain/fibromyalgia (9.4%) failed the NV-MSVT most frequently. No failures were observed among those with severe mental illness, although the group size was small ($n = 18$).

Table 20

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
NV-MSVT	*f*	%	Mild[a]	M-S[b]	NEU[c]	DEP[d]	ANX[e]	SMI[f]	CP/F[g]	ORT[h]	OTH[i]	χ^2	Φ^2
STN	308	18.7	28.0	4.9	19.8	18.9	14.3	21.9	23.9	26.5	17.4	23.91*	.015

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other.
[a]$n = 1645$. [b]$n = 107$. [c]$n = 41$. [d]$n = 86$. [e]$n = 482$. [f]$n = 483$. [g]$n = 32$. [h]$n = 259$. [i]$n = 34$. [j]$n = 121$.
*$p < .05$.

Table 21

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
NV-MSVT	*f*	%	Mild[a]	M-S[b]	NEU[c]	DEP[d]	ANX[e]	SMI[f]	CP/F[g]	ORT[h]	OTH[i]	χ^2	Φ^2
STN	83	7.3	9.5	3.0	8.7	7.1	6.0	0.0	9.4	5.3	9.7	-	-

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other.
[a]$n = 1141$. [b]$n = 63$. [c]$n = 33$. [d]$n = 69$. [e]$n = 325$. [f]$n = 350$. [g]$n = 18$. [h]$n = 171$. [i]$n = 19$. [j]$n = 93$.
*$p < .05$.

Tables 22 and 23 characterize NV-MSVT BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, NV-MSVT BR_{Fail} varied significantly across age groups, χ^2 $(4, N = 1642) = 20.79, p < .001$ (overall sample) and χ^2 $(4, N = 1141) = 23.63, p < .001$ (valid sample). However, effect sizes indicated that age accounted for only 1.3% (overall sample) to 2.1% (valid sample) of the variance in NV-MSVT BR_{Fail} (small effects).

In the overall sample, the youngest group demonstrated the lowest BR_{Fail} (9.7%), and BR_{Fail} increased with age. Almost one-third (29.6%) of individuals aged 60-69 years failed the

NV-MSVT. In the valid sample, BR_{Fail} remained stable at about 5% for participants between the ages of 16 and 49 years, with elevated BR_{Fail} in those aged 50 years and above.

Table 22

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
NV-MSVT	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
STN	308	18.7	9.7	18.5	18.7	20.3	29.6	20.79*	.013

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual.
[a]n = 1645. [b]n = 216. [c]n = 313. [d]n = 523. [e]n = 482. [f]n = 108.
*p < .05.

Table 23

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
NV-MSVT	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
STN	83	7.3	5.0	5.3	5.3	9.2	20.3	23.63*	.021

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual.
[a]n = 1141. [b]n = 179. [c]n = 207 [d]n = 358. [e]n = 327. [f]n = 69.
*p < .05.

Tables 24 and 25 show NV-MSVT BR_{Fail} as a function of education level in the overall and valid samples, respectively. Although BR_{Fail} varied significantly across education groups in the overall sample, χ^2 (5, N = 1645) = 28.17, p <.001, this was a small effect (Φ^2 = .017). The least educated groups failed the NV-MSVT most frequently (33.3% in those with ≤8 years of education and 23.4% in those with 9-11 years of education), while BR_{Fail} were lower and relatively similar (i.e., around 15%) across the remaining groups. There was no significant association between NV-MSVT failure and education level in the valid sample, χ^2 (5, N = 1140)

= 3.85, p = .571, and education level accounted for only 0.3% of the variance in NV-MSVT BR_{Fail} (very small effect).

Table 24

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
NV-MSVT	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
STN	308	18.7	33.3	23.4	15.8	14.4	18.3	14.8	28.17*	.017

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual.
[a]n = 1645. [b]n = 108. [c]n = 389. [d]n = 751. [e]n = 216. [f]n = 120. [g]n = 61.
*p < .05.

Table 25

Percentages of Participants Failing the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
NV-MSVT	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
STN	83	7.3	6.7	8.7	6.8	5.8	11.0	4.4	3.85	.003

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual.
[a]n = 1141. [b]n = 60. [c]n = 242. [d]n = 532. [e]n = 171. [f]n = 91. [g]n = 45.
*p < .05.

Tables 26 and 27 characterize NV-MSVT BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. Gender was significantly related to NV-MSVT BR_{Fail} in the overall sample, χ^2 (1, N = 1645) = 5.14, p = .023, such that males were 1.27 as likely to fail the NV-MSVT as females. No significant relationship was found between gender and NV-MSVT failure in the valid sample, χ^2 (1, N = 1141) = 0.03, p = .869 (LR = 1.04).

With regard to English language status, NSEs demonstrated significantly lower NV-MSVT BR_{Fail} than ESL patients in both samples, χ^2 (1, N = 1644) = 41.30, p <.001 (overall

sample) and χ^2 (1, N = 1140) = 5.89, p = .015 (valid sample). More specifically, NSEs were half as likely to fail the NV-MSVT than were ESL patients in either sample.

Table 26

Percentages of Participants Failing the Standard Cutoffs on the Non-Verbal Medical Symptom Validity Test in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
NV-MSVT	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
STN	308	18.7	20.5	16.1	5.14*	1.27	16.6	36.3	41.30*	0.46

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language.
[a]n = 1645. [b]n = 985. [c]n = 660. [d]n = 1462. [e]n = 182.
*p < .05.

Table 27

Percentages of Participants Failing the Standard Cutoffs on the Non-Verbal Medical Symptom Validity Test in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
NV-MSVT	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
STN	83	7.3	7.4	7.1	0.03	1.04	6.7	13.4	5.89*	0.50

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language.
[a]n = 1141. [b]n = 650. [c]n = 491. [d]n = 1043. [e]n = 228.
*p < .05.

TOMM Trial 1. Table 28 shows TOMM Trial 1 BR_{Fail}, across various cutoffs, as a function of WMT performance. Examinees who passed the WMT demonstrated significantly lower BR_{Fail} across TOMM Trial 1 cutoffs than those who failed the WMT. Not surprisingly, differences in cumulative percentages of failure were more pronounced at lower TOMM Trial 1 scores and less pronounced at higher TOMM Trial 1 scores. Individuals who passed the WMT were only 0.03 to 0.27 times as likely to score below previously published TOMM Trial 1 cutoffs (Trial 1 ≤35 to ≤45) as those who failed the WMT.

Table 28

Cumulative Percentages of Participants Failing Various TOMM Trial 1 Cutoffs as a Function of Word Memory Test Performance

TOMM	Overall Sample[a]		WMT			
Trial 1	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≤18	2	0.1	0.0	0.3	-	-
≤19	4	0.2	0.0	0.7	-	-
≤20	12	0.7	0.0	2.0	-	-
≤21	16	0.9	0.0	2.7	-	-
≤22	19	1.1	0.0	3.2	-	-
≤23	24	1.4	0.0	4.1	-	-
≤24	28	1.6	0.0	4.7	-	-
≤25	33	1.9	0.0	5.6	-	-
≤26	39	2.2	0.1	6.4	72.85*	0.02
≤27	47	2.7	0.1	7.8	89.42*	0.01
≤28	58	3.3	0.1	9.6	112.46*	0.01
≤29	71	4.0	0.3	11.5	128.20*	0.03
≤30	86	4.9	0.3	14.0	160.42*	0.02
≤31	100	5.7	0.3	16.4	191.04*	0.02
≤32	115	6.5	0.3	18.8	218.37*	0.02
≤33	125	7.1	0.4	20.3	234.86*	0.02
≤34	151	8.6	0.8	24.0	270.47*	0.03
≤35	168	9.6	0.9	26.7	304.22*	0.03
≤36	188	10.7	1.4	29.1	316.20*	0.05
≤37	207	11.8	1.6	31.8	344.34*	0.05
≤38	232	13.2	2.2	34.9	364.93*	0.06
≤39	258	14.7	3.1	37.6	372.79*	0.08
≤40	292	16.6	3.8	42.0	413.53*	0.09
≤41	333	18.9	5.3	45.9	420.37*	0.12
≤42	380	21.6	6.9	50.6	441.63*	0.14
≤43	449	25.5	9.8	56.5	449.60*	0.17
≤44	511	29.1	13.0	60.7	433.77*	0.21
≤45	605	34.4	18.0	66.8	415.12*	0.27
≤46	744	42.3	26.2	74.1	369.12*	0.35
≤47	893	50.8	36.3	79.4	291.06*	0.46
≤48	1088	61.9	49.9	85.4	210.01*	0.58
≤49	1372	78.0	69.9	93.9	131.28*	0.74
≤50	1759	100	100	100	-	-
Min	17			26	17	
Max	50			50	50	

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1759$. [b]$n = 1168$. [c]$n = 591$.
*$p < .05$.

Tables 29 and 30 characterize TOMM Trial 1 BR_{Fail} across clinical groups in the overall and valid samples, respectively. In the overall sample, BR_{Fail} varied significantly as a function of diagnosis across TOMM Trial 1 cutoffs. At previously published cutoffs (i.e., Trial 1 ≤35 through ≤45), diagnosis accounted for 2.6% to 3.8% of the variance in TOMM Trial 1 BR_{Fail} (small to medium effects): patients with mTBI and those with chronic pain/fibromyalgia tended to demonstrate the highest BR_{Fail}, while those with moderate-severe TBI, severe mental illness, orthopedic injury, and *other* diagnoses demonstrated lower BR_{Fail}.

In the valid sample, Fisher's exact tests and chi-square tests of independence failed to reach statistical significance where conducted (i.e., at cutoffs of ≤43 and higher). Effect sizes corresponding to previously identified cutoffs (Trial 1 ≤43 through ≤45) were very small (Φ^2 = .005-.006). Visual inspection of BR_{Fail} at lower cutoffs (e.g., Trial 1 ≤35 through ≤42) also did not reveal any consistent trends.

Table 29

Cumulative Percentages of Participants Failing Various Cutoffs on TOMM Trial 1 in the Overall Sample and Specific Clinical Groups

TOMM	Overall Sample[a]		TBI		Diagnostic Groups								
Trial 1	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18	2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.0	-	-
≤19	4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.4	0.0	-	-
≤20	12	0.7	0.0	0.0	0.0	0.6	0.2	0.0	1.8	0.6	0.0	-	-
≤21	16	0.9	0.0	0.0	0.0	0.9	0.4	0.0	2.2	0.6	0.0	-	-
≤22	19	1.1	0.0	0.0	0.0	0.9	0.9	0.0	2.4	0.6	0.0	-	-
≤23	24	1.4	0.0	0.0	0.0	1.2	0.9	0.0	2.9	1.1	0.6	-	-
≤24	28	1.6	0.0	0.0	0.0	1.5	1.3	0.0	3.1	1.1	0.6	-	-
≤25	33	1.9	0.0	0.0	0.0	1.5	1.3	0.0	4.0	1.1	1.2	-	-
≤26	39	2.2	0.0	0.0	0.0	2.1	1.7	0.0	4.4	1.1	1.2	-	-
≤27	47	2.7	1.6	0.0	0.0	2.9	1.9	0.0	5.1	1.1	1.2	-	-
≤28	58	3.3	3.2	0.0	0.0	3.8	2.4	7.7	5.8	1.1	1.8	-	-
≤29	71	4.0	4.8	0.0	0.0	4.7	2.8	7.7	6.9	1.7	2.4	-	-
≤30	86	4.9	4.8	0.0	2.0	5.9	3.2	7.7	8.4	1.7	3.0	FET*	.013
≤31	100	5.7	7.9	0.0	6.0	6.5	4.1	7.7	9.3	1.7	3.0	FET*	.014
≤32	115	6.5	9.5	0.0	6.0	7.1	4.7	7.7	11.3	1.7	3.0	FET*	.019
≤33	125	7.1	14.3	0.0	6.0	7.1	5.6	7.7	12.0	1.7	3.0	FET*	.021
≤34	151	8.6	19.0	3.6	8.0	8.8	6.6	7.7	13.6	2.2	4.2	39.68*	.023
≤35	168	9.6	20.6	3.6	8.0	9.1	7.5	7.7	15.8	2.8	4.2	48.10*	.027
≤36	188	10.7	22.2	3.6	8.0	10.3	8.5	7.7	17.1	4.4	4.8	46.04*	.026
≤37	207	11.8	25.4	3.6	8.0	10.9	10.0	7.7	18.2	5.6	5.4	46.95*	.027
≤38	232	13.2	25.4	3.6	10.0	13.3	10.9	7.7	20.0	6.1	7.1	44.87*	.026
≤39	258	14.7	28.6	7.1	12.0	14.5	12.6	7.7	22.2	6.1	7.1	52.04*	.030
≤40	292	16.6	30.2	10.7	14.0	16.2	14.7	7.7	25.1	6.7	7.7	57.16*	.032
≤41	333	18.9	33.3	10.7	16.0	18.9	17.1	7.7	28.4	7.8	8.3	65.55*	.037
≤42	380	21.6	36.5	14.3	18.0	21.8	19.4	15.4	31.8	11.1	8.3	67.80*	.038
≤43	449	25.5	39.7	17.9	24.0	26.3	23.1	30.8	35.6	13.3	13.1	60.87*	.035
≤44	511	29.1	41.3	17.9	26.0	31.3	26.7	30.8	39.1	14.4	17.9	59.51*	.034
≤45	605	34.4	47.6	21.4	28.0	35.7	32.7	30.8	44.9	19.4	23.8	56.94*	.032
≤46	744	42.3	57.1	21.4	32.0	44.2	40.2	53.8	52.2	29.4	31.5	53.26*	.030
≤47	893	50.8	65.1	28.6	44.0	49.6	48.9	69.2	60.4	40.6	42.3	43.43*	.025
≤48	1088	61.9	74.6	32.1	54.0	57.2	61.3	76.9	70.7	51.1	61.9	44.12*	.025
≤49	1372	78.0	83.1	64.3	74.0	72.3	79.7	84.6	82.0	74.4	78.6	18.07*	.010
≤50	1759	100	100	100	100	100	100	100	100	100	100	-	-
Min	17		27	34	30	20	20	28	17	20	23		
Max	50		50	50	50	50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 1759. [b]n = 63. [c]n = 28. [d]n = 50. [e]n = 339. [f]n = 468. [g]n = 13. [h]n = 450. [i]n = 180. [j]n = 168.
*p < .05.

Table 30

Cumulative Percentages of Participants Failing Various Cutoffs on TOMM Trial 1 in the Valid Sample and Specific Clinical Groups

TOMM	Valid Sample[a]		TBI		Diagnostic Groups								
Trial 1	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤25	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	-
≤26	1	0.1	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	-	-
≤27	-	-	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	-	-
≤28	-	-	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	-	-
≤29	3	0.3	0.0	0.0	0.0	0.5	0.3	0.0	0.0	0.0	0.8	-	-
≤30	-	-	0.0	0.0	0.0	0.5	0.3	0.0	0.0	0.0	0.8	-	-
≤31	-	-	0.0	0.0	0.0	0.5	0.3	0.0	0.0	0.0	0.8	-	-
≤32	4	0.3	0.0	0.0	0.0	0.5	0.3	0.0	0.4	0.0	0.8	-	-
≤33	5	0.4	0.0	0.0	0.0	0.5	0.6	0.0	0.4	0.0	0.8	-	-
≤34	9	0.8	0.0	0.0	0.0	0.9	1.0	0.0	1.1	0.0	0.8	-	-
≤35	10	0.9	0.0	0.0	0.0	0.9	1.3	0.0	1.1	0.0	0.8	-	-
≤36	16	1.4	0.0	0.0	0.0	1.4	1.6	0.0	1.1	1.9	1.6	-	-
≤37	19	1.6	0.0	0.0	0.0	1.4	1.9	0.0	1.5	1.9	2.4	-	-
≤38	26	2.2	0.0	0.0	0.0	2.3	2.3	0.0	2.7	1.9	3.2	-	-
≤39	36	3.1	0.0	4.5	2.5	2.8	3.2	0.0	4.2	1.9	3.2	-	-
≤40	44	3.8	0.0	4.5	5.0	3.2	4.2	0.0	5.4	1.9	3.2	-	-
≤41	62	5.3	3.2	4.5	5.0	5.1	6.1	0.0	8.0	1.9	3.2	-	-
≤42	81	6.9	3.2	9.1	5.0	6.9	7.4	0.0	10.3	4.5	3.2	-	-
≤43	115	9.8	6.5	13.6	12.5	9.7	9.7	12.5	12.6	7.1	7.1	FET	.005
≤44	152	13.0	6.5	13.6	15.0	14.8	12.3	12.5	16.1	8.4	11.9	7.27	.006
≤45	210	18.0	16.1	18.2	17.5	18.1	17.7	12.5	22.6	14.3	14.3	6.63	.006
≤46	306	26.2	35.5	18.2	20.0	30.1	24.5	25.0	30.7	24.0	18.3	12.23	.010
≤47	424	36.3	48.4	22.7	35.0	35.6	36.1	50.0	40.6	36.4	27.8	10.49	.009
≤48	583	49.9	61.3	27.3	47.5	44.4	49.7	62.5	55.9	46.8	52.4	14.02	.012
≤49	817	69.9	74.2	59.1	67.5	62.5	71.9	75.0	71.6	72.7	72.2	9.23	.008
≤50	1169	100	100	100	100	100	100	100	100	100	100	-	-
Min	26		41	39	39	26	29	43	32	36	29		
Max	50		50	50	50	50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1168$. [b]$n = 31$. [c]$n = 22$. [d]$n = 40$. [e]$n = 216$. [f]$n = 310$. [g]$n = 8$. [h]$n = 261$. [i]$n = 154$. [j]$n = 126$.
*$p < .05$.

Tables 31 and 32 characterize TOMM Trial 1 BR_{Fail} as a function of age in the overall and valid samples, respectively. In the overall sample, Fisher's exact tests and chi-square tests of independence were statistically significant at some values and non-significant at others. However, effect sizes were consistently very small and ranged from $\Phi^2 = .004$ to .008 at previously published cutoffs (Trial 1 ≤35 through ≤45). The youngest group (aged 16-29 years) tended to demonstrate the lowest BR_{Fail}, and BR_{Fail} generally increased with older age. Notably, individuals aged 60-69 years demonstrated somewhat lower BR_{Fail} than would be expected based on the trend at several cutoffs.

In the valid sample, Fisher's exact tests and chi-square tests of independence were largely non-significant across TOMM Trial 1 cutoffs, and effect sizes remained very small. Age accounted for 0.2% to 0.9% of the variance in TOMM Trial 1 BR_{Fail} at previously published cutoffs.

Table 31

Cumulative Percentages of Participants Failing Various TOMM Trial 1 cutoffs in the Overall Sample and across Several Age Groups

TOMM	Overall Sample[a]		Age (years)						
Trial 1	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18	2	0.1	0.4	0.0	0.0	0.3	0.0	-	-
≤19	4	0.2	0.4	0.0	0.2	0.5	0.0	-	-
≤20	12	0.7	0.7	0.2	0.7	1.0	1.4	FET	.001
≤21	16	0.9	0.7	0.2	1.1	1.5	1.4	FET	.003
≤22	19	1.1	0.7	0.2	1.2	2.0	1.4	FET	.004
≤23	24	1.4	0.7	0.7	1.6	2.3	1.4	FET	.003
≤24	28	1.6	0.7	0.9	1.6	3.0	1.4	FET	.004
≤25	33	1.9	0.7	1.6	1.8	3.3	1.4	6.48	.004
≤26	39	2.2	0.7	1.8	2.3	3.8	1.4	7.82	.004
≤27	47	2.7	0.7	2.2	2.5	4.8	2.7	11.10*	.006
≤28	58	3.3	1.5	2.7	3.4	5.3	2.7	8.35	.005
≤29	71	4.0	1.5	3.8	3.9	6.3	4.1	9.92*	.006
≤30	86	4.9	1.8	4.7	4.6	7.5	4.1	11.83*	.007
≤31	100	5.7	2.9	5.6	5.1	8.3	5.5	9.31	.005
≤32	115	6.5	2.9	6.0	5.8	10.6	5.5	17.18*	.010
≤33	125	7.1	4.0	6.5	6.2	11.1	5.5	14.75*	.008
≤34	151	8.6	5.9	7.6	7.8	12.6	6.8	11.99*	.007
≤35	168	9.6	6.2	8.3	9.2	13.3	9.6	11.02*	.006
≤36	188	10.7	7.7	9.2	10.8	14.1	9.6	8.56	.005
≤37	207	11.8	8.1	11.2	11.7	15.1	9.6	8.34	.005
≤38	232	13.2	8.4	12.3	13.6	17.1	9.6	11.97*	.007
≤39	258	14.7	9.9	14.1	14.9	18.8	9.6	12.22*	.007
≤40	292	16.6	10.3	16.3	17.5	20.6	11.0	14.64*	.008
≤41	333	18.9	12.1	18.8	20.4	22.4	13.7	13.47*	.008
≤42	380	21.6	14.3	22.4	23.2	24.1	16.4	12.29*	.007
≤43	449	25.5	18.3	25.1	28.7	27.4	19.2	12.75*	.007
≤44	511	29.1	20.9	28.9	32.2	30.9	24.7	12.95*	.007
≤45	605	34.4	27.8	34.0	36.8	35.9	32.9	7.19	.004
≤46	744	42.3	37.7	42.1	43.9	42.7	43.8	3.03	.002
≤47	893	50.8	46.5	51.0	52.7	49.7	53.4	3.23	.002
≤48	1088	61.9	54.9	62.6	62.8	63.8	63.0	6.55	.004
≤49	1372	78.0	75.1	80.8	77.9	76.9	78.1	3.62	.002
≤50	1759	100	100	100	100	100	100		
Min	17		18	20	19	17	20		
Max	50		50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1759$. [b]$n = 273$. [c]$n = 447$. [d]$n = 565$. [e]$n = 398$. [f]$n = 73$.
*$p < .05$.

Table 32

Cumulative Percentages of Participants Failing Various TOMM Trial 1 cutoffs in the Valid Sample and across Several Age Groups

TOMM	Valid Sample[a]		Age (years)						
Trial 1	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤25	0	0.0	0.0	0.0	0.0	0.0	0.0	-	-
≤26	1	0.1	0.0	0.0	0.0	0.4	0.0	-	-
≤27	-	-	0.0	0.0	0.0	0.4	0.0	-	-
≤28	-	-	0.0	0.0	0.0	0.4	0.0	-	-
≤29	3	0.3	0.0	0.3	0.3	0.4	0.0	-	-
≤30	-	-	0.0	0.3	0.3	0.4	0.0	-	-
≤31	-	-	0.0	0.3	0.3	0.4	0.0	-	-
≤32	4	0.3	0.0	0.3	0.3	0.9	0.0	-	-
≤33	5	0.4	0.0	0.3	0.3	1.3	0.0	-	-
≤34	9	0.8	0.0	0.7	0.8	1.7	0.0	-	-
≤35	10	0.9	0.0	0.7	0.8	1.7	2.4	-	-
≤36	16	1.4	0.5	0.7	2.1	1.7	2.4	FET	.004
≤37	19	1.6	0.5	1.3	2.3	1.7	2.4	FET	.003
≤38	26	2.2	1.0	1.7	3.1	2.6	2.4	FET	.003
≤39	36	3.1	1.9	2.6	4.1	3.0	2.4	2.62	.002
≤40	44	3.8	1.9	4.0	4.9	3.0	4.8	3.82	.003
≤41	62	5.3	2.9	4.6	7.8	4.3	4.8	7.77	.007
≤42	81	6.9	3.9	7.0	9.6	5.6	4.8	8.12	.007
≤43	115	9.8	6.3	9.9	13.5	7.3	7.1	10.62*	.009
≤44	152	13.0	9.7	13.6	16.6	9.9	9.5	8.83	.008
≤45	210	18.0	17.5	18.2	21.0	13.4	16.7	5.81	.005
≤46	306	26.2	26.7	26.2	28.0	22.8	26.2	2.01	.002
≤47	424	36.3	37.4	36.1	38.6	31.0	40.5	4.09	.003
≤48	583	49.9	47.1	51.0	50.8	49.1	52.4	1.07	.001
≤49	817	69.9	70.4	73.5	69.9	65.5	66.7	4.22	.004
≤50	1169	100	100	100	100	100	100	-	-
Min	26		36	29	29	26	35		
Max	50		50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1168$. [b]$n = 206$. [c]$n = 302$. [d]$n = 386$. [e]$n = 232$. [f]$n = 42$.
*$p < .05$.

Tables 33 and 34 characterize TOMM Trial 1 BR_{Fail} across education levels in the overall and valid samples, respectively. Although chi-square tests reached statistical significance at higher TOMM Trial 1 cutoffs in the overall sample, this pattern of findings was likely driven by the large sample size. Effect sizes indicate that education accounted for only 0.1% to 1.3% of the variance in TOMM Trial 1 BR_{Fail} across all cutoffs examined, and for 0.5% to 1.3% of the variance in BR_{Fail} at previously published cutoffs (Trial 1 ≤35 through ≤45; very small to small effect). Nonetheless, BR_{Fail} were highest among the group with ≤8 years of education and generally decreased with higher levels of education. Examinees with 16 and ≥17 years of education did, however, demonstrate higher BR_{Fail} than would be expected based on the trend at higher TOMM Trial 1 cutoffs.

In the valid sample, Fisher's exact tests and chi-square tests of independence failed to reach statistical significance where conducted (i.e., at cutoffs of Trial 1 ≤36 and above), and effect sizes remained small. Education level accounted for 0.4% to 1.0% of the variance in Trial 1 BR_{Fail} at previously published cutoffs.

Table 33

Cumulative Percentages of Participants Failing Various TOMM Trial 1 Cutoffs in the Overall Sample and across Several Education Levels

TOMM	Overall Sample[a]		Education (years)							
Trial 1	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18	2	0.1	0.0	0.2	0.0	0.0	0.0	1.4	-	-
≤19	4	0.2	0.0	0.2	0.2	0.4	0.0	1.4	-	-
≤20	12	0.7	1.4	0.6	0.6	0.7	0.0	1.4	-	-
≤21	16	0.9	3.4	0.8	0.6	0.7	0.0	1.4	-	-
≤22	19	1.1	3.4	1.2	0.8	0.7	0.0	1.4	-	-
≤23	24	1.4	3.4	1.2	1.2	1.1	0.9	1.4	FET	.003
≤24	28	1.6	3.4	1.8	1.2	1.5	0.9	1.4	FET	.002
≤25	33	1.9	4.1	2.4	1.4	1.5	0.9	1.4	FET	.004
≤26	39	2.2	4.1	2.6	1.8	2.2	0.9	1.4	FET	.002
≤27	47	2.7	4.1	2.9	2.4	2.2	1.8	2.8	FET	.001
≤28	58	3.3	4.8	3.7	2.9	2.6	3.7	2.8	FET	.001
≤29	71	4.0	6.1	5.3	3.4	2.6	3.7	2.8	6.25	.004
≤30	86	4.9	7.5	5.9	4.3	3.7	3.7	4.2	4.94	.003
≤31	100	5.7	8.8	6.7	5.3	3.7	3.7	5.6	6.52	.004
≤32	115	6.5	8.8	8.1	6.3	3.7	4.6	6.9	7.39	.004
≤33	125	7.1	10.2	8.3	7.0	4.1	5.5	6.9	7.19	.004
≤34	151	8.6	12.9	9.8	8.2	6.0	6.4	6.9	7.80	.004
≤35	168	9.6	14.3	11.0	9.0	6.7	8.3	6.9	8.50	.005
≤36	188	10.7	16.3	11.8	10.5	7.1	10.1	6.9	10.22	.006
≤37	207	11.8	18.4	13.0	11.5	7.9	11.0	8.3	11.73*	.007
≤38	232	13.2	20.4	14.5	12.8	8.6	11.9	11.1	12.88*	.007
≤39	258	14.7	21.1	17.1	13.9	9.0	14.7	12.5	14.70*	.008
≤40	292	16.6	23.1	19.8	15.0	11.2	15.6	16.7	15.29*	.009
≤41	333	18.9	25.9	23.6	16.3	12.0	18.3	22.2	23.54*	.013
≤42	380	21.6	28.6	25.1	19.1	15.7	23.9	23.6	16.38*	.009
≤43	449	25.5	33.3	30.5	22.1	18.4	28.4	27.8	23.07*	.013
≤44	511	29.1	35.4	34.4	26.4	20.6	31.2	30.6	21.66*	.012
≤45	605	34.4	40.1	39.9	32.4	25.8	33.9	34.7	18.80*	.011
≤46	744	42.3	49.0	48.5	40.2	33.7	39.4	40.3	20.57*	.012
≤47	893	50.8	59.2	55.4	49.3	42.7	47.7	48.6	16.60*	.009
≤48	1088	61.9	66.7	66.2	61.7	53.6	55.0	63.9	15.60*	.009
≤49	1372	78.0	80.3	80.9	78.5	71.9	71.6	80.6	11.78*	.007
≤50	1759	100	100	100	100	100	100	100	-	-
Min	17		20	17	19	19	23	18		
Max	50		50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1759$. [b]$n = 147$. [c]$n = 509$. [d]$n = 655$. [e]$n = 267$. [f]$n = 109$. [g]$n = 72$.
*$p < .05$.

Table 34

Cumulative Percentages of Participants Failing Various TOMM Trial 1 Cutoffs in the Valid Sample and across Several Education Levels

TOMM	Valid Sample[a]		Education (years)							
Trial 1	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤25	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	-
≤26	1	0.1	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤27	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤28	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤29	3	0.3	1.2	0.3	0.2	0.0	0.0	0.0	-	-
≤30	-	-	1.2	0.3	0.2	0.0	0.0	0.0	-	-
≤31	-	-	1.2	0.3	0.2	0.0	0.0	0.0	-	-
≤32	4	0.3	1.2	0.3	0.2	0.0	0.0	1.9	-	-
≤33	5	0.4	2.4	0.3	0.2	0.0	0.0	1.9	-	-
≤34	9	0.8	2.4	0.9	0.5	0.5	0.0	1.9	-	-
≤35	10	0.9	2.4	1.2	0.5	0.5	0.0	1.9	-	-
≤36	16	1.4	4.9	1.2	1.1	0.5	1.4	1.9	FET	.008
≤37	19	1.6	6.1	1.2	1.6	0.5	1.4	1.9	FET	.010
≤38	26	2.2	6.1	2.5	2.1	1.0	1.4	1.9	FET	.006
≤39	36	3.1	7.3	4.0	2.5	1.5	2.7	1.9	FET	.007
≤40	44	3.8	7.3	5.0	3.0	2.0	2.7	5.7	FET	.006
≤41	62	5.3	8.5	7.1	4.3	2.5	5.4	7.5	FET	.007
≤42	81	6.9	8.5	8.0	6.4	4.0	9.5	9.4	5.00	.004
≤43	115	9.8	12.2	12.4	8.9	5.0	12.2	13.2	9.60	.008
≤44	152	13.0	13.4	15.8	13.0	7.0	13.5	17.0	9.25	.008
≤45	210	18.0	17.1	20.4	18.5	13.1	16.2	20.8	5.15	.004
≤46	306	26.2	26.8	29.1	26.5	21.1	24.3	26.4	4.26	.004
≤47	424	36.3	41.5	39.0	36.6	31.2	29.7	37.7	5.70	.005
≤48	583	49.9	51.2	53.9	51.3	42.7	40.5	52.8	9.30	.008
≤49	817	69.9	68.3	74.0	71.2	64.3	59.5	73.6	10.13	.009
≤50	1169	100	100	100	100	100	100	100		
Min	26		29	29	26	34	36	32		
Max	50		50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1168$. [b]$n = 82$. [c]$n = 323$. [d]$n = 437$. [e]$n = 199$. [f]$n = 74$. [g]$n = 53$.
*$p < .05$.

Tables 35 and 36 characterize TOMM Trial 1 BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. BR_{Fail} were not related to gender at most TOMM Trial 1 cutoffs in either sample. Although some chi-square tests were statistically significant at higher cutoffs in both samples, these findings were likely related to the large sample size.

With regard to English language status, NSEs demonstrated lower BR_{Fail} than ESL patients across TOMM Trial 1 cutoffs in both samples. At previously published cutoffs, likelihood ratios ranged from 0.29 (Trial 1 ≤35) to 0.58 (Trial 1 ≤44) in the overall sample, and from 0.18 (Trial 1 ≤37) to 0.55 (Trial 1 ≤43) in the valid sample.

Table 35

Cumulative Percentages of Participants Failing Various TOMM Trial 1 Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

TOMM	Overall Sample[a]		Gender				English			
Trial 1	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤18	2	0.1	0.1	0.1	FET	1.00	0.0	0.8	-	-
≤19	4	0.2	0.1	0.4	FET	0.25	0.0	1.6	-	-
≤20	12	0.7	0.4	1.1	FET	0.36	0.3	3.3	FET*	0.09
≤21	16	0.9	0.8	1.1	0.54	0.73	0.5	3.3	FET*	0.15
≤22	19	1.1	1.0	1.2	0.32	0.83	0.7	3.7	FET*	0.19
≤23	24	1.4	1.3	1.5	0.24	0.87	0.8	4.9	FET*	0.16
≤24	28	1.6	1.4	1.8	0.35	0.78	0.9	5.3	FET*	0.17
≤25	33	1.9	1.8	1.9	0.29	0.95	1.1	6.5	FET*	0.17
≤26	39	2.2	2.3	2.1	0.11	1.10	1.3	7.3	36.02*	0.18
≤27	47	2.7	2.5	2.9	0.27	0.86	1.6	9.0	45.02*	0.18
≤28	58	3.3	3.1	3.6	0.37	0.86	1.9	11.4	60.53*	0.17
≤29	71	4.0	3.9	4.2	0.05	0.93	2.5	13.5	66.71*	0.19
≤30	86	4.9	5.0	4.7	0.08	1.06	3.1	15.5	70.17*	0.20
≤31	100	5.7	6.1	5.1	0.70	1.20	3.5	18.8	92.12*	0.19
≤32	115	6.5	7.0	5.8	1.02	1.21	4.4	19.6	80.25*	0.22
≤33	125	7.1	7.7	6.2	1.39	1.24	4.8	20.8	81.83*	0.23
≤34	151	8.6	9.3	7.5	1.87	1.24	6.2	22.9	74.44*	0.27
≤35	168	9.6	10.4	8.3	2.14	1.25	7.1	24.5	73.99*	0.29
≤36	188	10.7	11.7	9.3	2.49	1.26	8.2	25.7	67.67*	0.32
≤37	207	11.8	12.9	10.1	3.18	1.28	9.2	28.2	73.97*	0.33
≤38	232	13.2	14.6	11.1	4.68	1.32	10.4	30.2	72.17*	0.34
≤39	258	14.7	15.8	13.0	2.59	1.22	11.9	31.8	67.16*	0.37
≤40	292	16.6	17.9	14.7	3.18	1.22	13.5	35.5	73.56*	0.38
≤41	333	18.9	20.1	17.2	2.39	1.17	15.9	37.6	64.26*	0.42
≤42	380	21.6	22.8	19.8	2.26	1.15	18.6	40.0	56.78*	0.47
≤43	449	25.5	27.2	23.2	3.59	1.17	22.5	44.1	51.77*	0.51
≤44	511	29.1	30.6	26.8	3.09	1.14	25.5	50.2	62.34*	0.58
≤45	605	34.4	36.4	31.5	4.59*	1.16	30.6	57.1	65.67*	0.54
≤46	744	42.3	44.9	38.6	7.00*	1.16	38.5	64.9	60.08*	0.59
≤47	893	50.8	53.8	46.5	9.06*	1.16	47.3	71.4	48.90*	0.66
≤48	1088	61.9	64.5	58.0	7.79*	1.11	58.8	80.0	40.04*	0.74
≤49	1372	78.0	80.9	73.8	12.63*	1.10	75.9	90.2	24.97*	0.84
≤50	1759	100	100	100	-	-	100	100	-	-
Min	17		18	17			20	17		
Max	50		50	50			50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1759$. [b]$n = 1038$. [c]$n = 721$. [d]$n = 1508$. [e]$n = 245$.
*$p < .05$.

Table 36

Cumulative Percentages of Participants Failing Various TOMM Trial 1 Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

TOMM	Valid Sample[a]		Gender				English			
Trial 1	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤25	0	0.0	0.0	0.0	-	-	0.0	0.0	-	-
≤26	1	0.1	0.0	0.2	-	-	0.0	0.9	-	-
≤27	-	-	0.0	0.2	-	-	0.0	0.9	-	-
≤28	-	-	0.0	0.2	-	-	0.0	0.9	-	-
≤29	3	0.3	0.3	0.2	FET	1.50	0.1	1.7	FET*	0.06
≤30	-	-	0.3	0.2	FET	1.50	0.1	1.7	FET*	0.06
≤31	-	-	0.3	0.2	FET	1.50	0.1	1.7	FET*	0.06
≤32	4	0.3	0.3	0.4	FET	0.75	0.2	1.7	FET	0.12
≤33	5	0.4	0.5	0.4	FET	1.25	0.3	1.7	FET	0.18
≤34	9	0.8	0.6	1.0	FET	0.60	0.6	2.6	FET	0.23
≤35	10	0.9	0.8	1.0	FET	0.80	0.7	2.6	FET	0.27
≤36	16	1.4	1.1	1.8	1.10	0.61	1.0	4.3	FET*	0.23
≤37	19	1.6	1.4	2.0	0.68	0.70	1.1	6.1	FET*	0.18
≤38	26	2.2	2.1	2.4	0.09	0.88	1.7	7.0	FET*	0.24
≤39	36	3.1	2.6	3.8	1.35	0.68	2.5	8.7	FET*	0.29
≤40	44	3.8	3.6	4.0	0.09	0.90	3.0	10.4	FET*	0.29
≤41	62	5.3	4.7	6.1	1.19	0.77	4.6	12.2	11.89*	0.38
≤42	81	6.9	6.3	7.7	0.83	0.82	6.2	13.9	9.56*	0.45
≤43	115	9.8	9.8	9.9	0.00	0.99	9.0	16.5	6.56*	0.55
≤44	152	13.0	13.3	12.6	0.12	1.06	11.8	23.5	12.51*	0.50
≤45	210	18.0	18.6	17.2	0.37	1.08	16.3	33.0	19.77*	0.49
≤46	306	26.2	28.1	23.7	2.85	1.19	24.5	40.9	14.44*	0.60
≤47	424	36.3	39.1	32.6	5.27*	1.20	34.8	49.6	9.83*	0.70
≤48	583	49.9	52.3	46.8	3.38	1.12	48.3	64.3	10.70*	0.75
≤49	817	69.9	73.3	65.6	7.99*	1.12	68.7	80.9	7.33*	0.85
≤50	1169	100	100	100	-	-	100	100	-	-
Min	26		29	26			29	26		
Max	50		50	50			50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TOMM Trial 1 score in the respective group; Max = maximum TOMM Trial 1 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1168$. [b]$n = 662$. [c]$n = 506$. [d]$n = 1050$. [e]$n = 115$.
*$p < .05$.

TOMM Trial 2. Table 37 shows TOMM Trial 2 BR_{Fail}, across various cutoffs, as a function of WMT performance. Participants who passed the WMT demonstrated significantly lower BR_{Fail} across TOMM Trial 2 cutoffs than did those who failed the WMT. Those who passed the WMT were only 0.04 to 0.21 times as likely as those who failed the WMT to fail previously published Trial 2 cutoffs (Trial 2 ≤44 through ≤49).

Table 37

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

TOMM	Overall Sample[a]		WMT			
Trial 2	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≤18	7	0.4	0.1	1.0	FET*	0.10
≤19	-	-	0.1	1.0	FET*	0.10
≤20	8	0.5	0.1	1.2	FET*	0.08
≤21	10	0.6	0.1	1.5	FET*	0.07
≤22	14	0.8	0.1	2.2	FET*	0.05
≤23	17	1.0	0.1	2.7	28.18*	0.04
≤24	22	1.3	0.1	3.6	38.21*	0.03
≤25	28	1.6	0.1	4.6	50.35*	0.02
≤26	30	1.7	0.1	4.9	54.42*	0.02
≤27	34	1.9	0.1	5.6	62.58*	0.02
≤28	36	2.1	0.1	5.9	66.69*	0.02
≤29	43	2.4	0.1	7.1	81.12*	0.01
≤30	46	2.6	0.1	7.6	87.34*	0.01
≤31	54	3.1	0.1	9.0	104.05*	0.01
≤32	58	3.3	0.1	9.7	112.46*	0.01
≤33	60	3.4	0.1	10.0	116.69*	0.01
≤34	67	3.8	0.2	11.0	125.56*	0.02
≤35	73	4.2	0.2	12.0	138.36*	0.02
≤36	81	4.6	0.3	13.2	149.62*	0.02
≤37	90	5.1	0.3	14.6	163.22*	0.02
≤38	100	5.7	0.4	16.1	179.19*	0.02
≤39	118	6.7	0.5	19.0	213.18*	0.03
≤40	128	7.3	0.7	20.3	223.91*	0.03
≤41	144	8.2	0.9	22.7	248.55*	0.04
≤42	152	8.7	0.9	24.1	266.92*	0.04
≤43	167	9.5	0.9	26.4	295.95*	0.03
≤44	185	10.5	1.2	29.0	320.82*	0.04
≤45	199	11.3	1.5	30.8	336.76*	0.05
≤46	218	12.4	1.6	33.7	371.24*	0.05
≤47	250	14.2	2.4	37.6	398.14*	0.06
≤48	301	17.1	3.8	43.6	436.62*	0.09
≤49	448	25.5	11.3	53.6	367.80*	0.21
≤50	1756	100	100	100	-	-
Min	10		14	10		
Max	50		50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; FET = Fisher's exact test; LR = likelihood ratio; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1756$. [b]$n = 1166$. [c]$n = 590$.
*$p < .05$.

Tables 38 and 39 characterize TOMM Trial 2 BR_{Fail} across clinical groups in the overall and valid samples, respectively. In the overall sample, chi-square tests of independence were statistically significant where conducted (i.e., at Trial 2 cutoffs of ≤44 and higher). Effect sizes indicated that diagnosis accounted for 1.6 to 3.5% of the variance in BR_{Fail} at these cutoffs (small to medium effects). Patients with chronic pain/fibromyalgia and mTBI had the highest BR_{Fail}, and those with orthopedic injuries and *other* diagnoses had the lowest BR_{Fail}. Although BR_{Fail} were also quite low among those with moderate-severe TBI, severe mental illness, and neurological conditions, these groups were smaller in size.

In the valid sample, neither Fisher's exact tests nor chi-square tests of independence were conducted. However, as shown in Table 39, most participants achieved scores of 49 or 50 on TOMM Trial 2, across clinical groups. At Trial 2 ≤49, BR_{Fail} were highest among those with *other* diagnoses (15.9%) and lowest in those with orthopedic injury (9.7%), mTBI (9.7%), and anxiety (9.0%). No failures were observed in those with severe mental illness, although the group was small in size (n = 8).

Table 38

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Overall Sample and in Specific Clinical Groups

TOMM	Overall Sample[a]		TBI		Diagnostic Groups								
Trial 2	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18	7	0.4	0.0	0.0	0.0	0.3	0.0	0.0	1.1	0.0	0.6	-	-
≤19	-	-	0.0	0.0	0.0	0.3	0.0	0.0	1.1	0.0	0.6	-	-
≤20	8	0.5	0.0	0.0	0.0	0.3	0.0	0.0	1.3	0.0	0.6	-	-
≤21	10	0.6	0.0	0.0	0.0	0.9	0.0	0.0	1.3	0.0	0.6	-	-
≤22	14	0.8	0.0	0.0	0.0	0.9	0.0	0.0	2.2	0.0	0.6	-	-
≤23	17	1.0	0.0	0.0	0.0	1.2	0.0	0.0	2.4	0.6	0.6	-	-
≤24	22	1.3	0.0	0.0	0.0	1.2	0.4	0.0	2.9	1.1	0.6	-	-
≤25	28	1.6	0.0	0.0	0.0	1.5	0.6	0.0	3.8	1.1	0.6	-	-
≤26	30	1.7	0.0	0.0	0.0	1.8	0.9	0.0	3.8	1.1	0.6	-	-
≤27	34	1.9	0.0	0.0	0.0	1.8	1.1	0.0	4.4	1.1	1.2	-	-
≤28	36	2.1	0.0	0.0	0.0	1.8	1.1	0.0	4.7	1.1	1.2	-	-
≤29	43	2.4	1.6	0.0	0.0	2.1	1.5	0.0	5.3	1.1	1.2	-	-
≤30	46	2.6	1.6	0.0	0.0	2.4	1.9	0.0	5.8	1.1	1.2	-	-
≤31	54	3.1	3.2	0.0	0.0	3.6	1.9	0.0	6.2	1.7	1.2	-	-
≤32	58	3.3	3.2	0.0	0.0	3.9	1.9	0.0	6.2	1.7	1.8	-	-
≤33	60	3.4	3.2	0.0	0.0	3.9	2.1	0.0	6.4	1.7	1.8	-	-
≤34	67	3.8	4.8	0.0	0.0	4.2	2.4	0.0	7.3	1.7	1.8	-	-
≤35	73	4.2	4.8	0.0	0.0	4.5	2.8	0.0	8.0	1.7	1.8	-	-
≤36	81	4.6	4.8	0.0	0.0	4.7	3.9	0.0	8.4	1.7	1.8	-	-
≤37	90	5.1	6.3	0.0	0.0	5.0	4.1	0.0	9.3	1.7	3.0	-	-
≤38	100	5.7	6.3	0.0	2.0	5.9	4.7	0.0	10.0	1.7	3.0	-	-
≤39	118	6.7	12.7	0.0	2.0	7.1	5.1	0.0	11.8	1.7	3.0	-	-
≤40	128	7.3	12.7	0.0	2.0	7.4	5.8	0.0	12.9	2.2	3.0	-	-
≤41	144	8.2	14.3	0.0	4.0	8.0	6.9	0.0	14.0	2.2	4.2	-	-
≤42	152	8.7	14.3	0.0	4.0	8.3	7.3	0.0	15.1	2.2	4.2	-	-
≤43	167	9.5	15.9	3.6	6.0	9.2	8.4	0.0	15.8	2.8	4.2	-	-
≤44	185	10.5	17.5	3.6	6.0	9.5	9.9	7.7	17.1	3.3	4.8	42.96*	.024
≤45	199	11.3	17.5	3.6	6.0	11.0	10.3	7.7	18.2	3.9	5.4	43.33*	.025
≤46	218	12.4	20.6	7.1	6.0	11.9	11.8	7.7	19.1	4.4	6.0	42.58*	.024
≤47	250	14.2	27.0	7.1	10.0	13.6	12.6	7.7	21.8	5.0	7.7	51.15*	.029
≤48	301	17.1	28.6	7.1	10.0	19.0	14.6	7.7	26.0	6.1	8.9	61.64*	.035
≤49	448	25.5	33.3	21.7	20.0	27.6	23.1	7.7	33.8	13.3	19.6	40.69*	.016
≤50	1756	100	100	100	100	100	100	100	100	100	100	-	-
Min	10		29	43	38	16	24	44	10	23	14		
Max	50		50	50	50	50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.

[a]$n = 1756$. [b]$n = 63$. [c]$n = 28$. [d]$n = 50$. [e]$n = 337$. [f]$n = 467$. [g]$n = 13$. [h]$n = 450$. [i]$n = 180$. [j]$n = 168$.

*$p < .05$.

Table 39

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Valid Sample and in Specific Clinical Groups

TOMM	Valid Sample[a]		TBI		Diagnostic Groups								
Trial 2	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18	1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤19	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤20	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤21	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤22	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤23	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤24	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤25	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤26	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤27	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤28	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤29	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤30	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤31	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤32	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤33	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≤34	2	0.2	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.8	-	-
≤35	-	-	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.8	-	-
≤36	3	0.3	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.8	-	-
≤37	4	0.3	0.0	0.0	0.0	0.5	0.3	0.0	0.0	0.0	1.6	-	-
≤38	5	0.4	0.0	0.0	0.0	0.9	0.3	0.0	0.0	0.0	1.6	-	-
≤39	6	0.5	0.0	0.0	0.0	1.4	0.3	0.0	0.0	0.0	1.6	-	-
≤40	8	0.7	0.0	0.0	0.0	1.4	0.3	0.0	0.8	0.0	1.6	-	-
≤41	10	0.9	0.0	0.0	0.0	1.4	0.6	0.0	1.1	0.0	1.6	-	-
≤42	-	-	0.0	0.0	0.0	1.4	0.6	0.0	1.1	0.0	1.6	-	-
≤43	11	0.9	0.0	0.0	0.0	1.4	1.0	0.0	1.1	0.0	1.6	-	-
≤44	14	1.2	0.0	0.0	0.0	1.4	1.9	0.0	1.1	0.0	1.6	-	-
≤45	17	1.5	0.0	0.0	0.0	1.9	2.3	0.0	1.5	0.0	1.6	-	-
≤46	19	1.6	0.0	0.0	0.0	1.9	2.3	0.0	1.9	0.6	1.6	-	-
≤47	28	2.4	6.5	0.0	0.0	2.3	2.3	0.0	2.7	1.3	4.0	-	-
≤48	44	3.8	6.5	0.0	0.0	5.1	3.2	0.0	4.2	1.9	5.6	-	-
≤49	132	11.3	9.7	13.6	10.0	12.6	9.0	0.0	12.3	9.7	15.9	-	-
≤50	1166	100	100	100	100	100	100	100	100	100	100	-	-
Min	14		47	49	49	34	36	50	40	46	14		
Max	40		50	50	50	50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1166$. [b]$n = 31$. [c]$n = 22$. [d]$n = 40$. [e]$n = 214$. [f]$n = 310$. [g]$n = 8$. [h]$n = 261$. [i]$n = 154$. [j]$n = 126$.
*$p < .05$.

Tables 40 and 41 characterize TOMM Trial 2 BR_{Fail} as a function of age in the overall and valid samples, respectively. Although chi-square tests of independence were significant at higher cutoffs in the overall sample, this pattern of results was likely driven by the large sample size. Effect sizes were consistently small (Φ^2 = .003 - .010) across cutoffs examined, and age accounted for only 0.8 to 1.0% of the variance in TOMM Trial 2 BR_{Fail} at previously published cutoffs (≤44 through ≤49). BR_{Fail} tended to be lowest in the youngest group and increase with older age across TOMM Trial 2 cutoffs. A spike in BR_{Fail} was, however, observed in individuals 50-59 years of age.

In the valid sample, most participants achieved scores of 49 or 50 on TOMM Trial 2 across age groups. Chi-square tests of independence and Fisher's exact tests were conducted at previously published cutoffs (i.e., ≤44 and higher). Age accounted for only 0.1 to 1.6% of the variance in Trial 2 BR_{Fail} at these levels of performance (very small to small effects). No consistent trends in BR_{Fail} were observed across age groups at cutoffs of Trial 2 ≤48 and lower. At Trial 2 ≤49, BR_{Fail} were stable at about 10% for examinees between aged 16-59 years, and an elevated BR_{Fail} (31.0%) was observed in the oldest group (60-69 years).

Table 40

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Overall Sample and across Several Age Groups

TOMM	Overall Sample[a]		Age (years)						
Trial 2	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18	7	0.4	0.4	0.2	0.2	1.0	0.0	-	-
≤19	-	-	0.4	0.2	0.2	1.0	0.0	-	-
≤20	8	0.5	0.7	0.2	0.2	1.0	0.0	-	-
≤21	10	0.6	0.7	0.4	0.4	1.0	0.0	-	-
≤22	14	0.8	0.7	0.9	0.5	1.3	0.0	-	-
≤23	17	1.0	0.7	0.9	1.1	1.3	0.0	-	-
≤24	22	1.3	0.7	1.1	1.4	1.8	0.0	-	-
≤25	28	1.6	0.7	1.6	1.6	2.5	0.0	-	-
≤26	30	1.7	0.7	1.6	1.6	3.0	0.0	-	-
≤27	34	1.9	0.7	1.8	1.9	3.3	0.0	-	-
≤28	36	2.1	0.7	1.8	2.1	3.5	0.0	-	-
≤29	43	2.4	1.1	2.5	2.3	3.8	1.4	5.42	.003
≤30	46	2.6	1.1	2.7	2.5	4.0	1.4	6.06	.003
≤31	54	3.1	1.5	3.4	2.7	4.8	1.4	7.44	.004
≤32	58	3.3	1.8	3.6	2.8	5.0	1.4	6.95	.004
≤33	60	3.4	1.8	3.6	3.2	5.5	1.4	9.05	.005
≤34	67	3.8	2.2	3.8	3.4	6.0	2.7	8.15	.005
≤35	73	4.2	2.2	4.3	3.4	6.8	2.7	10.85*	.006
≤36	81	4.6	2.6	4.9	3.5	7.1	5.5	9.67*	.005
≤37	90	5.1	2.9	4.9	4.8	7.3	5.5	6.76	.004
≤38	100	5.7	3.3	5.4	5.3	8.3	5.5	8.21	.005
≤39	118	6.7	4.0	6.3	6.2	9.8	6.8	9.62*	.005
≤40	128	7.3	4.4	7.4	6.5	10.3	6.8	9.28	.005
≤41	144	8.2	4.8	7.9	7.8	11.3	8.2	9.71*	.005
≤42	152	8.7	4.8	8.3	8.1	12.3	8.2	12.38*	.007
≤43	167	9.5	5.1	9.4	8.8	13.6	8.2	14.30*	.008
≤44	185	10.5	5.9	10.8	9.6	14.9	9.6	14.93*	.008
≤45	199	11.3	6.2	11.7	10.6	15.6	9.6	14.94*	.008
≤46	218	12.4	7.0	13.3	11.7	16.6	9.6	15.08*	.009
≤47	250	14.2	8.8	13.7	14.9	18.4	9.6	13.84*	.008
≤48	301	17.1	10.6	16.2	18.2	21.2	15.1	13.71*	.008
≤49	448	25.5	18.3	25.6	25.5	27.2	41.1	17.40*	.010
≤50	1756	100	100	100	100	100	100	-	-
Min	10		16	14	12	10	29		
Max	50		50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1756$. [b]$n = 273$. [c]$n = 445$. [d]$n = 565$. [e]$n = 397$. [f]$n = 73$.
*$p < .05$.

Table 41

Cumulative Percentages of Participants Failing Various TOMM Trial 2 cutoffs in the Valid Sample and across Several Age Groups

TOMM	Valid Sample[a]		Age (years)						
Trial 2	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18	1	0.1	0.0	0.3	0.0	0.0	0.0	-	-
≤19	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤20	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤21	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤22	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤23	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤24	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤25	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤26	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤27	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤28	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤29	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤30	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤31	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤32	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤33	-	-	0.0	0.3	0.0	0.0	0.0	-	-
≤34	2	0.2	0.0	0.3	0.0	0.4	0.0	-	-
≤35	-	-	0.0	0.3	0.0	0.4	0.0	-	-
≤36	3	0.3	0.5	0.3	0.0	0.4	0.0	-	-
≤37	4	0.3	0.5	0.3	0.3	0.4	0.0	-	-
≤38	5	0.4	0.5	0.3	0.5	0.4	0.0	-	-
≤39	6	0.5	0.5	0.3	0.8	0.4	0.0	-	-
≤40	8	0.7	0.5	0.3	1.3	0.4	0.0	-	-
≤41	10	0.9	0.5	0.3	1.3	1.3	0.0	-	-
≤42	-	-	0.5	0.3	1.3	1.3	0.0	-	-
≤43	11	0.9	0.5	0.7	1.3	1.3	0.0	-	-
≤44	14	1.2	0.5	1.0	1.3	1.7	2.4	FET	.002
≤45	17	1.5	1.0	1.3	1.6	1.7	2.4	FET	.001
≤46	19	1.6	1.0	1.3	1.8	2.2	2.4	FET	.001
≤47	28	2.4	1.9	1.3	3.4	2.6	2.4	FET	.003
≤48	44	3.8	2.9	3.0	5.2	3.5	2.4	3.32	.003
≤49	132	11.3	8.7	11.0	11.9	9.5	31.0	18.41*	.016
≤50	1166	100	100	100	100	100	100	-	-
Min	14		36	14	37	34	44		
Max	40		50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1166$. [b]$n = 206$. [c]$n = 301$. [d]$n = 386$. [e]$n = 231$. [f]$n = 42$.
*$p < .05$.

Tables 42 and 43 characterize TOMM Trial 2 BR_{Fail} as a function of education level in the overall and valid samples, respectively. In the overall sample, Fisher's exact tests and chi-square tests of independence were not statistically significant where conducted, and education level accounted for only 0.3% to 0.5% of the variance in BR_{Fail} at previously published Trial 2 cutoffs (≤44 through ≤49). Nonetheless, visual inspection of BR_{Fail} revealed a trend involving higher BR_{Fail} among those with lower levels of education. Individuals with 16 and ≥17 years of education demonstrated somewhat higher BR_{Fail} at some cutoffs than would be expected based on the trend.

In the valid sample, Fisher's exact tests and chi-square tests of independence were conducted at Trial 2 ≤48 and ≤49. These tests failed to reach statistical significance, and the associated effect sizes indicated that education accounted for only 0.5 to 0.9% of the variance at these levels of performance (small effects).

Table 42

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Overall Sample and across Several Education Levels

TOMM	Overall Sample[a]		Education (years)							
Trial 2	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18	7	0.4	1.4	0.4	0.5	0.0	0.0	0.0	-	-
≤19	-	-	1.4	0.4	0.5	0.0	0.0	0.0	-	-
≤20	8	0.5	1.4	0.4	0.5	0.0	0.0	1.4	-	-
≤21	10	0.6	2.0	0.4	0.6	0.0	0.0	1.4	-	-
≤22	14	0.8	2.7	0.4	0.8	0.8	0.0	1.4	-	-
≤23	17	1.0	2.7	0.4	1.1	1.1	0.0	1.4	-	-
≤24	22	1.3	2.7	1.0	1.2	1.5	0.0	1.4	-	-
≤25	28	1.6	3.4	1.8	1.4	1.5	0.0	1.4	-	-
≤26	30	1.7	3.4	2.2	1.5	1.5	0.0	1.4	-	-
≤27	34	1.9	3.4	2.6	1.5	1.9	0.0	1.4	-	-
≤28	36	2.1	4.1	2.6	1.7	1.9	0.0	1.4	-	-
≤29	43	2.4	4.8	3.1	2.0	1.9	0.9	1.4	FET	.004
≤30	46	2.6	4.8	3.1	2.3	1.9	1.8	1.4	FET	.003
≤31	54	3.1	5.4	3.7	2.8	2.2	1.8	1.4	FET	.003
≤32	58	3.3	5.4	4.3	2.9	2.2	1.8	1.4	FET	.004
≤33	60	3.4	5.4	4.5	2.9	2.6	1.8	1.4	FET	.004
≤34	67	3.8	6.1	4.7	3.5	3.0	1.8	1.4	6.43	.004
≤35	73	4.2	6.1	5.3	3.7	3.0	3.7	1.4	6.21	.003
≤36	81	4.6	7.5	5.5	4.1	3.4	4.6	1.4	5.83	.003
≤37	90	5.1	8.2	5.7	4.9	3.7	5.5	1.4	6.63	.004
≤38	100	5.7	8.8	7.1	5.2	3.7	5.5	1.4	6.34	.004
≤39	118	6.7	10.2	8.1	6.4	4.1	7.3	1.4	9.18	.005
≤40	128	7.3	10.2	8.8	7.1	4.9	7.3	1.4	10.59	.006
≤41	144	8.2	10.9	9.0	8.3	5.6	7.3	6.9	9.74	.005
≤42	152	8.7	11.6	9.4	8.7	6.0	7.3	8.3	4.51	.003
≤43	167	9.5	12.9	10.2	9.8	6.7	7.3	8.3	4.61	.003
≤44	185	10.5	14.3	11.6	10.6	7.5	8.3	9.7	5.45	.003
≤45	199	11.3	15.6	12.0	11.2	8.6	10.1	11.1	6.07	.003
≤46	218	12.4	16.3	13.0	12.7	9.0	10.1	13.9	5.08	.003
≤47	250	14.2	19.0	13.9	14.4	11.2	13.8	16.7	5.84	.003
≤48	301	17.1	23.8	18.3	16.3	13.5	16.5	18.1	5.18	.003
≤49	448	25.5	34.7	28.7	22.9	21.7	23.9	25.0	8.01	.005
≤50	1756	100	100	100	100	100	100	100	-	-
Min	10		16	16	10	22	29	20		
Max	50		50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1756$. [b]$n = 147$. [c]$n = 509$. [d]$n = 652$. [e]$n = 267$. [f]$n = 109$. [g]$n = 72$.
*$p < .05$.

Table 43

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Valid Sample and across Several Education Levels

TOMM	Valid Sample[a]		Education (years)							
Trial 2	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18	1	0.1	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤19	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤20	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤21	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤22	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤23	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤24	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤25	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤26	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤27	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤28	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤29	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤30	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤31	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤32	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤33	-	-	0.0	0.0	0.2	0.0	0.0	0.0	-	-
≤34	2	0.2	0.0	0.0	0.5	0.0	0.0	0.0	-	-
≤35	-	-	0.0	0.0	0.5	0.0	0.0	0.0	-	-
≤36	3	0.3	0.0	0.3	0.5	0.0	0.0	0.0	-	-
≤37	4	0.3	1.2	0.3	0.5	0.0	0.0	0.0	-	-
≤38	5	0.4	1.2	0.6	0.5	0.0	0.0	0.0	-	-
≤39	6	0.5	1.2	0.6	0.7	0.0	0.0	0.0	-	-
≤40	8	0.7	1.2	1.2	0.7	0.0	0.0	0.0	-	-
≤41	10	0.9	1.2	1.2	0.7	0.5	0.0	1.9	-	-
≤42	-	-	1.2	1.2	0.7	0.5	0.0	1.9	-	-
≤43	11	0.9	1.2	1.5	0.7	0.5	0.0	1.9	-	-
≤44	14	1.2	2.4	2.2	0.7	0.5	0.0	1.9	-	-
≤45	17	1.5	2.4	2.2	0.9	1.5	0.0	1.9	-	-
≤46	19	1.6	2.4	2.5	1.1	1.5	0.0	1.9	-	-
≤47	28	2.4	4.9	3.1	1.8	2.0	0.0	3.8	-	-
≤48	44	3.8	6.1	5.3	2.5	3.5	2.7	3.8	FET	.005
≤49	132	11.3	18.3	14.2	8.7	10.1	9.5	11.3	10.19	.009
≤50	1166	100	100	100	100	100	100	100	-	-
Min	14		37	36	14	41	48	41		
Max	40		50	50	50	50	50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); FET = Fisher's exact test; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1166$. [b]$n = 82$. [c]$n = 323$. [d]$n = 435$. [e]$n = 199$. [f]$n = 74$. [g]$n = 53$.
*$p < .05$.

Tables 44 and 45 report TOMM Trial 2 BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. In the overall sample, Fisher's exact tests and chi-square tests of independence largely failed to reach statistical significance across TOMM Trial 2 cutoffs, suggesting no meaningful relationship between TOMM Trial 2 performance and gender. The isolated exception to this finding was observed at Trial 2 ≤45, χ^2 $(1, N = 1756) = 4.33, p = .037$. Given that the LR (1.34) was comparable to those observed at lower Trial 2 cutoffs, this significant finding was likely the result of the large sample size. In the valid sample, there was no significant relationship between TOMM Trial 2 failure and gender across cutoffs, and likelihood ratios ranged from 0.88 to 1.07, indicating largely comparable BR_{Fail} between males and females.

With regard to English language status, in the overall sample, NSEs consistently demonstrated lower BR_{Fail} than ESL patients, resulting in statistically significant Fisher's exact tests and chi-square tests across TOMM Trial 2 cutoffs. NSEs were 0.28 to 0.46 times as likely as ESL patients to fail previously published Trial 2 cutoffs (≤44 through ≤49). A similar pattern was observed across most TOMM Trial 2 scores in the valid sample, although contrasts only reached statistical significance at ≤45 and above (likely due to the larger sample size at these levels of performance). NSEs were approximately one-quarter to one-half as likely as ESL patients to fail previously published Trial 2 cutoffs in the valid sample.

Table 44

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

TOMM	Overall Sample[a]		Gender				English			
Trial 2	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤18	7	0.4	0.3	0.6	FET	0.50	0.1	2.5	FET*	0.04
≤19	-	-	0.3	0.6	FET	0.50	0.1	2.5	FET*	0.04
≤20	8	0.5	0.4	0.6	FET	0.67	0.1	2.9	FET*	0.03
≤21	10	0.6	0.5	0.7	FET	0.71	0.2	2.9	FET*	0.07
≤22	14	0.8	0.7	1.0	0.47	0.70	0.2	4.5	FET*	0.04
≤23	17	1.0	0.8	1.3	1.01	0.62	0.4	4.5	FET*	0.09
≤24	22	1.3	1.2	1.4	0.18	0.86	0.7	4.9	FET*	0.14
≤25	28	1.6	1.6	1.5	0.04	1.07	1.1	4.9	FET*	0.22
≤26	30	1.7	1.7	1.7	0.01	1.00	1.1	4.9	FET*	0.22
≤27	34	1.9	1.9	1.9	0.00	1.00	1.4	4.9	FET*	0.29
≤28	36	2.1	2.0	2.1	0.01	0.95	1.5	5.3	FET*	0.28
≤29	43	2.4	2.6	2.2	0.26	1.18	1.7	6.6	20.92*	0.26
≤30	46	2.6	2.8	2.4	0.32	1.17	1.8	7.4	26.14*	0.24
≤31	54	3.1	3.3	2.8	0.36	1.18	2.0	9.4	39.52*	0.21
≤32	58	3.3	3.7	2.8	1.05	1.32	2.3	9.4	34.24*	0.24
≤33	60	3.4	3.8	2.9	0.93	1.31	2.3	9.8	36.37*	0.23
≤34	67	3.8	4.3	3.1	1.92	1.39	2.6	11.1	41.57*	0.23
≤35	73	4.2	4.6	3.5	1.44	1.31	2.8	11.1	48.10*	0.25
≤36	81	4.6	5.	3.9	1.45	1.28	3.1	12.3	52.10*	0.25
≤37	90	5.1	5.7	4.3	1.69	1.33	3.5	13.5	59.66*	0.26
≤38	100	5.7	6.3	4.9	1.58	1.29	4.1	15.2	48.01*	0.27
≤39	118	6.7	7.3	5.8	1.53	1.26	4.8	18.0	58.52*	0.27
≤40	128	7.3	8.1	6.1	2.51	1.33	5.3	19.3	60.71*	0.27
≤41	144	8.2	9.0	7.1	2.02	1.27	5.8	22.5	78.02*	0.26
≤42	152	8.7	9.4	7.6	1.60	1.24	6.1	24.2	86.98*	0.25
≤43	167	9.5	10.4	8.2	2.46	1.27	6.8	25.8	88.11*	0.26
≤44	185	10.5	11.7	8.9	3.51	1.31	7.8	27.5	86.52*	0.28
≤45	199	11.3	12.6	9.4	4.33*	1.34	8.4	29.1	82.37*	0.29
≤46	218	12.4	13.6	10.7	3.32	1.27	9.4	30.7	87.77*	0.31
≤47	250	14.2	15.1	13.1	1.40	1.15	11.1	33.6	87.24*	0.33
≤48	301	17.1	18.2	15.6	2.16	1.17	13.6	38.9	94.79*	0.35
≤49	448	25.5	26.9	23.5	2.67	1.14	21.9	47.5	72.63*	0.46
≤50	1756	100	100	100	-	-	100	100	-	-
Min	10		12	10			14	10		
Max	50		50	50			50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1756$. [b]$n = 1036$. [c]$n = 720$. [d]$n = 1506$. [e]$n = 244$.
*$p < .05$.

Table 45

Cumulative Percentages of Participants Failing Various TOMM Trial 2 Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

TOMM	Valid Sample[a]		Gender				English			
Trial 2	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤18	1	0.1	0.2	0.0	-	-	0.1	0.0	-	-
≤19	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤20	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤21	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤22	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤23	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤24	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤25	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤26	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤27	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤28	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤29	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤30	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤31	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤32	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤33	-	-	0.2	0.0	-	-	0.1	0.0	-	-
≤34	2	0.2	0.2	0.2	FET	1.00	0.1	0.9	FET	0.11
≤35	-	-	0.2	0.2	FET	1.00	0.1	0.9	FET	0.11
≤36	3	0.3	0.2	0.4	FET	0.50	0.2	0.9	FET	0.22
≤37	4	0.3	0.3	0.4	FET	0.75	0.3	0.9	FET	0.33
≤38	5	0.4	0.3	0.6	FET	0.50	0.4	0.9	FET	0.44
≤39	6	0.5	0.5	0.6	FET	0.83	0.5	0.9	FET	0.56
≤40	8	0.7	0.6	0.8	FET	0.75	0.7	0.9	FET	0.78
≤41	10	0.9	0.6	1.2	FET	0.50	0.8	1.8	FET	0.44
≤42	-	-	0.6	1.2	FET	0.50	0.8	1.8	FET	0.44
≤43	11	0.9	0.8	1.2	FET	0.67	0.9	1.8	FET	0.50
≤44	14	1.2	1.2	1.2	0.00	1.00	1.0	2.6	FET	0.38
≤45	17	1.5	1.5	1.4	0.03	1.07	1.1	4.4	FET*	0.25
≤46	19	1.6	1.7	1.6	0.01	1.06	1.3	4.4	FET*	0.30
≤47	28	2.4	2.3	2.6	0.11	0.88	2.0	6.1	FET*	0.33
≤48	44	3.8	3.8	3.8	0.00	1.00	3.2	8.8	FET*	0.36
≤49	132	11.3	11.6	10.9	0.16	1.06	10.6	18.4	6.28*	0.58
≤50	1166	100	100	100	-	-	100	100	-	-
Min	14		14	34			14	34		
Max	40		50	50			50	50		

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TOMM Trial 2 score in the respective group; Max = maximum TOMM Trial 2 score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1166$. [b]$n = 661$. [c]$n = 505$. [d]$n = 1049$. [e]$n = 114$.
*$p < .05$.

RDS. Table 46 shows RDS BR_{Fail} at various cutoffs as a function of WMT performance. Fisher's exact tests and chi-square tests of independence were statistically significant at all RDS cutoffs. Participants who passed the WMT consistently demonstrated lower BR_{Fail} than those who failed the WMT. Those who passed the WMT were 0.22 to 0.39 times as likely as those who failed the WMT to score below previously published RDS cutoffs (i.e., RDS ≤6 and ≤7).

Table 46

Cumulative Percentages of Participants Failing Various Reliable Digit Span cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
RDS	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≤4	12	0.6	0.1	1.8	FET*	0.06
≤5	53	2.6	0.7	7.0	65.07*	0.10
≤6	155	7.7	3.7	16.9	104.58*	0.22
≤7	425	21.2	14.3	36.9	130.74*	0.39
≤8	800	39.9	32.1	57.6	114.88*	0.56
Min	0		4	0		
Max	17		17	17		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 2003$. [b]$n = 1388$. [c]$n = 615$.
*$p < .05$.

Tables 47 and 48 characterize RDS BR_{Fail} across clinical groups in the overall and valid samples, respectively. In the overall sample, chi-square tests reached statistical significance where conducted (i.e., at cutoffs between RDS ≤6 and ≤8). However, diagnosis only accounted for 1.3% to 1.4% of the variance in RDS BR_{Fail} at previously published cutoffs (RDS ≤6 and ≤7; small effects). At RDS ≤6, BR_{Fail} were highest among those with orthopedic injuries (17.0%), mTBI (12.4%), and chronic pain/fibromyalgia (11.5%), and lowest among those with severe mental illness (2.6%) and *other* diagnoses (3.2%). At RDS ≤7, the largest BR_{Fail} were observed

for patients with orthopedic injuries (32.1%), chronic pain/fibromyalgia (28.0%), mTBI (27.5%), and neurological conditions (26.1%), while patients with anxiety (16.5%) and moderate-severe TBI (16.7%) had the lowest BR_{Fail}.

In the valid sample, chi-square tests of independence or Fisher's exact tests were not conducted at RDS ≤6. However, visual inspection of BR_{Fail} at this cutoff revealed that patients with orthopedic injuries had the highest BR_{Fail} (9.1%), and those with severe mental illness (0.0%), moderate-severe TBI (1.9%), anxiety (2.0%), and *other* diagnoses (2.5%) had the lowest BR_{Fail}. Although a chi-square test of independence did reach statistical significance at RDS ≤7, $\chi^2(8, N = 1388) = 16.10$, $p = .041$, diagnosis only accounted for 1.2% of the variance in BR_{Fail} at this cutoff (small effect). Nevertheless, patients with neurological conditions (21.3%) and chronic pain/fibromyalgia (20.2%) had the highest BR_{Fail}; moderate-severe TBI patients had the lowest BR_{Fail} (9.3%), and the remaining clinical groups had intermediate BR_{Fail}.

Table 47

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
RDS	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤4	12	0.6	1.3	1.5	0.0	0.0	1.1	0.0	0.6	0.0	0.6	-	-
≤5	53	2.6	4.6	4.5	1.8	1.1	3.3	0.0	3.1	9.4	1.3	-	-
≤6	155	7.7	12.4	9.1	6.3	7.0	5.9	2.6	11.5	17.0	3.2	26.95*	.013
≤7	425	21.2	27.5	16.7	26.1	19.6	16.5	21.1	28.0	32.1	18.1	27.72*	.014
≤8	800	39.9	43.1	36.4	42.3	36.8	37.5	34.2	47.4	50.9	38.7	15.54*	.008
Min	0		0	4	5	5	2	6	2	5	2		
Max	17		15	15	15	16	17	14	15	14	14		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 2003$. [b]$n = 153$. [c]$n = 66$. [d]$n = 111$. [e]$n = 560$. [f]$n = 546$. [g]$n = 38$. [h]$n = 321$. [i]$n = 53$. [j]$n = 155$.
*$p < .05$.

Table 48

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
RDS	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤4	1	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	-	-
≤5	10	0.7	1.1	0.0	1.1	0.5	0.8	0.0	0.0	9.1	0.0	-	-
≤6	51	3.7	6.9	1.9	5.6	3.7	2.0	0.0	5.3	9.1	2.5	-	-
≤7	198	14.3	16.1	9.3	21.3	12.6	10.9	13.0	20.2	12.1	16.7	16.10*	.012
≤8	446	32.1	29.9	29.6	38.2	28.3	31.3	26.1	38.0	39.4	34.2	9.16	.007
Min	4		5	6	5	5	4	7	6	5	6		
Max	17		15	15	15	16	17	14	15	14	14		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 1388. [b]n = 87. [c]n = 54. [d]n = 89. [e]n = 381. [f]n = 393. [g]n = 23. [h]n = 208. [i]n = 33. [j]n = 120.
*$p < .05$.

Tables 49 and 50 characterize RDS BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, the youngest group demonstrated the lowest BR_{Fail}, and BR_{Fail} tended to increase with age. In the overall sample, RDS BR_{Fail} were significantly associated with age between RDS ≤5 and ≤8, and age accounted for 1.3 to 2.7% of the variance in BR_{Fail} at previously published cutoffs (i.e., RDS ≤6 and ≤7; small effects). In the valid sample, although chi-square tests reached statistical significance at higher cutoffs (likely due to the increase in sample size), age accounted for only 0.3% of the variance in BR_{Fail} at RDS ≤6 (very small effect), and for 1.9% of the variance in BR_{Fail} at RDS ≤7 (small effect).

Table 49

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
RDS	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤4	12	0.6	0.0	0.0	0.2	1.4	2.4	-	-
≤5	53	2.6	1.9	2.0	1.3	3.8	7.9	22.67*	.011
≤6	155	7.7	3.8	5.0	6.7	11.4	12.7	25.82*	.013
≤7	425	21.2	10.7	17.4	18.6	28.8	33.3	54.66*	.027
≤8	800	39.9	29.4	34.3	39.1	45.8	57.1	41.43*	.021
Min	0		5	5	2	0	2		
Max	17		17	16	15	17	14		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 2003$. [b]$n = 262$. [c]$n = 397$. [d]$n = 639$. [e]$n = 579$. [f]$n = 126$.
*$p < .05$.

Table 50

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
RDS	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤4	1	0.1	0.0	0.0	0.0	0.3	0.0	-	-
≤5	10	0.7	0.5	0.7	0.2	1.0	2.5	FET	.004
≤6	51	3.7	2.8	2.9	3.0	4.9	6.2	4.54	.003
≤7	198	14.3	8.8	12.1	11.3	19.9	24.7	26.55*	.019
≤8	446	32.1	26.9	27.9	30.6	37.0	45.7	16.38*	.012
Min	4		5	5	5	4	5		
Max	17		17	16	15	15	14		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1388$. [b]$n = 216$. [c]$n = 272$. [d]$n = 432$. [e]$n = 387$. [f]$n = 81$.
*$p < .05$.

Tables 51 and 52 characterize RDS BR_{Fail} as a function of education level in the overall and valid samples, respectively. In both samples, BR_{Fail} varied significantly across education groups at all RDS cutoffs where Fisher's exact tests or chi-square tests were conducted. The highest BR_{Fail} were observed in participants with ≤8 years of education, and BR_{Fail} tended to

decrease with higher levels of education. Participants with 16 years of education did, however, exhibit somewhat elevated BR_{Fail} than would be expected based on the trend at some cutoffs. At previously published cutoffs (RDS ≤6 and ≤7), education accounted for 1.3% to 2.5% of the variance in BR_{Fail} in the overall sample (small effects), and 1.0% to 2.3% of the variance in BR_{Fail} in the valid sample (small effects).

Table 51

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
RDS	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤4	12	0.6	3.2	0.8	0.1	0.7	0.7	0.0	-	-
≤5	53	2.6	8.7	3.7	1.4	1.6	2.6	4.1	FET*	.014
≤6	155	7.7	18.3	9.5	6.2	5.9	7.2	5.4	26.57*	.013
≤7	425	21.2	42.1	25.7	17.8	16.4	20.9	16.2	49.97*	.025
≤8	800	39.9	64.3	48.4	35.4	34.1	34.0	32.4	61.23*	.031
Min	0		2	2	4	0	4	5		
Max	17		14	17	17	15	16	15		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 2003. [b]n = 126. [c]n = 486. [d]n = 859. [e]n = 305. [f]n = 153. [g]n = 74.
*p < .05.

Table 52

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
RDS	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤4	1	0.1	1.5	0.0	0.0	0.0	0.0	0.0	-	-
≤5	10	0.7	4.5	1.0	0.3	0.4	0.0	1.9	-	-
≤6	51	3.7	11.9	4.2	3.1	2.6	3.4	1.9	FET*	.010
≤7	198	14.3	34.3	17.9	12.3	11.1	13.8	5.6	32.49*	.023
≤8	446	32.1	56.7	40.6	28.5	30.2	25.0	18.5	40.12*	.029
Min	4		4	5	5	5	6	5		
Max	17		14	14	17	15	16	15		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 1388. [b]n = 67. [c]n = 308. [d]n = 608. [e]n = 235. [f]n = 116. [g]n = 54.
*p < .05.

Tables 53 and 54 characterize RDS BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. With regard to gender, Fisher's exact tests and chi-square tests of independence largely failed to reach statistical significance in either sample, suggesting no meaningful relationship between gender and RDS performance. Likelihood ratios at previously published cutoffs (i.e., RDS ≤6 and ≤7) also suggest comparable BR_{Fail} between males and females in both the overall sample (LR = 1.17-1.20) and the valid sample (LR = 1.10-1.18).

With regard to English language status, NSEs demonstrated lower BR_{Fail} than did ESL patients across RDS cutoffs in both samples, and Fisher's exact tests and chi-square tests of independence reached statistical significance in nearly all cases. NSEs were roughly one third to one half as likely as ESL patients to fail previously published cutoffs in both samples.

Table 53

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
RDS	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤4	12	0.6	0.8	0.2	FET	4.00	0.3	2.6	FET*	0.12
≤5	53	2.6	2.8	2.5	0.18	1.12	2.1	6.9	18.24*	0.30
≤6	155	7.7	8.3	6.9	1.39	1.20	6.1	20.2	57.09*	0.30
≤7	425	21.2	22.5	19.3	2.98	1.17	18.5	42.1	68.52*	0.44
≤8	800	39.9	41.8	37.1	4.45*	1.13	36.9	62.7	56.74*	0.59
Min	0		0	2			2	0		
Max	17		17	16			17	14		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language. FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 2003. [b]n = 1190. [c]n = 813. [d]n = 1770. [e]n = 233.
*$p < .05$.

Table 54

Cumulative Percentages of Participants Failing Various Reliable Digit Span Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
RDS	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤4	1	0.1	0.1	0.0	-	-	0.1	0.0	-	-
≤5	10	0.7	0.9	0.5	FET	1.80	0.6	1.6	FET	0.38
≤6	51	3.7	3.9	3.3	0.36	1.18	3.2	8.8	FET*	0.36
≤7	198	14.3	14.9	13.5	0.54	1.10	12.9	28.0	21.19*	0.46
≤8	446	32.1	33.8	30.0	2.32	1.13	30.1	52.8	26.91*	0.57
Min	4		4	5			4	5		
Max	17		17	16			17	13		

Note. RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum RDS score in the respective group; Max = maximum RDS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 1388$. [b]$n = 787$. [c]$n = 601$. [d]$n = 1263$. [e]$n = 125$.
*$p < .05$.

WCST FMS. Table 55 shows WCST FMS BR_{Fail}, at various cutoffs, as a function of WMT performance. Participants who passed the WMT demonstrated lower BR_{Fail} than those who failed the WMT across cutoffs. At the previously published cutoffs of FMS ≥2 and ≥3, likelihood ratios were 0.49 and 0.66, respectively.

Table 55

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

WCST	Overall Sample[a]		WMT			
FMS	*f*	%	Pass	Fail	χ^2	LR
≥7	4	0.7	0.4	1.2	FET	0.33
≥6	11	1.8	1.5	2.3	FET	0.65
≥5	26	4.2	2.4	8.4	11.61*	0.29
≥4	57	9.1	5.9	16.8	18.40*	0.35
≥3	117	18.6	14.3	29.1	18.61*	0.49
≥2	199	31.5	27.5	41.4	11.47*	0.66
≥1	366	57.8	56.1	62.1	1.87	0.90
Min	0		0	0		
Max	8		7	8		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; FET = Fisher's exact test; LR = likelihood ratio; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 634$. [b]$n = 455$. [c]$n = 179$.
*$p < .05$.

Tables 56 and 57 characterize WCST FMS BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. In both samples, chi-square tests of independence conducted at FMS cutoffs of ≥1 and ≥2 failed to reach statistical significance. Diagnosis accounted for only 1.6 to 1.9% of the variance in BR_{Fail} at these cutoffs in the overall sample (small effects), and 1.4 to 3.3% of the variance in BR_{Fail} at these cutoffs in the valid sample (small to medium effects). Chi-square tests were not conducted at the previously published cutoff of FMS ≥3. However, visual inspection revealed that patients with mild TBI and orthopedic injury tended to

demonstrate high BR_{Fail} relative to other groups. Patients with anxiety demonstrated the lowest BR_{Fail}, although the group was somewhat small in size (n = 24 in the overall sample, n = 18 in the valid sample).

Table 56

Cumulative Percentages of Participants Failing Various WCST FMS cutoffs in the Overall Sample and in Specific Clinical Groups

WCST	Overall Sample[a]		TBI		Diagnostic Groups								
FMS	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥7	4	0.7	0.0	3.3	1.0	0.0	0.0	0.0	0.0	0.0	0.8	-	-
≥6	11	1.8	0.7	5.0	1.0	1.0	0.0	4.0	2.8	0.0	2.4	-	-
≥5	26	4.2	5.7	6.7	1.0	4.1	0.0	4.0	5.6	4.3	4.0	-	-
≥4	57	9.1	10.7	11.7	8.6	7.2	0.0	8.0	11.2	13.0	8.0	-	-
≥3	117	18.6	28.0	21.7	17.2	10.3	0.0	20.0	19.5	26.0	15.1	-	-
≥2	199	31.5	38.8	36.7	32.4	22.8	12.5	36.0	30.6	30.3	29.4	12.10	.019
≥1	366	57.8	64.7	66.7	54.3	50.9	50.0	64.0	50.0	65.1	54.8	9.90	.016
Min	0		0	0	0	0	0	0	0	0	0		
Max	8		6	7	8	6	2	6	6	5	7		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]n = 634. [b]n = 139. [c]n = 60. [d]n = 105. [e]n = 96. [f]n = 24. [g]n = 25. [h]n = 36. [i]n = 23. [j]n = 126.
*p < .05.

Table 57

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Valid Sample and in Specific Clinical Groups

WCST	Valid Sample[a]		TBI		Diagnostic Groups								
FMS	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥7	2	0.4	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	-	-
≥6	7	1.5	0.0	4.0	0.0	1.3	0.0	6.7	0.0	0.0	3.2	-	-
≥5	11	2.4	1.3	4.0	0.0	2.6	0.0	6.7	4.3	0.0	4.3	-	-
≥4	27	5.9	5.1	6.0	7.1	2.6	0.0	13.4	4.3	13.3	7.5	-	-
≥3	65	14.3	21.8	16.2	15.4	6.4	0.0	20.1	13.0	20.0	13.8	-	-
≥2	125	27.4	32.1	30.5	29.7	20.5	11.1	33.4	21.7	26.7	29.8	6.43	.014
≥1	255	56.1	68.0	67.5	51.1	47.4	44.4	66.7	39.1	46.7	58.0	15.19	.033
Min	0		0	0	0	0	0	0	0	0	0		
Max	7		5	7	4	6	2	6	5	4	7		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; TBI = traumatic brain injury; Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 455$. [b]$n = 78$. [c]$n = 49$. [d]$n = 84$. [e]$n = 78$. [f]$n = 18$. [g]$n = 15$. [h]$n = 23$. [i]$n = 15$. [j]$n = 95$.
*$p < .05$.

Tables 58 and 59 characterize WCST FMS BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, Fisher's exact tests and chi-square tests of independence failed to reach statistical significance where conducted. At previously published cutoffs (WCST ≥2 and ≥3), age accounted for only 0.1 to 0.3% of the variance in BR_{Fail} in the overall sample (very small effects), and only 0.2% to 0.5% of the variance in BR_{Fail} in the valid sample (very small effects).

Table 58

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Overall Sample and across Several Age Groups

WCST	Overall Sample[a]		Age (years)						
FMS	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥7	4	0.7	1.6	1.7	0.0	0.4	0.0	-	-
≥6	11	1.8	3.2	2.6	1.1	1.3	1.8	FET	.003
≥5	26	4.2	4.8	6.1	4.0	3.5	1.8	FET	.003
≥4	57	9.1	7.9	10.4	9.1	8.0	10.6	0.78	.001
≥3	117	18.6	23.8	19.1	18.2	16.5	19.4	1.78	.003
≥2	199	31.5	28.6	31.3	33.6	30.4	31.7	0.76	.001
≥1	366	57.8	58.7	60.0	58.7	57.8	49.2	2.09	.003
Min	0		0	0	0	0	0		
Max	8		7	7	6	8	6		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; FET = Fisher's exact test; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 634$. [b]$n = 63$. [c]$n = 115$. [d]$n = 175$. [e]$n = 223$. [g]$n = 57$.
*$p < .05$.

Table 59

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Valid Sample and across Several Age Groups

WCST	Valid Sample[a]		Age (years)						
FMS	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥7	2	0.4	1.9	1.3	0.0	0.0	0.0	-	-
≥6	7	1.5	3.8	2.6	1.7	0.6	0.0	-	-
≥5	11	2.4	5.8	3.9	2.5	1.2	0.0	-	-
≥4	27	5.9	7.7	5.2	6.7	5.4	5.3	FET	.001
≥3	65	14.3	21.2	13.0	14.2	13.1	13.2	2.35	.005
≥2	125	27.4	25.0	24.7	30.0	28.0	26.4	0.89	.002
≥1	255	56.1	57.7	58.5	57.5	56.0	44.8	2.31	.005
Min	0		0	0	0	0	0		
Max	7		7	7	6	6	4		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; FET = Fisher's exact test; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 455$. [b]$n = 52$. [c]$n = 77$. [d]$n = 120$. [e]$n = 168$. [f]$n = 38$.
*$p < .05$.

Tables 60 and 61 characterize WCST FMS BR_{Fail} as a function of education level in the overall and valid samples, respectively. Chi-square tests of independence conducted at previously published cutoffs (WCST FMS ≥2 and ≥3) were statistically significant in the overall sample but failed to reach statistical significance in the valid sample. Education accounted for 1.9 to 2.6% of the variance in BR_{Fail} at these cutoffs in the overall sample (small effects), and 2.2 to 2.4% of the variance in BR_{Fail} at these cutoffs in the valid sample (small effects).

Estimates of BR_{Fail} for patients with ≤8 years of formal education were likely unreliable due to the small group size. Patients with 9-11 years of education demonstrated the highest BR_{Fail} at previously published cutoffs in both samples, however, no consistent trends in BR_{Fail} were noted across the remaining groups (12 years through ≥17 years of education).

Table 60

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Overall Sample and across Several Education Levels

WCST	Overall Sample[a]		Education (years)							
FMS	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥7	4	0.7	0.0	1.6	0.4	0.0	0.0	2.5	-	-
≥6	11	1.8	0.0	2.4	2.1	1.6	0.0	2.5	-	-
≥5	26	4.2	0.0	4.0	5.9	3.9	1.3	2.5	-	-
≥4	57	9.1	4.8	12.8	9.7	8.6	6.3	2.5	5.62	.009
≥3	117	18.6	14.3	28.8	17.6	14.0	15.1	15.0	11.94*	.019
≥2	199	31.5	23.8	44.8	31.8	24.1	27.6	22.5	16.29*	.026
≥1	366	57.8	66.7	70.4	55.6	52.0	53.9	52.5	12.08*	.019
Min	0		0	0	0	0	0	0		
Max	8		4	7	7	6	5	8		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 634$. [b]$n = 21$. [c]$n = 125$. [d]$n = 239$. [e]$n = 129$. [f]$n = 80$. [g]$n = 40$.
*$p < .05$.

Table 61

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Valid Sample and across Several Education Levels

WCST	Valid Sample[a]		Education (years)							
FMS	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥7	2	0.4	0.0	1.3	0.6	0.0	0.0	0.0	-	-
≥6	7	1.5	0.0	2.6	1.7	2.1	0.0	0.0	-	-
≥5	11	2.4	0.0	10.1	3.4	3.2	0.0	0.0	-	-
≥4	27	5.9	0.0	17.6	5.7	8.5	1.6	0.0	-	-
≥3	65	14.3	11.1	32.6	13.6	11.7	9.4	9.7	9.96	.022
≥2	125	27.4	22.2	46.4	30.0	20.2	23.5	16.1	10.79	.024
≥1	255	56.1	77.8	76.4	57.7	48.9	46.9	48.4	12.00*	.026
Min	0		0	0	0	0	0	0		
Max	7		3	7	7	6	4	3		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 455$. [b]$n = 9$. [c]$n = 80$. [d]$n = 177$. [e]$n = 94$. [f]$n = 64$. [g]$n = 31$.
*$p < .05$.

Tables 62 and 63 characterize WCST FMS performance as a function of gender and English language status in the overall and valid samples, respectively. BR_{Fail} on WCST FMS were not significantly related to gender at the previously published cutoff of WCST FMS ≥2 in either sample (LR = 1.09 in the overall sample, LR = 0.98 in the valid sample). However, males were 1.44 times as likely as females to fail WCST FMS ≥3 in the overall sample and 1.36 times as likely as females to fail WCST FMS ≥3 in the valid sample.

BR_{Fail} on WCST FMS were unrelated to English language status in either sample. Likelihood ratios at previously published cutoffs ranged from 0.91 to 1.04 in the overall sample, and from 0.90 to 1.18 in the valid sample.

Table 62

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

WCST	Overall Sample[a]		Gender				English			
FMS	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥7	4	0.7	0.9	0.3	FET	3.00	0.5	1.4	FET	0.36
≥6	11	1.8	1.5	2.0	FET	0.75	1.4	4.3	FET	0.33
≥5	26	4.2	2.5	4.8	0.81	0.52	3.9	5.7	0.57	0.68
≥4	57	9.1	9.6	8.3	0.25	1.16	8.5	12.9	1.55	0.66
≥3	117	18.6	21.4	14.9	4.20*	1.44	18.2	20.1	0.17	0.91
≥2	199	31.5	32.6	29.9	0.49	1.09	31.5	30.2	0.03	1.04
≥1	366	57.8	60.3	54.6	1.97	1.10	57.7	57.7	0.00	1.00
Min	0		0	0			0	0		
Max	8		8	7			7	8		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 634$. [b]$n = 347$. [c]$n = 287$. [d]$n = 565$. [e]$n = 69$.
*$p < .05$.

Table 63

Cumulative Percentages of Participants Failing Various WCST FMS Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

WCST	Valid Sample[a]		Gender				English			
FMS	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥7	2	0.4	0.8	0.0	-	-	0.5	0.0	-	-
≥6	7	1.5	1.2	1.9	FET	0.63	1.5	2.6	FET	0.58
≥5	11	2.4	2.4	2.4	0.00	1.00	2.2	5.2	FET	0.42
≥4	27	5.9	6.5	5.3	0.31	1.23	5.6	10.5	FET	0.53
≥3	65	14.3	16.3	12.0	1.71	1.36	14.2	15.8	0.08	0.90
≥2	125	27.4	27.3	27.8	0.02	0.98	27.9	23.7	0.30	1.18
≥1	255	56.1	58.8	54.1	0.61	1.09	56.7	50.0	0.62	1.13
Min	0		0	0			0	0		
Max	7		7	6			7	6		

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum FMS score in the respective group; Max = maximum FMS score in the respective group. Boldface font denotes previously published cutoffs.
[a]$n = 455$. [b]$n = 246$. [c]$n = 209$. [d]$n = 417$. [e]$n = 38$.
*$p < .05$.

CT-TE. Table 64 shows CT-TE BR_{Fail}, at various cutoffs, as a function of WMT performance. Participants who passed the WMT demonstrated lower BR_{Fail} than those who failed the WMT at various CT-TE scores, and chi-square tests reached statistical significance at all CT-TE cutoffs examined. Likelihood ratios ranged from 0.41 to 0.47.

Table 64

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
CT-TE	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≥123	51	2.0	1.5	3.3	7.29*	0.45
≥113	123	4.8	3.4	8.3	22.48*	0.41
≥105	229	9.9	7.1	17.1	48.33*	0.42
≥98	338	14.7	10.8	24.8	68.41*	0.44
≥85	551	25.0	19.1	40.6	108.41*	0.47
Min	3		3	10		
Max	150		150	146		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio. Min = minimum CT-TE score in the respective group. Max = maximum CT-TE score in the respective group.
[a]$n = 2199$. [b]$n = 1595$. [c]$n = 604$.
*$p < .05$.

Tables 65 and 66 characterize CT-TE BR_{Fail} across various clinical groups in the overall and valid samples, respectively. In the overall sample, chi-square tests were statistically significant at CT-TE ≥105 and ≥85. However, diagnosis accounted for only 0.3 to 0.8% of the variance in CT-TE BR_{Fail} across cutoffs examined (very small effects). The highest BR_{Fail} tended to be observed in patients with severe mental illness and neurological conditions, while patients with depression and mTBI demonstrated the lowest BR_{Fail}. A similar pattern was observed in the valid sample. Chi-square tests were statistically significant at lower (i.e., more liberal) cutoffs. However, diagnosis accounted for only 0.4 to 1.6% of the variance in CT-TE BR_{Fail} (very small to small effects).

Table 65

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI	Diagnostic Groups									
CT-TE	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEPe	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥123	51	2.0	1.3	1.6	1.9	1.5	2.7	3.7	3.5	2.2	2.3	FET	.003
≥113	123	4.8	3.6	6.5	6.0	3.5	4.4	5.6	6.4	7.8	4.7	7.48	.003
≥105	229	9.9	7.7	12.4	14.8	7.5	9.2	16.7	11.7	14.4	8.5	18.75*	.008
≥98	338	14.7	13.6	15.6	20.8	11.0	14.0	20.4	17.5	16.7	14.0	14.52	.007
≥85	551	25.0	22.8	25.8	34.3	23.5	22.9	37.0	24.0	28.9	22.5	18.40*	.008
Min	3		9	9		3	8	4	5	12	5		
Max	150		133	138	146	150	128	126	136	129	130		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum CT-TE score in the respective group; Max = maximum CT-TE score in the respective group.
[a]n = 2199. [b]n = 531. [c]n = 186. [d]n = 216. [e]n = 400. [f]n = 293. [g]n = 54. [h]n = 171. [i]n = 90. [j]n = 258.
*p < .05.

Table 66

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
CT-TE	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥123	24	1.5	0.6	2.0	0.0	1.0	2.4	5.0	1.7	1.5	2.9	-	-
≥113	54	3.4	2.0	6.1	4.7	2.8	2.8	5.0	3.3	4.4	3.9	7.25	.004
≥105	113	7.1	4.6	10.1	11.8	4.5	5.2	17.5	8.3	8.8	7.8	22.18*	.014
≥98	172	10.8	8.4	13.5	17.2	7.3	8.1	20.0	11.7	10.3	13.6	20.78*	.013
≥85	305	19.1	14.7	23.0	29.0	17.1	14.7	32.5	15.8	22.1	20.9	26.19*	.016
Min	3		9	9	11	3	8	4	5	12	5		
Max	150		126	138	121	150	128	126	136	129	130		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum CT-TE score in the respective group; Max = maximum CT-TE score in the respective group.
[a]n = 1595. [b]n = 347. [c]n = 148. [d]n = 169. [e]n = 286. [f]n = 211. [g]n = 40. [h]n = 120. [i]n = 68. [j]n = 206.
*$p < .05$.

Tables 67 and 68 characterize CT-TE BR_{Fail} as a function of age in the overall and valid samples, respectively. Chi-square tests were statistically significant at nearly all cutoffs examined in both samples. Effect sizes indicated that age accounted for 0.6% to 6.1% of the variance in CT-TE BR_{Fail} across cutoffs in the overall sample (very small to medium effects), and 0.3% to 5.5% of the variance in CT-TE BR_{Fail} in the valid sample (very small to medium effects). BR_{Fail} were lowest in the youngest age group (16-29 years) and increased with age.

Table 67

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Overall Sample and across Various Age Groups

	Overall Sample[a]		Age (years)						
CT-TE	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥123	51	2.0	0.3	1.2	1.9	2.9	4.4	12.45*	.006
≥113	123	4.8	1.6	2.7	4.9	6.4	12.3	29.77*	.013
≥105	229	9.9	2.5	5.1	8.8	15.2	28.1	94.34*	.043
≥98	338	14.7	4.1	7.4	14.1	22.6	36.8	123.66*	.056
≥85	551	25.0	9.2	16.2	25.4	34.9	50.9	133.81*	.061
Min	3		5	5	3	10	15		
Max	150		124	133	136	150	136		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); Min = minimum CT-TE score in the respective group; Max = maximum CT-TE score in the respective group.
[a]*n* = 2199. [b]*n* = 315. [c]*n* = 489. [d]*n* = 697. [e]*n* =579. [f]*n* = 114.
*$p < .05$.

Table 68

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Valid Sample and across Various Age Groups

	Valid Sample[a]		Age (years)						
CT-TE	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥123	24	1.5	0.4	1.1	1.6	2.4	1.4	4.67	.003
≥113	54	3.4	1.2	1.7	3.2	6.0	7.0	18.07*	.011
≥105	113	7.1	1.6	2.8	6.0	13.0	21.1	65.23*	.041
≥98	172	10.8	2.4	4.5	9.9	19.0	29.6	88.05*	.055
≥85	305	19.1	6.8	12.1	18.5	28.6	42.3	84.89*	.053
Min	3		5	5	3	10	15		
Max	150		124	133	136	150	127		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); Min = minimum CT-TE score in the respective group; Max = maximum CT-TE score in the respective group.
[a]*n* = 1595. [b]*n* = 249. [c]*n* = 354. [d]*n* = 503. [e]*n* = 416. [f]*n* = 71.
*$p < .05$.

Tables 69 and 70 characterize CT-TE BR_{Fail} across education levels in the overall and valid samples, respectively. Chi-square and Fisher's exact tests were statistically significant at all CT-TE scores in both samples. Education accounted for 0.6% (small effect) to 4.5% (small to medium effect) of the variance in CT-TE BR_{Fail} across cutoffs in the overall sample, and 1.0% (small effect) to 3.2% (small to medium effect) of the variance in CT-TE BR_{Fail} across cutoffs in the valid sample. BR_{Fail} were highest in the least educated group (≤8 years) and decreased with higher levels of education.

Table 69

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
CT-TE	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥123	51	2.0	5.7	2.6	2.2	0.8	0.9	0.8	FET*	.006
≥113	123	4.8	7.6	6.8	5.1	3.1	1.9	1.7	15.53*	.007
≥105	229	9.9	29.5	14.1	7.7	6.2	6.1	6.7	71.28*	.032
≥98	338	14.7	36.2	20.5	12.9	9.3	8.5	10.1	72.87*	.033
≥85	551	25.0	56.2	33.1	21.6	18.9	15.1	21.0	98.61*	.045
Min	3		9	5	5	3	4	11		
Max	150		130	150	138	136	141	130		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); FET = Fisher's exact test; Min = Minimum CT-TE score in the respective group; Max = Maximum CT-TE score in the respective group.
[a]$n = 2199$. [b]$n = 105$. [c]$n = 547$. [d]$n = 830$. [e]$n = 386$. [f]$n = 212$. [g]$n = 119$.
*$p < .05$.

Table 70

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
CT-TE	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥123	24	1.5	5.1	2.8	1.1	1.0	0.0	1.0	-	-
≥113	54	3.4	6.8	5.8	3.6	2.0	0.6	1.0	15.70*	.010
≥105	113	7.1	23.7	10.9	5.5	5.0	4.3	5.2	39.06*	.024
≥98	172	10.8	32.2	15.9	9.2	7.7	6.1	7.3	47.04*	.030
≥85	305	19.1	44.1	26.7	17.0	14.7	11.0	15.6	50.64*	.032
Min	3		9	5	5	3	4	11		
Max	150		129	150	138	136	117	130		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); Min = Minimum CT-TE score in the respective group; Max = Maximum CT-TE score in the respective group.
[a]n = 1595. [b]n = 59. [c]n = 359. [d]n = 618. [e]n = 300. [f]n = 163. [g]n = 96.
*p < .05.

Tables 71 and 72 characterize CT-TE BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. In the overall sample, BR_{Fail} were generally comparable between males and females, and chi-square and Fisher's exact tests failed to reach statistical significance at the CT-TE cutoffs examined. Likelihood ratios ranged from 0.62 to 0.89. BR_{Fail} between males and females were also generally comparable in the valid sample (LR = 0.73 – 1.06). Although chi-square tests were statistically significant at more liberal cutoffs in the valid sample (i.e., CT-TE ≥98 and ≥85), this was likely due to the large sample size at these cutoffs.

With regard to English language status, NSEs demonstrated lower BR_{Fail} than ESL patients across nearly all CT-TE cutoffs in both samples, and chi-square tests were significant in a number of cases. NSEs were only about one-half as likely as ESL patients to fail CT-TE ≥85 through ≥113 in the overall sample, and two-thirds to three-quarters as likely as ESL patients to fail CT-TE ≥85 through ≥113 in the valid sample.

Table 71

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
CT-TE	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥123	51	2.0	1.6	2.6	2.26	0.62	2.0	2.2	FET	0.91
≥113	123	4.8	4.5	5.2	0.67	0.87	4.3	7.8	5.51*	0.55
≥105	229	9.9	9.4	10.6	0.83	0.89	9.0	15.5	10.10*	0.58
≥98	338	14.7	13.5	16.5	3.82	0.82	13.2	25.9	26.74*	0.51
≥85	551	25.0	23.7	26.9	2.79	0.88	23.0	41.4	37.60*	0.56
Min	3		4	3			3	9		
Max	150		146	150			150	136		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum CT-TE score in the respective group. Max = maximum CT-TE score in the respective group.
[a]*n* = 2199. [b]*n* = 1340. [c]*n* = 859. [d]*n* = 1947. [e]*n* = 232.
*$p < .05$.

Table 72

Cumulative Percentages of Participants Failing Various Category Test Total Errors Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
CT-TE	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥123	24	1.5	1.3	1.8	0.71	0.72	1.6	0.7	FET	2.29
≥113	54	3.4	3.5	3.3	0.06	1.06	3.4	4.3	FET	0.79
≥105	113	7.1	6.3	8.3	2.25	0.76	6.9	9.3	1.14	0.74
≥98	172	10.8	9.4	12.8	4.57*	0.73	10.3	16.4	5.05*	0.63
≥85	305	19.1	17.1	21.9	5.81*	0.78	18.2	28.6	8.97*	0.64
Min	3		4	3			3	9		
Max	150		133	150			150	124		

Note. CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum CT-TE score in the respective group; Max = maximum CT-TE score in the respective group.
[a]*n* = 1595. [b]*n* = 932. [c]*n* = 663. [d]*n* = 1443. [e]*n* = 140.
*$p < .05$.

TMT-A. Table 73 shows TMT-A BR_{Fail}, across various cutoffs, as a function of WMT performance. Chi-square tests were statistically significant at all TMT-A cutoffs, suggesting an association between TMT-A performance and WMT classification. Those who passed the WMT demonstrated lower TMT-A BR_{Fail} than those who failed the WMT across cutoffs. Likelihood ratios ranged from 0.03 to 0.42.

Table 73

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
TMT-A	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≥92	47	1.9	0.2	6.0	88.63*	0.03
≥73	120	4.9	1.9	11.9	112.77*	0.16
≥57	251	10.2	5.7	20.8	129.34*	0.27
≥50	372	15.1	8.8	29.9	179.01*	0.29
≥43	599	24.3	17.1	41.0	160.58*	0.42
Min	9		9	12		
Max	302		188	302		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 2463$. [b]$n = 1724$. [c]$n = 739$.
*$p < .05$.

Tables 74 and 75 characterize TMT-A BR_{Fail} across clinical groups in the overall and valid samples, respectively. Fisher's exact tests and chi-square tests were statistically significant where conducted in both samples, suggesting that TMT-A BR_{Fail} varied as a function of diagnosis. However, diagnosis only accounted for 0.8% to 2.1% of the variance across TMT-A cutoffs in the overall sample (small effects), and for 2.2% to 3.0% of the variance between cutoffs of TMT-A ≥43 and ≥57 in the valid sample (small to medium effects).

In the overall sample, the highest BR_{Fail} were most consistently observed in patients with neurological conditions and those with chronic pain/fibromyalgia, although those with moderate-severe TBI also had higher BR_{Fail} at more liberal TMT-A cutoff scores. The lowest BR_{Fail} were observed in patients with anxiety. In the valid sample, patients with neurological conditions had the highest BR_{Fail}, and at more liberal TMT-A cutoffs, elevated BR_{Fail} were observed in patients with moderate-severe TBI, chronic pain/fibromyalgia, and severe mental illness. Patients with anxiety and mTBI demonstrated the lowest BR_{Fail}.

Table 74

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
TMT-A	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ
≥92	47	1.9	2.3	2.6	2.2	0.6	0.6	1.7	6.6	1.3	1.5	FET*	.01
≥73	120	4.9	5.9	4.6	6.6	3.8	2.4	3.4	9.4	4.0	3.8	19.22*	.00
≥57	251	10.2	12.0	9.2	17.5	8.1	5.6	5.2	15.5	10.7	8.0	35.37*	.01
≥50	372	15.1	16.0	16.8	23.1	11.8	9.4	17.2	20.7	16.0	14.0	30.78*	.01
≥43	599	24.3	23.7	32.1	35.8	21.2	14.2	25.9	31.5	25.3	23.5	51.27*	.02
Min	9		12	11	14	12	13	16	14	9	11		
Max	302		302	188	152	213	252	119	238	92	144		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 2463$. [b]$n = 557$. [c]$n = 196$. [d]$n = 229$. [e]$n = 532$. [f]$n = 339$. [g]$n = 58$. [h]$n = 213$. [i]$n = 75$. [j]$n = 264$.
*$p < .05$.

Table 75

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
TMT-A	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥92	4	0.2	0.0	1.3	0.0	0.0	0.0	0.0	0.7	0.0	0.5	-	-
≥73	33	1.9	0.9	3.2	4.6	1.4	1.2	0.0	2.9	0.0	2.0	-	-
≥57	98	5.7	3.9	7.8	13.7	4.6	2.7	2.5	10.2	2.0	4.4	37.39*	.022
≥50	152	8.8	5.4	11.7	19.4	6.5	6.2	10.0	12.4	3.9	9.4	39.64*	.023
≥43	295	17.1	11.0	25.3	30.9	14.9	10.4	20.0	22.6	15.7	17.7	52.11*	.030
Min	9		12	11	14	14	13	16	14	9	11		
Max	188		86	188	90	90	75	65	109	57	111		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]n = 1724. [b]n = 335. [c]n = 154. [d]n = 175. [e]n = 370. [f]n = 259. [g]n = 40. [h]n = 137. [i]n = 51. [j]n = 203.
*$p < .05$.

Tables 76 and 77 characterize TMT-A BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, chi-square and Fisher's exact tests were statistically significant across TMT-A cutoffs. Age accounted for 1.5% (small effect) to 5.2% (small to medium effect) of the variance in TMT-A BR_{Fail} in the overall sample, and 0.8% (very small effect) to 4.2% (small to medium effect) of the variance in TMT-A BR_{Fail} in the valid sample. In both samples, BR_{Fail} tended to increase with age and were often highest among oldest age group (60-69 years).

Table 76

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
TMT-A	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥92	47	1.9	0.0	1.5	2.4	2.4	2.7	-	-
≥73	120	4.9	1.2	3.5	4.2	6.7	12.8	37.22*	.015
≥57	251	10.2	3.7	5.6	9.4	14.1	25.5	77.57*	.032
≥50	372	15.1	6.2	9.3	13.9	20.0	37.6	106.91*	.044
≥43	599	24.3	13.2	14.3	23.8	31.6	51.0	127.75*	.052
Min	9		9	12	12	14	17		
Max	302		81	238	258	219	252		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 2463$. [b]$n = 325$. [c]$n = 518$. [d]$n = 791$. [e]$n = 674$. [f]$n = 149$.
*$p < .05$.

Table 77

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
TMT-A	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥92	4	0.2	0.0	0.0	0.2	0.6	0.0	-	-
≥73	33	1.9	0.8	1.1	1.1	3.6	3.3	FET*	.008
≥57	98	5.7	2.7	2.2	4.9	9.0	14.4	35.41*	.020
≥50	152	8.8	5.0	4.5	7.1	12.4	27.8	62.94*	.036
≥43	295	17.1	10.5	9.7	15.4	23.9	40.0	71.59*	.042
Min	9		9	12	14	14	17		
Max	188		81	90	109	188	90		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 1724$. [b]$n = 258$. [c]$n = 359$. [d]$n = 546$. [e]$n = 468$. [f]$n = 90$.
*$p < .05$.

Tables 78 and 79 characterize TMT-A BR_{Fail} as a function of education level in the overall and valid samples, respectively. In the overall sample, all of the chi-square and Fisher's exact tests were statistically significant, and education level accounted for 1.6% to 2.9% of the variance in TMT-A BR_{Fail} across cutoffs (small effects). The highest BR_{Fail} were observed in the group with the least education (≤8 years), and BR_{Fail} tended to decrease with higher levels of education. Individuals with ≥17 years of education demonstrated somewhat higher BR_{Fail} than would be expected based on the trend.

In the valid sample, chi-square and Fisher's exact tests largely failed to reach statistical significance, and education level accounted for only 0.2% to 1.2% of the variance in TMT-A BR_{Fail} across cutoffs (very small to small effects). Although the chi-square test at TMT-A ≥43 was statistically significant, given the effect size, this was likely due to the increase in sample size at this cutoff. Visual inspection of BR_{Fail} revealed higher BR_{Fail} among the two least educated groups (≤8 and 9-11 years of education) with lower and more similar BR_{Fail} among the remaining groups.

Table 78

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
TMT-A	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥92	47	1.9	8.7	1.8	0.9	2.2	2.2	1.7	FET*	.016
≥73	120	4.9	15.9	4.9	3.4	3.7	5.6	6.6	43.25*	.018
≥57	251	10.2	26.8	11.2	8.0	8.4	10.3	10.7	49.01*	.020
≥50	372	15.1	36.2	17.1	11.4	15.0	13.4	15.7	60.88*	.025
≥43	599	24.3	49.3	28.6	19.0	22.9	21.1	27.3	70.01*	.029
Min	9		12	11	11	12	9	14		
Max	302		302	121	258	144	219	104		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 2463$. [b]$n = 138$. [c]$n = 597$. [d]$n = 968$. [e]$n = 407$. [f]$n = 232$. [g]$n = 121$.
*$p < .05$.

Table 79

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
TMT-A	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥92	4	0.2	0.0	0.3	0.1	0.3	0.6	0.0	-	-
≥73	33	1.9	2.9	1.6	1.8	1.3	1.7	4.6	FET	.002
≥57	98	5.7	13.0	6.8	5.2	4.4	5.0	4.6	9.53	.005
≥50	152	8.8	18.8	10.0	8.0	7.7	7.3	9.2	10.85	.006
≥43	295	17.1	33.3	20.5	14.9	15.4	14.0	19.5	20.56*	.012
Min	9		15	11	11	12	9	14		
Max	188		79	111	109	109	188	90		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-A score in the respective group; Max = maximum TMT-A score in the respective group.
[a]$n = 1724$. [b]$n = 69$. [c]$n = 380$. [d]$n = 711$. [e]$n = 298$. [f]$n = 179$. [g]$n = 66$. [h]$n = 87$.
*$p < .05$.

Tables 80 and 81 characterize TMT-A BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. Chi-square and Fisher's exact tests involving gender did not reach statistical significance in either sample, and likelihood ratios suggested largely comparable BR_{Fail} between males and females in both samples.

With regard to English language status, NSEs consistently demonstrated lower BR_{Fail} across TMT-A cutoffs than did ESL patients, and chi-square and Fisher's exact tests were statistically significant in all but one cutoff (TMT-A ≥92 in the valid sample, $p = .050$). NSEs were 0.09 to 0.43 times as likely as ESL patients to fail various TMT-A cutoffs in the overall sample, and 0.08 to 0.47 times as likely as ESL patients to fail various TMT-A cutoffs in the valid sample.

Table 80

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
TMT-A	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥92	47	1.9	2.1	1.7	0.63	1.24	0.9	9.5	102.61*	0.09
≥73	120	4.9	5.0	4.7	0.10	1.06	3.1	17.3	117.15*	0.18
≥57	251	10.2	10.1	10.4	0.08	0.97	7.8	27.8	116.44*	0.28
≥50	372	15.1	15.5	14.5	0.47	1.07	12.1	36.6	125.52*	0.33
≥43	599	24.3	24.6	23.8	0.18	1.03	20.9	48.4	110.20*	0.43
Min	9		11	9			9	12		
Max	302		302	258			302	258		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language. Min = minimum TMT-A score in the respective group. Max = maximum TMT-A score in the respective group.
[a]n = 2463. [b]n = 1506. [c]n = 957. [d]n = 2139. [e]n = 306.
*p < .05.

Table 81

Cumulative Percentages of Participants Failing Various Trail Making Test Part A Completion Time Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
TMT-A	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥92	4	0.2	0.3	0.1	FET	3.00	0.1	1.2	FET	0.08
≥73	33	1.9	2.1	1.6	0.59	1.31	1.4	6.0	FET*	0.23
≥57	98	5.7	5.3	6.2	0.65	0.85	4.7	15.6	33.27*	0.30
≥50	152	8.8	9.2	8.2	0.56	1.12	7.6	20.4	30.19*	0.37
≥43	295	17.1	17.5	16.5	0.27	1.06	15.5	32.9	32.37*	0.47
Min	9		11	9			9	12		
Max	188		188	111			111	188		

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TMT-A score in the respective group. Max = maximum TMT-A score in the respective group.
[a]n = 1724. [b]n = 1017. [c]n = 707. [d]n = 1546. [e]n = 167.
*p < .05.

TMT-B. Table 82 shows TMT-B BR_{Fail}, across various cutoffs, as a function of WMT performance. Chi-square tests of independence were statistically significant at each TMT-B cutoff, suggesting that performance on TMT-B was related to WMT classification. Participants who passed the WMT demonstrated lower TMT-B BR_{Fail} than those who failed the WMT across cutoffs. Likelihood ratios ranged from 0.12 to 0.39.

Table 82

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
TMT-B	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≥304	49	2.0	0.6	5.2	57.14*	0.12
≥208	122	5.0	2.0	12.0	108.08*	0.17
≥150	244	10.0	5.1	21.6	154.99*	0.24
≥127	364	14.9	8.6	29.8	180.83*	0.29
≥101	620	25.4	17.4	44.3	194.21*	0.39
Min	26		26	27		
Max	863		863	601		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; LR = likelihood ratio; WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]$n = 2442$. [b]$n = 1715$. [c]$n = 727$.
*$p < .05$.

Tables 83 and 84 characterize TMT-B BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. Fisher's exact tests and chi-square tests of independence were statistically significant at nearly all TMT-B cutoffs in both samples, suggesting an association between TMT-B performance and diagnosis. Diagnosis accounted for 0.6% to 1.6% of the variance in TMT-B BR_{Fail} across cutoffs (very small to small effects) in the overall sample, and for 1.2% to 2.1% of the variance in TMT-B BR_{Fail} across cutoffs (small effects) in the valid sample.

In the overall sample, patients with moderate-severe TBI had the highest BR_{Fail} across TMT-B cutoffs, and patients with anxiety had the lowest BR_{Fail}. In the valid sample, the highest BR_{Fail} were most consistently observed in patients with moderate-severe TBI, neurological conditions, and severe mental illness, while the lowest rates were observed in patients with anxiety.

Table 83

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
TMT-B	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥304	49	2.0	2.7	5.7	2.2	0.8	0.3	3.4	2.8	4.1	2.7	FET*	.006
≥208	122	5.0	6.3	10.4	6.6	2.7	1.5	5.2	7.1	9.5	4.9	25.36*	.010
≥150	244	10.0	12.1	26.4	15.8	6.7	3.9	8.6	13.7	12.2	10.3	36.45*	.015
≥127	364	14.9	17.8	40.4	20.2	11.2	7.1	20.7	18.5	18.9	14.4	36.19*	.015
≥101	620	25.4	27.0	62.2	34.2	21.5	15.2	29.3	30.3	31.1	25.5	39.84*	.016
Min	26		26	26	33	27	26	31	35	29	28		
Max	863		601	500	863	600	315	370	540	448	600		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]n = 2442. [b]n = 552. [c]n = 194. [d]n = 228. [e]n = 526. [f]n = 336. [g]n = 58. [h]n = 211. [i]n = 74. [j]n = 263.
*p < .05.

Table 84

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
TMT-B	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥304	10	0.6	0.3	1.3	1.1	0.0	0.0	2.5	0.0	2.0	1.5	-	-
≥208	34	2.0	0.9	4.6	4.0	1.4	0.4	5.0	0.7	4.0	3.0	FET*	.012
≥150	87	5.1	3.6	8.5	9.8	3.5	1.2	7.5	6.6	6.0	6.9	25.82*	.015
≥127	147	8.6	6.6	15.0	13.8	6.8	3.5	17.5	9.6	8.0	10.3	30.67*	.018
≥101	298	17.4	13.8	27.5	25.9	15.2	9.7	22.5	19.9	16.0	20.2	36.64*	.021
Min	26		26	26	33	27	26	31	35	29	28		
Max	863		380	500	863	294	208	370	223	448	600		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]n = 1715. [b]n = 334. [c]n = 153. [d]n = 174. [e]n = 368. [f]n = 257. [g]n = 40. [h]n = 136. [i]n = 50. [j]n = 203.
*$p < .05$.

Tables 85 and 86 characterize TMT-B BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, BR_{Fail} varied significantly across age groups at all TMT-B cutoffs. Age accounted for 0.7% to 6.3% of the variance in TMT-B BR_{Fail} across cutoffs in the overall sample (small to medium effects), and for 1.0% to 7.0% of the variance in TMT-B BR_{Fail} in the valid sample (small to medium effects). In both samples, TMT-B BR_{Fail} increased with older age.

Table 85

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
TMT-B	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥304	49	2.0	0.6	1.2	1.5	2.5	5.6	17.24*	.007
≥208	122	5.0	2.2	2.1	4.0	6.7	16.8	59.28*	.024
≥150	244	10.0	3.4	5.8	8.3	14.2	27.3	89.46*	.037
≥127	364	14.9	6.6	10.4	12.4	20.0	38.5	106.35*	.044
≥101	620	25.4	11.6	18.7	22.4	33.1	58.7	153.36*	.063
Min	26		26	26	27	31	30		
Max	863		462	480	601	863	540		

Note. TMT-B = Trail Makimg Test (Reitan & Wolfson, 1985) Part B completion time in seconds; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]n = 2442. [b]n = 320. [c]n = 518. [d]n = 784. [e]n = 671. [f]n = 143.
*p < .05.

Table 86

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
TMT-B	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥304	10	0.6	0.4	0.0	0.0	1.3	2.3	-	-
≥208	34	2.0	1.2	0.8	1.1	3.4	5.7	FET*	.010
≥150	87	5.1	2.0	2.8	2.9	8.6	17.0	52.77*	.031
≥127	147	8.6	4.3	5.6	5.7	12.9	28.4	70.86*	.041
≥101	298	17.4	7.4	11.7	13.8	24.9	51.1	118.88*	.070
Min	26		26	26	29	31	30		
Max	863		462	298	296	863	500		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]n = 1715. [b]n = 256. [c]n = 359. [d]n = 543. [e]n = 466. [f]n = 88.
*p < .05.

Tables 87 and 88 characterize TMT-B BR_{Fail} across education levels in the overall and valid samples, respectively. Fisher's exact tests and chi-square tests of independence were statistically significant at nearly all TMT-B cutoffs in both samples. Education level accounted for 0.7% (small effect) to 5.1% (small to medium effect) of the variance in TMT-B BR_{Fail} across cutoffs in the overall sample, and for 0.8% (very small effect) to 3.8% (small to medium effect) of the variance in TMT-B BR_{Fail} across cutoffs in the valid sample. In both samples, the least educated group demonstrated the highest BR_{Fail} across TMT-B cutoffs, and BR_{Fail} decreased with higher levels of education. Patients with 12 and more years of education demonstrated generally comparable BR_{Fail}, although individuals with ≥17 years of education had larger BR_{Fail} at some TMT-B cutoffs than would be expected based on the trend.

Table 87

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
TMT-B	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥304	49	2.0	10.0	2.4	0.5	2.2	2.6	0.8	FET*	.023
≥208	122	5.0	18.5	6.8	2.7	4.4	3.9	3.3	66.38*	.007
≥150	244	10.0	35.4	13.4	5.9	7.9	6.5	12.4	124.41*	.051
≥127	364	14.9	43.1	19.1	9.2	13.1	13.8	18.2	116.01*	.048
≥101	620	25.4	59.2	30.5	19.8	21.7	22.0	28.1	106.94*	.044
Min	26		27	27	27	26	26	26		
Max	863		600	863	601	600	600	396		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]$n = 2442$. [b]$n = 130$. [c]$n = 591$. [d]$n = 963$. [e]$n = 405$. [f]$n = 232$. [g]$n = 121$.
*$p < .05$.

Table 88

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
TMT-B	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥304	10	0.6	3.0	1.3	0.1	0.3	0.6	0.0	-	-
≥208	34	2.0	7.5	2.7	1.7	1.0	1.7	1.1	FET	.008
≥150	87	5.1	22.4	6.9	3.9	2.4	3.4	5.7	51.87*	.030
≥127	147	8.6	28.4	11.7	6.5	6.1	8.4	6.9	44.40*	.026
≥101	298	17.4	50.7	22.0	14.7	12.8	15.6	13.8	66.46*	.038
Min	26		45	27	27	26	26	26		
Max	863		500	863	462	308	370	281		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time in seconds; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group; Max = maximum TMT-B score in the respective group.
[a]$n = 1715$. [b]$n = 67$. [c]$n = 377$. [d]$n = 709$. [e]$n = 296$. [f]$n = 179$. [g]$n = 87$.
*$p < .05$.

Tables 89 and 90 characterize TMT-B performance as a function of gender and English language status in the overall and valid samples, respectively. Chi-square tests of independence largely failed to reach statistical significance, suggesting no meaningful association between TMT-B BR_{Fail} and gender. Although a chi-square test was statistically significant at TMT-B ≥101 in the overall sample, this was likely driven by large sample at this cutoff. Likelihood ratios ranged from 0.95 to 1.29 in the overall sample and from 0.44 to 1.18 in the valid sample.

With regard to English language status, NSEs demonstrated lower BR_{Fail} than ESL patients at nearly all TMT-B cutoffs in both samples, resulting in statistically significant Fisher's exact tests and chi-square tests of independence. Likelihood ratios ranged from 0.26 to 0.45 in the overall sample and from 0.32 to 1.00 in the valid sample.

Table 89

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

	Overall Sample[a]		Gender				English			
TMT-B	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥304	49	2.0	1.9	2.0	0.01	0.95	1.5	5.4	20.52*	0.28
≥208	122	5.0	5.4	4.2	1.88	1.29	3.7	14.2	61.56*	0.26
≥150	244	10.0	10.5	9.2	0.97	1.14	7.7	26.4	101.22*	0.29
≥127	364	14.9	15.5	14.1	0.93	1.10	12.2	34.9	104.88*	0.35
≥101	620	25.4	26.9	23.1	4.43*	1.16	22.2	49.2	99.42*	0.45
Min	26		26	26			26	32		
Max	863		863	600			863	600		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language. Min = minimum TMT-B score in the respective group. Max = maximum TMT-B score in the respective group.
[a]$n = 2442$. [b]$n = 1490$. [c]$n = 952$. [d]$n = 2129$. [e]$n = 295$.
*$p < .05$.

Table 90

Cumulative Percentages of Participants Failing Various Trail Making Test Part B Completion Time Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

	Valid Sample[a]		Gender				English			
TMT-B	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥304	10	0.6	0.4	0.9	FET	0.44	0.6	0.6	FET	1.00
≥208	34	2.0	2.1	1.8	0.11	1.17	1.7	4.8	FET*	0.35
≥150	87	5.1	5.0	5.1	0.00	0.98	4.2	13.3	25.53*	0.32
≥127	147	8.6	8.5	8.8	0.05	0.97	7.5	19.4	26.41*	0.39
≥101	298	17.4	18.6	15.8	2.31	1.18	15.6	35.2	39.50*	0.44
Min	26		26	26			26	32		
Max	863		863	600			863	380		

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TMT-B score in the respective group. Max = maximum TMT-B score in the respective group.
[a]$n = 1715$. [b]$n = 1011$. [c]$n = 704$. [d]$n = 1539$. [e]$n = 165$.
*$p < .05$.

TMT A+B. Table 91 shows TMT A+B BR_{Fail}, at various cutoffs, as a function of WMT performance. Chi-square tests of independence were statistically significant across TMT A+B cutoffs. Participants who passed the WMT demonstrated lower BR_{Fail} than those who failed the WMT across TMT A+B cutoffs (LR = 0.12 – 0.37).

Table 91

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

TMT	Overall Sample[a]		WMT			
A+B	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≥391	49	2.0	0.6	5.2	57.24*	0.12
≥271	122	5.0	1.9	12.3	116.95*	0.15
≥202	247	10.1	4.8	22.6	178.50*	0.21
≥175	366	15.0	8.2	31.3	213.14*	0.26
≥143	610	25.0	16.7	44.8	214.47*	0.37
Min	38		38	49		
Max	928		928	902		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]n = 2441. [b]n = 1715. [c]n = 726.
*p < .05.

Tables 92 and 93 characterize TMT A+B BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. Fisher's exact tests and chi-square tests of independence were statistically significant across TMT A+B cutoffs in both samples. Diagnosis accounted for 0.5% to 1.7% of the variance in TMT A+B BR_{Fail} in the overall sample (small effects), and for 2.0% to 2.1% of the variance in TMT A+B BR_{Fail} in the valid sample (small effects). The highest BR_{Fail} were most consistently observed in patients with neurological conditions, although those with moderate-to-severe TBI also had elevated BR_{Fail} at more liberal cutoffs in the valid sample. Patients with anxiety demonstrated the lowest BR_{Fail}.

Table 92

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Overall Sample and in Specific Clinical Groups

TMT	Overall Sample[a]		TBI		Diagnostic Groups								
A+B	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥391	49	2.0	2.5	2.6	2.6	1.0	0.3	3.4	2.8	2.7	2.7	FET*	.005
≥271	122	5.0	6.3	6.7	6.6	2.9	0.9	5.2	8.5	8.1	4.9	28.83*	.012
≥202	247	10.1	12.9	10.4	17.1	7.0	3.9	10.3	12.8	12.2	9.1	39.18*	.016
≥175	366	15.0	17.4	18.1	22.4	11.6	8.0	17.2	17.5	18.9	13.7	33.70*	.014
≥143	610	25.0	25.7	30.6	32.9	21.3	14.6	29.3	31.8	29.7	25.9	40.89*	.017
Min	38		42	38	47	43	45	49	54	42	47		
Max	928		902	544	928	813	534	426	646	503	711		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]n = 2441. [b]n = 552. [c]n = 193. [d]n = 228. [e]n = 526. [f]n = 336. [g]n = 58. [h]n = 211. [i]n = 74. [j]n = 263.
*$p < .05$.

Table 93

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Valid Sample and in Specific Clinical Groups

TMT	Valid Sample[a]		TBI		Diagnostic Groups								
A+B	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≥391	10	0.6	0.3	2.0	1.1	0.0	0.0	2.5	0.0	2.0	1.0	-	-
≥271	33	1.9	0.3	3.9	4.6	1.4	0.0	5.0	1.5	4.0	3.0	-	-
≥202	82	4.8	3.3	7.2	12.1	3.0	1.6	7.5	4.4	6.0	5.9	33.76*	.020
≥175	141	8.2	5.7	15.0	16.1	6.2	3.9	12.5	8.8	6.0	7.9	36.11*	.021
≥143	286	16.7	13.2	26.8	24.1	13.9	9.7	22.5	19.9	12.0	20.2	36.79*	.021
Min	38		42	38	47	43	45	48	54	42	47		
Max	928		448	544	928	360	256	425	332	503	711		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]n = 1715. [b]n = 334. [c]n = 153. [d]n = 174. [e]n = 368. [f]n = 257. [g]n = 40. [h]n = 136. [i]n = 50. [j]n = 203.
*$p < .05$.

Tables 94 and 95 characterize TMT A+B BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, chi-square and Fisher's exact tests were statistically significant across TMT A+B cutoffs. Age accounted for 0.5% (very small effect) to 6.7% (medium effect) of the variance in TMT A+B BR_{Fail} across cutoffs in the overall sample, and 1.4% (small effect) to 6.8% (medium effect) of the variance in TMT A+B BR_{Fail} in the valid sample. At nearly all TMT A+B cutoffs, the lowest BR_{Fail} were observed in the youngest age group (16-19 years). BR_{Fail} increased with age, and the oldest age group (60-69 years) consistently had the highest BR_{Fail}.

Table 94

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Overall Sample and across Several Age Groups

TMT	Overall Sample[a]		Age (years)						
A+B	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥391	49	2.0	0.6	1.2	1.7	2.5	4.9	13.20*	.005
≥271	122	5.0	1.6	2.3	4.1	6.4	17.5	69.24*	.029
≥202	247	10.1	3.1	6.0	8.5	13.7	29.4	98.27*	.040
≥175	366	15.0	6.2	10.4	12.8	20.1	37.8	103.14*	.042
≥143	610	25.0	11.2	16.6	22.1	34.3	56.6	162.80*	.067
Min	38		38	42	43	49	47		
Max	928		514	646	813	928	619		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]$n = 2441$. [b]$n = 320$. [c]$n = 517$. [d]$n = 784$. [e]$n = 671$. [f]$n = 143$.
*$p < .05$.

Table 95

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Valid Sample and across Several Age Groups

TMT	Valid Sample[a]		Age (years)						
A+B	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≥391	10	0.6	0.4	0.0	0.0	1.3	2.3	-	-
≥271	33	1.9	0.8	0.8	0.9	3.2	6.8	FET*	.014
≥202	82	4.8	2.0	2.2	2.4	8.2	18.2	62.34*	.036
≥175	141	8.2	3.9	5.3	5.5	12.0	27.3	67.60*	.040
≥143	286	16.7	7.0	10.6	12.5	25.5	46.6	116.95*	.068
Min	38		38	42	43	50	47		
Max	928		514	371	350	928	544		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); FET = Fisher's exact test; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]n = 1715. [b]n = 256. [c]n = 359. [d]n = 543. [e]n = 466. [f]n = 88.
*p < .05.

Tables 96 and 97 characterize TMT A+B BR_{Fail} as a function of education level in the overall and valid samples, respectively. Chi-square and Fisher's exact tests were statistically significant at all TMT A+B cutoffs in the overall sample and nearly all cutoffs in the valid sample. Education level accounted for 1.9% (small effect) to 4.5% (small to medium effect) of the variance in TMT A+B BR_{Fail} across cutoffs in the overall sample, and for 0.8% (very small effect) to 3.0% (small to medium effect) of the variance in TMT A+B BR_{Fail} across cutoffs in the valid sample.

In both samples, the highest BR_{Fail} were consistently observed in the group with the lowest level of education (≤8 years), and BR_{Fail} decreased with higher levels of education. Patients with ≥12 years of education demonstrated relatively comparable BR_{Fail} across TMT A+B cutoffs, although BR_{Fail} in individuals with ≥17 years of education were higher at some cutoffs than would be expected based on the trend.

Table 96

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Overall Sample and across Several Education Levels

TMT	Overall Sample[a]		Education (years)							
A+B	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥391	49	2.0	9.2	2.2	0.5	2.2	3.0	0.8	FET*	.019
≥271	122	5.0	19.2	6.1	2.7	4.9	3.9	4.1	68.96*	.028
≥202	247	10.1	32.3	12.4	6.5	8.1	9.5	10.7	89.30*	.036
≥175	366	15.0	41.5	18.8	9.8	12.8	13.4	20.7	103.88*	.042
≥143	610	25.0	56.9	31.5	18.7	22.0	20.7	28.1	108.96*	.045
Min	38		66	47	43	38	42	45		
Max	928		902	928	798	652	813	500		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); FET = Fisher's exact test; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]$n = 2441$. [b]$n = 130$. [c]$n = 591$. [d]$n = 962$. [e]$n = 405$. [f]$n = 232$. [g]$n = 121$.
*$p < .05$.

Table 97

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Valid Sample and across Several Education Levels

TMT	Valid Sample[a]		Education (years)							
A+B	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≥391	10	0.6	3.0	1.3	0.1	0.0	1.1	0.0	-	-
≥271	33	1.9	7.5	2.1	1.6	1.0	1.7	2.3	FET	.008
≥202	82	4.8	19.4	5.0	4.5	2.0	4.5	4.6	36.61*	.021
≥175	141	8.2	25.4	10.6	6.8	5.1	7.3	8.0	35.30*	.020
≥143	286	16.7	44.8	21.5	13.7	13.5	14.5	13.8	52.17*	.030
Min	38		66	47	43	38	42	45		
Max	928		544	928	514	367	448	371		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); FET = Fisher's exact test; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]$n = 1715$. [b]$n = 67$. [c]$n = 377$. [d]$n = 709$. [e]$n = 296$. [f]$n = 179$. [g]$n = 87$.
*$p < .05$.

Tables 98 and 99 characterize TMT A+B BR_{Fail} as a function of gender and English language status in the overall and valid samples, respectively. In both samples, chi-square tests and Fisher's exact tests generally failed to reach statistical significance, suggesting no relationship between gender and TMT A+B BR_{Fail}. Although a chi-square test was significant at TMT A+B ≥143 in the valid sample, χ^2 (1, N = 2441) = 4.98, p = .026, this finding is likely due to the increase in sample size at that cutoff. Likelihood ratios ranged from 1.05 to 1.20 in the overall sample and from 0.71 to 1.31 in the valid sample.

With regard to English language status, chi-square tests were statistically significant at all TMT A+B cutoffs in the overall sample and nearly all TMT A+B cutoffs in the valid sample. NSEs demonstrated lower BR_{Fail} than ESL patients across all TMT A+B cutoffs in both samples. Likelihood ratios ranged from 0.25 to 0.44 in the overall sample, and from 0.31 to 0.45 in the valid sample.

Table 98

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

TMT	Overall Sample[a]		Gender				English			
A+B	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥391	49	2.0	2.0	1.9	0.05	1.05	1.5	5.8	24.74*	0.26
≥271	122	5.0	5.3	4.4	0.99	1.20	3.6	14.6	66.09*	0.25
≥202	247	10.1	10.3	9.7	0.30	1.06	7.8	24.4	99.40*	0.32
≥175	366	15.0	15.9	13.7	2.32	1.16	11.9	37.6	133.65*	0.32
≥143	610	25.0	26.6	22.6	4.98*	1.18	21.7	49.5	106.38*	0.44
Min	38		38	42			38	50		
Max	928		928	798			928	813		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]n = 2441. [b]n = 1489. [c]n = 952. [d]n = 2128. [e]n = 295.
*p < .05.

Table 99

Cumulative Percentages of Participants Failing Various Trail Making Test A+B Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

TMT	Valid Sample[a]		Gender				English			
A+B	*f*	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≥391	10	0.6	0.5	0.7	FET	0.71	0.5	1.2	FET	0.42
≥271	33	1.9	2.1	1.6	0.60	1.31	1.6	4.2	FET*	0.38
≥202	82	4.8	4.6	5.0	0.10	0.92	4.1	11.5	17.92*	0.36
≥175	141	8.2	8.4	7.8	0.20	1.08	6.8	21.8	44.82*	0.31
≥143	286	16.7	18.0	14.8	3.12	1.22	15.0	33.3	35.82*	0.45
Min	38		38	42			38	50		
Max	928		928	711			928	448		

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TMT A+B score in the respective group; Max = maximum TMT A+B score in the respective group.
[a]$n = 1715$. [b]$n = 1011$. [c]$n = 704$. [d]$n = 1539$. [e]$n = 165$.
*$p < .05$.

TMT-B/A. Table 100 shows TMT-B/A BR_{Fail}, at various cutoffs, as a function of WMT performance. BR_{Fail} were generally comparable between those who passed the WMT and those who failed the WMT across TMT-B/A cutoffs, and chi-square tests failed to reach statistical significance. Likelihood ratios ranged from 1.04 to 1.35.

Table 100

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

TMT-B/A	Overall Sample[a]		WMT			
	f	%	Pass[b]	Fail[c]	χ^2	LR
≤1.26	51	2.1	2.3	1.7	0.96	1.35
≤1.40	122	5.0	5.2	4.5	0.30	1.16
≤1.57	244	10.0	10.3	9.2	0.68	1.12
≤1.71	373	15.3	15.5	14.9	0.13	1.04
≤1.91	608	24.9	25.4	23.8	0.64	1.07
Min	0.24		0.47	0.24		
Max	15.41		13.28	15.41		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio. Min = minimum TMT-B/A score in the respective group. Max = maximum TMT-B/A score in the respective group.
[a]n = 2441. [b]n = 1715. [c]n = 726.
*p < .05.

Tables 101 and 102 characterize TMT-B/A BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. In both samples, chi-square and Fisher's exact tests failed to reach statistical significance. Effect sizes indicated that diagnosis accounted for only 0.2% to 0.4% of the variance in TMT-B/A BR_{Fail} in the overall sample (very small effects) and for 0.3% to 0.5% of the variance in TMT-B/A BR_{Fail} in the valid sample (very small effects).

Table 101

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Overall Sample and in Specific Clinical Groups

TMT	Overall Sample[a]		TBI		Diagnostic Groups								
B/A	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤1.26	51	2.1	1.8	2.1	2.6	2.3	1.6	5.2	2.4	5.4	1.1	FET	.004
≤1.40	122	5.0	5.1	7.3	4.8	4.4	4.5	10.3	4.7	5.4	4.2	5.86	.002
≤1.57	244	10.0	7.8	11.9	10.1	10.3	9.8	17.2	10.0	12.2	10.6	7.73	.003
≤1.71	373	15.3	12.9	16.1	18.9	16.0	14.3	19.0	13.7	14.9	17.1	6.97	.003
≤1.91	608	24.9	23.9	25.4	28.9	25.1	24.4	32.8	23.7	21.6	23.6	5.12	.002
Min	0.24		.93	1.19	1.05	0.24	.87	1.06	0.47	1.02	1.14		
Max	15.41		15.41	11.36	13.28	6.77	7.44	6.73	7.87	9.23	11.92		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT-B/A score in the respective group; Max = maximum TMT-B/A score in the respective group.
[a]n = 2441. [b]n = 552. [c]n = 193. [d]n = 228. [e]n = 526. [f]n = 336. [g]n = 58. [h]n = 211. [i]n = 74. [j]n = 263.
*$p < .05$.

Table 102

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Valid Sample and in Specific Clinical Groups

TMT	Valid Sample[a]		TBI		Diagnostic Groups								
B/A	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤1.26	39	2.3	2.1	2.6	2.9	2.2	1.6	5.0	2.2	6.0	1.5	FET	.003
≤1.40	86	5.0	6.0	8.5	4.6	4.3	3.9	12.5	4.4	6.0	3.9	9.27	.005
≤1.57	177	10.3	8.7	11.8	10.9	10.3	8.2	17.5	11.8	10.0	11.8	5.70	.003
≤1.71	266	15.5	12.9	15.0	19.5	16.3	13.2	20.0	14.7	14.0	17.7	6.70	.004
≤1.91	436	25.4	24.0	24.2	31.6	24.7	24.1	32.5	26.5	22.0	24.6	5.86	.003
Min	0.47		0.93	1.19	1.05	1.00	0.87	1.06	0.47	1.02	1.14		
Max	13.28		6.79	11.36	13.28	5.33	5.64	6.73	7.87	8.15	8.88		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum TMT-B/A score in the respective group; Max = maximum TMT-B/A score in the respective group.
[a]n = 1715. [b]n = 334. [c]n = 153. [d]n = 174. [e]n = 418. [f]n = 257. [g]n = 40. [h]n = 136. [i]n = 50. [j]n = 203.
*$p < .05$.

Tables 103 and 104 characterize TMT-B/A BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, chi-square and Fisher's exact tests failed to reach statistical significance. Across cutoffs, age accounted for only 0.1% to 0.3% of the variance in TMT-B/A BR_{Fail} in both samples (small effects).

Table 103

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Overall Sample and across Several Age Groups

TMT	Overall Sample[a]		Age (years)						
B/A	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤1.26	51	2.1	1.3	3.1	1.9	2.2	0.7	5.18	.002
≤1.40	122	5.0	4.4	5.8	5.5	4.8	2.1	3.94	.002
≤1.57	244	10.0	9.7	10.8	10.5	9.8	6.3	2.81	.001
≤1.71	373	15.3	15.3	16.2	16.6	14.5	9.1	5.97	.003
≤1.91	608	24.9	25.6	26.1	26.0	23.8	18.9	4.18	.002
Min	0.24		1.11	0.87	0.24	0.47	1.20		
Max	15.41		8.88	8.24	15.41	13.28	11.36		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); Min = minimum TMT-B/A score in the respective group. Max = maximum TMT-B/A score in the respective group.
[a]$n = 2441$. [b]$n = 320$. [c]$n = 571$. [d]$n = 784$. [e]$n = 671$. [f]$n = 143$.
*$p < .05$.

Table 104

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Valid Sample and across Several Age Groups

TMT-	Valid Sample[a]		Age (years)						
B/A	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤1.26	39	2.3	1.2	3.6	2.2	2.1	1.1	4.88	.003
≤1.40	86	5.0	3.9	6.4	6.3	4.3	2.3	5.57	.003
≤1.57	177	10.3	9.4	11.7	10.9	9.7	8.0	1.91	.001
≤1.71	266	15.5	14.5	16.4	16.8	14.8	10.2	3.15	.002
≤1.91	436	25.4	25.8	26.2	26.5	24.2	20.5	1.96	.001
Min	0.47		1.11	0.87	1.02	0.47	1.20		
Max	13.28		8.88	7.10	7.87	13.28	11.36		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); Min = minimum TMT-B/A score in the respective group. Max = maximum TMT-B/A score in the respective group.
[a]$n = 1715$. [b]$n = 256$. [c]$n = 359$. [d]$n = 543$. [e]$n = 466$. [f]$n = 88$.
*$p < .05$.

Tables 105 and 106 characterize TMT-B/A BR_{Fail} as a function of education level in the overall and valid samples, respectively. Chi-square and Fisher's exact tests reached significance at more liberal cutoffs in both samples, likely due to the increase in sample size at those cutoffs. However, education level accounted for only 0.1% to 1.0% of the variance in TMT-B/A BR_{Fail} in the overall sample (small effects), and for 0.2% to 1.1% of the variance in TMT-B/A BR_{Fail} in the valid sample (very small to small effects). Nevertheless, in both samples, the lowest BR_{Fail} were observed in the least educated group (≤8 years of formal education), and BR_{Fail} tended to increase with higher levels of education.

Table 105

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Overall Sample and across Several Education Levels

TMT B/A	Overall Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤1.26	51	2.1	1.5	1.9	1.9	2.2	3.4	2.5	FET	.001
≤1.40	122	5.0	2.3	4.4	4.5	5.7	7.8	7.4	9.39	.004
≤1.57	244	10.0	3.8	8.3	10.5	9.4	15.5	12.4	16.45*	.007
≤1.71	373	15.3	9.2	13.2	15.7	14.6	23.3	15.7	17.42*	.007
≤1.91	608	24.9	16.2	20.6	25.6	26.2	34.9	26.4	24.22*	.010
Min	0.24		0.24	0.87	0.47	1.05	1.10	1.16		
Max	15.41		11.36	13.28	15.41	11.92	8.06	8.24		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); FET = Fisher's exact test; Min = minimum B/A score in the respective group; Max = maximum B/A score in the respective group.
[a]$n = 2441$. [b]$n = 130$. [c]$n = 591$. [d]$n = 962$. [e]$n = 405$. [f]$n = 232$. [g]$n = 121$.
*$p < .05$.

Table 106

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Valid Sample and across Several Education Levels

TMT-B/A	Valid Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤1.26	39	2.3	1.5	2.4	1.8	2.0	3.9	3.4	FET	.002
≤1.40	86	5.0	1.5	4.8	4.5	5.7	8.4	6.9	7.72	.004
≤1.57	177	10.3	4.5	8.2	10.6	10.1	15.6	11.5	9.93	.006
≤1.71	266	15.5	10.4	12.7	15.2	14.9	24.6	16.1	14.97*	.009
≤1.91	436	25.4	14.9	19.9	25.4	28.4	34.6	27.6	19.59*	.011
Min	0.47		1.15	0.87	0.47	1.19	1.10	1.16		
Max	13.28		11.36	13.28	8.88	6.52	6.73	5.39		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); FET = Fisher's exact test; Min = minimum B/A score in the respective group; Max = maximum B/A score in the respective group.
[a]n = 1715. [b]n = 67. [c]n = 377. [d]n = 709. [e]n = 296. [f]n = 179. [g]n = 87.
*p < .05.

Tables 107 and 108 characterize TMT-B/A BR_{Fail} across gender and English language status in the overall and valid samples, respectively. Chi-square tests largely failed to reach statistical significance across TMT-B/A cutoffs in either sample, suggesting no meaningful relationship between TMT-B/A BR_{Fail} and gender. Although chi-square tests were statistically significant at TMT-B/A ≤1.91 in both samples, these findings were likely due to the increased sample size. Likelihood ratios ranged from 0.84 to 0.91 in the overall sample, and from 0.82 to 1.25 in the valid sample.

TMT-B/A BR_{Fail} were unrelated to English language status across cutoffs in either sample. Likelihood ratios ranged from 0.65 to 1.00 in the overall sample, and from 0.93 to 0.99 in the valid sample.

Table 107

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Overall Sample and as a Function of Gender and English Language Status

TMT-B/A	Overall Sample[a]		Gender				English			
	f	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤1.26	51	2.1	2.0	2.2	0.10	0.91	2.0	3.1	1.46	0.65
≤1.40	122	5.0	4.7	5.5	1.00	0.85	4.7	6.8	2.51	0.69
≤1.57	244	10.0	9.5	10.8	1.18	0.88	9.8	11.5	0.83	0.85
≤1.71	373	15.3	14.4	16.7	2.44	0.86	15.2	16.6	0.41	0.92
≤1.91	608	24.9	23.2	27.5	5.70*	0.84	25.0	25.1	0.00	1.00
Min	0.24		0.24	0.47			0.24	0.47		
Max	15.41		15.41	11.92			15.41	9.69		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; Min = minimum TMT-B/A score in the respective group; Max = maximum TMT-B/A score in the respective group.
[a]n = 2441. [b]n = 1489. [c]n = 952. [d]n = 2128. [e]n = 295.
*p < .05.

Table 108

Cumulative Percentages of Participants Failing Various Trail Making Test B/A Cutoffs in the Valid Sample and as a Function of Gender and English Language Status

TMT-B/A	Valid Sample[a]		Gender				English			
	f	%	Male[b]	Female[c]	χ^2	LR	NSE[d]	ESL[e]	χ^2	LR
≤1.26	39	2.3	2.5	2.0	0.44	1.25	2.3	2.4	FET	0.96
≤1.40	86	5.0	5.3	5.0	0.03	1.06	5.1	5.5	0.06	0.93
≤1.57	177	10.3	10.0	10.8	0.29	0.93	10.3	10.9	0.07	0.94
≤1.71	266	15.5	14.5	16.8	1.57	0.86	15.4	16.4	0.11	0.94
≤1.91	436	25.4	23.3	28.3	5.32*	0.82	25.3	25.5	0.00	0.99
Min	0.47		0.87	0.47			0.87	0.47		
Max	13.28		13.28	11.36			13.28	6.00		

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); LR = likelihood ratio; NSE = native speakers of English; ESL = English as a second language; FET = Fisher's exact test; Min = minimum TMT-B/A score in the respective group; Max = maximum TMT-B/A score in the respective group.
[a]n = 1715. [b]n = 1101. [c]n = 704. [d]n = 1539. [e]n = 165.
*p < .05.

FTT-DH$_M$. Table 109 shows FTT-DH$_M$ BR$_{Fail}$, at various cutoffs, as a function of WMT performance. Participants who passed the WMT demonstrated significantly lower BR$_{Fail}$ than those who failed the WMT across FTT-DH$_M$ cutoffs (LR = 0.09 - 0.36).

Table 109

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
FTT-DH$_M$	*f*	%	Pass[d]	Fail[e]	χ^2	LR
≤18.0	11	2.0	0.5	5.6	FET*	0.09
≤27.6	27	4.9	2.6	10.5	15.54*	0.25
≤36.5	56	10.1	6.1	19.1	23.33*	0.32
≤42.0	82	14.8	9.0	29.0	36.50*	0.31
≤45.0	139	25.1	16.6	45.7	51.39*	0.36
Min	4.4		17.4	4.4		
Max	71.2		71.2	65.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH$_M$ = average number of taps achieved with the dominant hand (males only); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; FET = Fisher's exact test; LR = likelihood ratio; Min = minimum FTT-DH$_M$ score in the respective group; Max = maximum FTT-DH$_M$ score in the respective group.
[a]n = 553. [b]n = 391. [c]n = 162.
*p < .05.

Tables 110 and 111 characterize FTT-DH$_M$ BR$_{Fail}$ as a function of diagnosis in the overall and valid samples, respectively. The only Fisher's exact test conducted in the overall sample (i.e., at FTT-DH$_M$ ≤45) failed to reach statistical significance, and the corresponding effect size indicated that diagnosis accounted for only 1.9% of the variance in BR$_{Fail}$ at that cutoff (small effect). Chi-square or Fisher's exact tests were not conducted in the valid sample. No comments are made about trends in BR$_{Fail}$ due to the small sizes of several clinical groups.

Table 110

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Dominant Hand Cutoffs in the Overall Sample and in Specific Clinical Groups

FTT-DH_M	Overall Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18.0	11	2.0	2.4	3.0	1.6	4.2	0.0	0.0	0.0	0.0	0.0	-	-
≤27.6	27	4.9	4.4	6.9	8.2	8.3	0.0	0.0	0.0	0.0	2.8	-	-
≤36.5	56	10.1	8.0	14.9	13.1	12.5	0.0	0.0	20.0	20.0	7.0	-	-
≤42.0	82	14.8	12.0	17.8	19.7	25.0	0.0	8.3	40.0	26.7	9.9	-	-
≤45.0	139	25.1	22.7	27.7	24.6	33.3	12.5	16.7	50.0	46.7	22.5	FET	.019
Min	4.4		4.4	9.0	14.6	17.4	45.0	38.8	30.2	29.2	18.2		
Max	71.2		71.0	65.0	63.2	71.2	67.2	59.2	56.2	65.2	68.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]n = 553. [b]n = 251. [c]n = 101. [d]n = 61. [e]n = 24. [f]n = 8. [g]n = 12. [h]n = 10. [i]n = 15. [j]n = 71.
*p < .05.

Table 111

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Dominant Hand Cutoffs in the Valid Sample and in Specific Clinical Groups

FTT-DH_M	Valid Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18.0	2	0.5	0.0	1.3	0.0	6.3	0.0	0.0	0.0	0.0	0.0	-	-
≤27.6	10	2.6	1.2	2.7	6.5	12.5	0.0	0.0	0.0	0.0	1.8	-	-
≤36.5	24	6.1	1.9	10.7	8.7	18.8	0.0	0.0	16.7	23.1	3.6	-	-
≤42.0	35	9.0	3.7	12.0	13.0	25.0	0.0	0.0	50.0	30.8	3.6	-	-
≤45.0	65	16.6	10.5	20.0	19.6	25.0	0.0	10.0	50.0	46.2	17.9	-	-
Min	17.4		24.8	18.0	18.8	17.4	49.4	42.8	32.2	29.2	24.2		
Max	71.2		71.0	65.0	63.2	71.2	67.2	59.2	56.2	65.2	68.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]n = 391. [b]n = 162. [c]n = 75. [d]n = 46. [e]n = 16. [f]n = 7. [g]n = 10. [h]n = 6. [i]n = 13. [j]n = 56.
*p < .05.

Tables 112 and 113 characterize FTT-DH_M BR_{Fail} as a function of age in the overall and valid samples, respectively. In the overall sample, chi-square and Fisher's exact tests were statistically significant at nearly all cutoffs examined, and age accounted for 1.4% to 7.1% of the variance in FTT-DH_M BR_{Fail} across cutoffs (small to medium effects). BR_{Fail} tended to increase with age, and the oldest group (60-69 years) consistently demonstrated the highest BR_{Fail}. In the valid sample, chi-square tests reached statistical significance at more liberal cutoffs (likely due to the small sample sizes at more conservative cutoffs), and age accounted for 2.3% (small effect) to 3.9% (small to medium effect) of the variance in BR_{Fail}. Again, BR_{Fail} tended to increase with age. The oldest age group was, however, relatively small in size (n = 18).

Table 112

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand Cutoffs in the Overall Sample and across Several Age Groups

FTT-DH_M	Overall Sample[a]		Age (years)						
	f	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18.0	11	2.0	0.9	0.8	1.2	4.1	6.5	FET	.014
≤27.6	27	4.9	1.9	3.1	3.7	9.0	12.9	12.14*	.022
≤36.5	56	10.1	7.5	5.5	9.2	13.1	29.0	17.51*	.032
≤42.0	82	14.8	9.3	9.4	11.0	21.3	48.4	39.23*	.071
≤45.0	139	25.1	17.8	17.3	25.2	32.8	51.6	22.59*	.041
Min	4.4		18.0	13.4	14.5	4.4	9.0		
Max	71.2		71.0	67.8	69.2	71.2	58.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group. Max = maximum FTT-DH_M score in the respective group.
[a]n = 553. [b]n = 107. [c]n = 127. [d]n = 163. [e]n = 122. [f]n = 31.
*p < .05.

Table 113

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand cutoffs in the Valid Sample and across Several Age Groups

FTT-DH_M	Valid Sample[a]		Age (years)						
	f	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18.0	2	0.5	1.2	0.0	0.0	1.2	0.0	-	-
≤27.6	10	2.6	1.2	2.2	2.7	4.8	0.0	-	-
≤36.5	24	6.1	3.7	2.2	7.1	9.6	16.7	8.80	.023
≤42.0	35	9.0	3.7	5.4	8.8	14.5	27.8	15.12*	.039
≤45.0	65	16.6	8.5	9.7	20.4	24.1	33.3	15.15*	.039
Min	17.4		18.0	18.8	20.4	17.4	28.0		
Max	71.2		71.0	67.8	69.2	71.2	58.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); Min = minimum FTT-DH_M score in the respective group. Max = maximum FTT-DH_M score in the respective group.
[a]$n = 391$. [b]$n = 82$. [c]$n = 93$. [d]$n = 113$. [e]$n = 83$. [f]$n = 18$.
*$p < .05$.

Tables 114 and 115 characterize FTT-DH_M BR_{Fail} as a function of education level in the overall and valid samples, respectively. In the overall sample, Fisher's exact tests failed to reach statistical significance where conducted, and education level accounted for only 1.3% to 1.7% of the variance in FTT-DH_M BR_{Fail} across cutoffs (small effects). BR_{Fail} tended to be highest in the least educated groups and decrease with higher levels of education. However, the most educated group (≥17 years of education) demonstrated higher BR_{Fail} than would be expected based on the trend. At several cutoffs, BR_{Fail} among those with ≥17 years of education exceeded those of the least educated groups as well.

In the valid sample, cell sizes at more conservative cutoffs (e.g., ≤18.0, ≤27.6, and ≤36.5) were quite small, resulting in potentially unreliable estimates of BR_{Fail} across education groups at these cutoffs. A chi-square test conducted at FTT-DH_M ≤45.0 was statistically significant, χ^2 (5, $N = 391$) = 12.97, $p = .024$, and education accounted for 3.3% of the variance at this cutoff (small to medium effect). A trend similar to the overall sample was observed: BR_{Fail} tended to

decrease with higher levels of education, with a spike in BR_{Fail} in those with ≥17 years of education.

Table 114

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand cutoffs in the Overall Sample and across Several Education Levels

FTT-DH_M	Overall Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18.0	11	2.0	5.4	0.6	2.2	0.0	4.4	6.9	-	-
≤27.6	27	4.9	10.8	2.9	5.6	2.2	6.7	10.3	FET	.015
≤36.5	56	10.1	10.8	8.8	12.3	5.4	8.9	20.7	FET	.013
≤42.0	82	14.8	16.2	16.5	14.0	9.7	8.9	31.0	FET	.016
≤45.0	139	25.1	32.4	28.8	23.5	20.4	13.3	37.9	FET	.017
Min	4.4		4.4	9.0	13.4	18.8	14.6	14.5		
Max	71.2		62.8	65.0	71.2	65.2	68.2	69.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]*n* = 553. [b]*n* = 37. [c]*n* = 170. [d]*n* = 179. [e]*n* = 93. [f]*n* = 45. [g]*n* = 29.
*$p < .05$.

Table 115

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand Cutoffs in the Valid Sample and across Several Education Levels

FTT-DH_M	Valid Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18.0	2	0.5	0.0	0.0	0.7	0.0	0.0	4.5	-	-
≤27.6	10	2.6	5.9	0.9	3.6	1.4	0.0	9.1	-	-
≤36.5	24	6.1	5.9	4.7	8.6	1.4	0.0	22.7	-	-
≤42.0	35	9.0	11.8	9.4	10.7	2.8	0.0	27.3	-	-
≤45.0	65	16.6	17.6	17.9	18.6	11.3	2.9	36.4	12.97*	.033
Min	17.4		20.4	22.2	18.0	18.8	43.1	17.4		
Max	71.2		56.8	65.0	71.2	65.2	68.2	69.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]*n* = 391. [b]*n* = 17. [c]*n* = 106. [d]*n* = 140. [e]*n* = 71. [f]*n* = 35. [g]*n* = 22.
*$p < .05$.

Tables 116 and 117 characterize FTT-DH_M BR_{Fail} as a function of English language status in the overall and valid samples, respectively. In the overall sample, NSEs demonstrated lower BR_{Fail} than ESL patients across FTT-DH_M cutoffs, although Fisher's exact tests and chi-square tests of independence only reached statistical significance at FTT-DH_M ≤18.0 and ≤45.0. Likelihood ratios ranged from 0.14 to 0.64 across cutoffs. In the valid sample, sample sizes were quite small at many cutoffs, and Fisher's exact tests and chi-square tests consistently failed to reach statistical significance across cutoffs. Likelihood ratios ranged from 0.14 to 1.05.

Table 116

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand Cutoffs in the Overall Sample and as a Function of English Language Status

	Overall Sample[a]		English			
FTT-DH_M	*f*	%	NSE[b]	ESL[c]	χ^2	LR
≤18.0	11	2.0	1.1	8.1	FET*	0.14
≤27.6	27	4.9	4.0	9.5	FET	0.42
≤36.5	56	10.1	9.1	16.2	3.51	0.56
≤42.0	82	14.8	13.8	21.6	3.06	0.64
≤45.0	139	25.1	23.0	39.2	8.92*	0.59
Min	4.4		9.0	4.4		
Max	71.2		71.2	71.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]n = 553. [b]n = 470. [c]n = 74.
*p < .05.

Table 117

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Dominant Hand Cutoffs in the Valid Sample and as a Function of English Language Status

	Valid Sample[a]		English			
FTT-DH_M	*f*	%	NSE[b]	ESL[c]	χ^2	LR
≤18.0	2	0.5	0.3	2.2	FET	0.14
≤27.6	10	2.6	2.3	2.2	FET	1.05
≤36.5	24	6.1	5.6	8.9	FET	0.63
≤42.0	35	9.0	8.5	11.1	FET	0.77
≤45.0	65	16.6	15.5	24.4	2.28	0.64
Min	17.4		17.4	18.0		
Max	71.2		71.2	71.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-DH_M score in the respective group; Max = maximum FTT-DH_M score in the respective group.
[a]$n = 391$. [b]$n = 341$. [c]$n = 45$.
*$p < .05$.

FTT-NDH$_M$. Table 118 shows FTT-NDH$_M$ BR$_{Fail}$ as a function of WMT performance. Chi-square and Fisher's exact tests were statistically significant across FTT-NDH$_M$ cutoffs, suggesting that FTT-NDH$_M$ BR$_{Fail}$ were related to WMT classification. Those who passed the WMT consistently demonstrated lower BR$_{Fail}$ than those who failed the WMT across FTT-NDH$_M$ cutoffs. Likelihood ratios ranged from 0.09 to 0.45.

Table 118

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

FTT-NDH$_M$	Overall Sample[a]		WMT			
	f	%	Pass[b]	Fail[c]	χ^2	LR
≤18.4	11	2.0	0.5	5.7	FET*	0.09
≤27.0	28	5.0	3.1	9.4	9.57*	0.33
≤35.2	54	9.9	6.2	18.9	20.19*	0.33
≤38.4	82	15.0	10.9	25.2	17.96*	0.43
≤41.6	140	25.7	18.9	42.1	31.83*	0.45
Min	5.0		15.0	5.0		
Max	68.0		68.0	61.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH$_M$ = average number of taps achieved with the non-dominant hand (males only); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-NDH$_M$ score in the respective group; Max = maximum FTT-NDH$_M$ score in the respective group.
[a]n = 545. [b]n = 386. [c]n = 159.
*p < .05.

Tables 119 and 120 characterize FTT-NDH_M BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. Chi-square and Fisher's exact tests were not conducted at any of the cutoffs, and trends in BR_{Fail} were not identified given the small sizes of several clinical groups.

Table 119

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Overall Sample and in Specific Clinical Groups

FTT-NDH_M	Overall Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18.4	11	2.0	2.4	2.0	3.4	4.2	0.0	0.0	0.0	0.0	0.0	-	-
≤27.0	28	5.0	3.6	8.1	6.8	8.3	0.0	0.0	0.0	20.0	1.4	-	-
≤35.2	54	9.9	8.0	15.2	11.9	12.5	12.5	0.0	0.0	26.7	5.7	-	-
≤38.4	82	15.0	12.0	22.2	16.9	25.0	12.5	0.0	22.2	33.3	8.6	-	-
≤41.6	140	25.7	22.5	29.3	32.2	33.0	25.0	0.0	44.4	46.7	21.4	-	-
Min	5.0		5.0	17.8	13	16.2	34.4	41.8	36.0	19.2	18.8		
Max	68.0		67.4	60.6	60.2	63.2	59.4	62.4	53.8	64.6	68.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]$n = 545$. [b]$n = 249$. [c]$n = 99$. [d]$n = 59$. [e]$n = 24$. [f]$n = 8$. [g]$n = 12$. [h]$n = 9$. [i]$n = 15$. [j]$n = 70$.
*$p < .05$.

Table 120

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Valid Sample and in Specific Clinical Groups

FTT-NDH_M	Valid Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤18.4	2	0.5	0.0	0.0	2.2	6.3	0.0	0.0	0.0	0.0	0.0	-	-
≤27.0	12	3.1	0.6	5.4	4.4	12.5	0.0	0.0	0.0	23.1	0.0	-	-
≤35.2	24	6.2	3.1	8.1	8.9	18.8	14.3	0.0	0.0	30.8	1.8	-	-
≤38.4	42	10.9	5.0	16.2	11.1	37.5	14.3	0.0	40.0	38.5	5.5	-	-
≤41.6	73	18.9	11.8	24.3	24.4	37.5	28.6	0.0	40.0	46.5	16.4	-	-
Min	15.0		23.0	19.3	15.0	16.2	34.4	41.8	36.0	19.2	34.2		
Max	68.0		67.4	60.6	60.2	63.2	59.4	62.4	53.8	64.6	68.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]n = 386. [b]n = 161. [c]n = 74. [d]n = 45. [e]n = 16. [f]n = 7. [g]n = 10. [h]n = 5. [i]n = 13. [j]n = 55.
*p < .05.

Tables 121 and 122 characterize FTT-NDH_M BR_{Fail} as a function of age in the overall and valid samples, respectively. Chi-square and Fisher's exact tests were statistically significant at nearly all FTT-NDH_M cutoffs in either sample. Age accounted for 3.1% (small to medium effect) to 6.4% (medium effect) of the variance in FTT-NDH_M BR_{Fail} across cutoffs in the overall sample, and for 2.1% (small effect) to 6.9% (medium effect) of the variance in FTT-NDH_M BR_{Fail} across cutoffs in the valid sample. Visual inspection revealed that BR_{Fail} tended to increase with older age in both samples. The youngest age group (16-29 years) demonstrated slightly higher BR_{Fail} at some of the cutoffs than would be expected based on the trend. Notably, the oldest group (60-69 years) was small in size in the valid sample (n = 18).

Table 121

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
FTT-NDH_M	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18.4	11	2.0	0.0	1.6	1.3	5.0	3.3	-	-
≤27.0	28	5.0	1.9	2.4	3.8	9.2	16.7	17.55*	.032
≤35.2	54	9.9	7.5	4.0	10.0	13.3	26.7	16.85*	.031
≤38.4	82	15.0	9.4	6.3	16.3	22.5	33.3	23.44*	.043
≤41.6	140	25.7	17.6	15.9	23.1	37.5	56.7	34.29*	.064
Min	5.0		19.3	12.8	13.0	5.0	17.8		
Max	68.0		63.2	67.4	64.8	68.0	52.6		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]n = 545. [b]n = 106. [c]n = 126. [d]n = 160. [e]n = 120. [f]n = 30.
*p < .05.

Table 122

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
FTT-NDH_M	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤18.4	2	0.5	0.0	0.0	0.0	2.4	0.0	-	-
≤27.0	12	3.1	2.5	1.1	1.8	6.1	11.1	FET	.021
≤35.2	24	6.2	2.5	1.1	7.2	9.8	22.2	14.32*	.037
≤38.4	42	10.9	4.9	3.3	12.6	19.5	27.8	20.31*	.053
≤41.6	73	18.9	12.3	7.6	18.9	31.7	44.4	26.52*	.069
Min	15.0		19.3	20.6	19.2	15.0	25.6		
Max	68.0		63.2	67.4	64.8	68.0	51.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]n = 386. [b]n = 81. [c]n = 92. [d]n = 111. [e]n = 82. [f]n = 18.
*p < .05.

Tables 123 and 124 characterize FTT-NDH_M BR_{Fail} as a function of education level in the overall and valid samples, respectively. In the overall sample, education level accounted for 1.6% (small effect) to 4.9% (small to medium effect) of the variance in FTT-NDHM BR_{Fail} across cutoffs, and Fisher's exact tests reached statistical significance at FTT-NDH_M ≤38.4 and ≤41.6. The group with the least education (≤8 years) tended to have the highest BR_{Fail}, and BR_{Fail} generally decreased with higher levels of education. The most educated group (≥17 years) did, however, exhibit considerably higher BR_{Fail} than would be expected based on the trends at a number of FTT-NDH_M cutoffs.

In the valid sample, sample sizes were quite small at many of the cutoffs examined. Nonetheless, a pattern similar to that of the overall sample was observed at more liberal cutoffs. Chi-square and Fisher's exact tests reached statistical significance at FTT-NDH_M ≤38.4 (p = .027, Fisher's exact test) and ≤41.6, χ^2 (5, N = 386) = 17.24, p = .004. Education accounted for 3.1% and 4.5% of the variance in BR_{Fail} at these cutoffs, respectively (small to medium effects).

Table 123

Cumulative Percentages of Male Participants Failing Finger Tapping Test Non-dominant Hand Cutoffs in the Overall Sample and across Several Education Levels

FTT-NDH_M	Overall Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18.4	11	2.0	8.1	0.0	1.7	2.2	4.5	3.4	-	-
≤27.0	28	5.0	10.8	1.8	6.8	3.3	6.8	6.9	FET	.016
≤35.2	54	9.9	18.9	9.6	11.3	5.4	6.8	10.3	FET	.012
≤38.4	82	15.0	27.0	18.1	14.1	6.5	6.8	24.1	FET*	.025
≤41.6	140	25.7	48.6	33.1	19.8	18.5	11.4	34.5	FET*	.049
Min	5.0		5.0	22.2	12.8	15.0	13.0	16.2		
Max	68.0		58.4	67.4	64.6	63.2	68.0	64.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]n = 545. [b]n = 37. [c]n = 166. [d]n = 177. [e]n = 92. [f]n = 44. [g]n = 29.
*p < .05.

Table 124

Cumulative Percentages of Male Participants Failing Finger Tapping Test Non-dominant Hand Cutoffs in the Valid Sample and across Several Education Levels

FTT-NDH_M	Valid Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤18.4	2	0.5	0.0	0.0	0.0	1.4	0.0	4.5	-	-
≤27.0	12	3.1	0.0	1.0	5.0	2.9	0.0	9.1	-	-
≤35.2	24	6.2	17.6	4.9	8.6	2.9	0.0	9.1	-	-
≤38.4	42	10.9	23.5	12.6	11.5	4.3	2.9	22.7	FET*	.031
≤41.6	73	18.9	29.4	26.2	16.5	12.9	2.9	36.4	17.24*	.045
Min	15.0		33.2	26.2	19.2	15.0	38.0	16.2		
Max	68.0		55.2	67.4	64.6	63.2	68.0	64.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = average number of taps achieved with the non-dominant hand (males only); FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]n = 386. [b]n = 17. [c]n = 103. [d]n = 139. [e]n = 70. [f]n = 35. [g]n = 22.
*p < .05.

Tables 125 and 126 characterize FTT-NDH_M BR_{Fail} as a function of English language status in the overall and valid samples, respectively. In both samples, NSEs demonstrated lower BR_{Fail} than ESL patients across most FTT-NDH_M cutoffs. All of the chi-square and Fisher's exact tests were statistically significant in the overall sample (LR = 0.27 – 0.65), although they did not reach statistical significance in the valid sample, likely due to the small sample sizes across cutoffs (LR = 0.40 – 0.74).

Table 125

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Overall Sample and as a Function of English Language Status

FTT-NDH_M	Overall Sample[a]		English			
	f	%	NSE[b]	ESL[c]	χ^2	LR
≤18.4	11	2.0	1.5	5.5	FET*	0.27
≤27.0	28	5.0	4.1	11.0	FET*	0.37
≤35.2	54	9.9	8.9	16.4	4.07*	0.54
≤38.4	82	15.0	13.8	23.3	4.40*	0.59
≤41.6	140	25.7	24.0	37.0	5.59*	0.65
Min	5.0		13.0	5.0		
Max	68.0		68.0	60.8		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = Average number of taps achieved with the non-dominant hand (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]$n = 545$. [b]$n = 463$. [c]$n = 73$.
*$p < .05$.

Table 126

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test Non-dominant Hand Cutoffs in the Valid Sample and as a Function of English Language Status

FTT-NDH_M	Valid Sample[a]		English			
	f	%	NSE[b]	ESL[c]	χ^2	LR
≤18.4	2	0.5	0.6	0.0	-	-
≤27.0	12	3.1	2.7	6.8	FET	0.40
≤35.2	24	6.2	5.6	11.4	FET	0.49
≤38.4	42	10.9	10.1	18.2	FET	0.55
≤41.6	73	18.9	18.4	25.0	1.10	0.74
Min	15.0		15.0	19.6		
Max	68.0		68.0	60.8		

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH_M = Average number of taps achieved with the non-dominant hand (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-NDH_M score in the respective group; Max = maximum FTT-NDH_M score in the respective group.
[a]$n = 386$. [b]$n = 337$. [c]$n = 44$.
*$p < .05$.

FTT-C_M. Table 127 shows FTT-C_M BR_{Fail} as a function of WMT performance. Chi-square and Fisher's exact tests were statistically significant at all FTT-C_M cutoffs, suggesting a relationship between FTT-C_M performance and WMT classification. Those who passed the WMT consistently demonstrated lower FTT-C_M BR_{Fail} than those who failed the WMT. Likelihood ratios ranged from 0.16 to 0.38.

Table 127

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

	Overall Sample[a]		WMT			
FTT-C_M	*f*	%	Pass[b]	Fail[c]	χ^2	LR
≤39.4	11	2.0	0.8	5.0	FET*	0.16
≤54.8	27	5.0	3.1	9.4	9.75*	0.33
≤72.5	55	10.1	6.0	20.1	24.91*	0.30
≤79.6	82	15.0	9.1	29.6	37.00*	0.31
≤87.4	136	25.0	16.8	44.7	46.52*	0.38
Min	9.4		33.6	9.4		
Max	134.4		134.4	121.8		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of the average number of taps achieved with the dominant and non-dominant hands (males only); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 545. [b]n = 386. [c]n = 159.
*p < .05.

Tables 128 and 129 characterize FTT-C_M BR_{Fail} as a function of diagnosis in the overall and valid samples, respectively. In the overall sample, a Fisher's exact test conducted at FTT-CM ≤87.4 did not reach statistical significance; diagnosis only accounted for 1.9% of the variance in BR_{Fail} at this cutoff (small effect). In the valid sample, no chi-square tests or Fisher's exact tests were conducted. Due to the small sizes of several clinical groups, no comments are made about trends in BR_{Fail}.

Table 128

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Overall Sample and in Specific Clinical Groups

	Overall Sample[a]		TBI		Diagnostic Groups								
FTT-C_M	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤39.4	11	2.0	2.4	2.0	1.7	4.2	0.0	0.0	0.0	0.0	1.4	-	-
≤54.8	27	5.0	4.0	8.1	5.1	8.3	0.0	0.0	0.0	20.0	1.4	-	-
≤72.5	55	10.1	8.4	15.2	11.9	12.5	0.0	0.0	22.2	26.7	4.3	-	-
≤79.6	82	15.0	11.6	21.2	20.3	20.8	0.0	0.0	33.3	26.7	11.4	-	-
≤87.4	136	25.0	22.1	27.3	28.8	33.3	25.0	8.3	44.4	46.7	21.4	FET	.019
Min	9.4		9. 4	35.6	39.4	33.6	83.8	86.0	68.2	51.4	37.0		
Max	134.4		133.2	125.6	121.0	134.4	126.6	117.2	110.0	128.1	129.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 545. [b]n = 249. [c]n = 99. [d]n = 59. [e]n = 24. [f]n = 8. [g]n = 12. [h]n = 9. [i]n = 15. [j]n = 70.
*p < .05.

Table 129

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Valid Sample and in Specific Clinical Groups

	Valid Sample[a]		TBI		Diagnostic Groups								
FTT-C_M	*f*	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤39.4	3	0.8	0.0	1.4	2.2	6.3	0.0	0.0	0.0	0.0	0.0	-	-
≤54.8	12	3.1	1.2	5.4	2.2	12.5	0.0	0.0	0.0	23.1	0.0	-	-
≤72.5	23	6.0	1.9	9.5	8.9	18.8	0.0	0.0	20.0	30.8	1.8	-	-
≤79.6	35	9.1	2.5	14.9	15.6	25.0	0.0	0.0	40.0	30.8	5.5	-	-
≤87.4	65	16.8	8.7	21.6	24.4	31.3	14.3	0.0	40.0	46.2	18.2	-	-
Min	33.6		47.8	37.6	39.4	33.6	83.8	89.7	68.2	51.4	71.4		
Max	134.4		133.2	125.6	121.0	134.4	126.6	117.2	110.0	128.1	129.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 386. [b]n = 161. [c]n = 74. [d]n = 45. [e]n = 16. [f]n = 7. [g]n = 10. [h]n = 5. [i]n = 13. [j]n = 55.
*p < .05.

Tables 130 and 131 characterize FTT-C_M BR_{Fail} as a function of age in the overall and valid samples, respectively. In both samples, chi-square and Fisher's exact tests were statistically significant at some cutoffs and non-significant at others. Age accounted for 1.2% (small effect) to 5.9% (medium effect) of the variance in FTT-C_M BR_{Fail} across cutoffs in the overall sample, and for 2.1% (small effect) to 6.1% (medium effect) of the variance in FTT-C_M BR_{Fail} across cutoffs in the valid sample. Although sample sizes were small in the oldest group (60-69 years; n = 30 in the overall sample, n = 18 in the valid sample) and at more conservative cutoffs in both samples, at more liberal cutoffs, the youngest group (16-29 years) tended to have the lowest FTT-C_M BR_{Fail}, and BR_{Fail} increased with older age at each cutoff.

Table 130

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Overall Sample and across Several Age Groups

	Overall Sample[a]		Age (years)						
FTT-C_M	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤39.4	11	2.0	0.9	2.4	0.6	3.3	6.7	FET	.012
≤54.8	27	5.0	1.9	2.4	3.1	9.2	20.0	23.85*	.044
≤72.5	55	10.1	6.6	6.3	10.0	14.2	20.0	8.90	.016
≤79.6	82	15.0	9.4	9.5	14.4	20.0	40.0	22.72*	.042
≤87.4	136	25.0	16.0	15.1	26.3	33.3	56.7	31.85*	.059
Min	9.4		37.6	26.2	27.6	9.4	35.6		
Max	134.4		127.4	128.8	133.2	134.4	107.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 545. [b]n = 106. [c]n = 126. [d]n = 160. [e]n = 120. [f]n = 30.
*p < .05.

Table 131

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Valid Sample and across Several Age Groups

	Valid Sample[a]		Age (years)						
FTT-C_M	*f*	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤39.4	3	0.8	1.2	1.1	0.0	1.2	0.0	-	-
≤54.8	12	3.1	2.5	1.1	1.8	6.1	11.1	FET	.021
≤72.5	23	6.0	2.5	3.3	6.3	11.0	11.1	FET	.020
≤79.6	35	9.1	3.7	5.4	10.8	12.2	27.8	13.26*	.035
≤87.4	65	16.8	6.2	8.7	21.6	25.6	38.9	23.41*	.061
Min	33.6		37.6	39.4	51.4	33.6	54.8		
Max	134.4		127.4	128.8	133.2	134.4	107.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 386. [b]n = 81. [c]n = 92. [d]n = 111. [e]n = 82. [f]n = 18.
*p < .05.

Tables 132 and 133 characterize FTT-C_M BR_{Fail} as a function of education level in the overall and valid samples, respectively. In the overall sample, Fisher's exact tests were statistically significant at FTT-C_M ≤54.8 and ≤87.4, and education level accounted for only 1.1% to 2.8% of the variance in FTT-C_M BR_{Fail} across cutoffs (small effects). In the valid sample, a chi-square test was statistically significant at FTT-C_M ≤87.4, and education accounted for 3.4% of the variance at this cutoff (small to medium effect). With regard to trends in BR_{Fail}, there appeared to be a negative relationship between BR_{Fail} and education level at more liberal FTT-C_M cutoffs, but the most educated group (≥17 years) exhibited considerably larger BR_{Fail} than would be expected based on the trend. Notably, the least and most educated groups were small in size, and sample size was also quite small at several of the more conservative cutoffs.

Table 132

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Overall Sample and across Several Education Levels

	Overall Sample[a]		Education (years)							
FTT-C_M	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤39.4	11	2.0	5.4	0.0	2.8	2.2	2.3	3.4	-	-
≤54.8	27	5.0	8.1	1.8	7.3	2.2	6.8	10.3	FET*	.018
≤72.5	55	10.1	13.5	9.0	12.4	5.4	6.8	17.2	FET	.011
≤79.6	82	15.0	21.6	17.5	14.1	9.8	6.8	27.6	FET	.018
≤87.4	136	25.0	37.8	30.1	19.8	22.8	11.4	37.9	FET*	.028
Min	9.4		9.4	42.2	26.2	36.8	35.6	33.6		
Max	134.4		121.2	128.8	134.4	126.4	129.0	133.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 545. [b]n = 37. [c]n = 166. [d]n = 177. [e]n = 92. [f]n = 44. [g]n = 29.
*p < .05.

Table 133

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Valid Sample and across Several Education Levels

	Valid Sample[a]		Education (years)							
FTT-C_M	*f*	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤39.4	3	0.8	0.0	0.0	0.7	1.4	0.0	4.5	-	-
≤54.8	12	3.1	0.0	1.0	5.8	1.4	0.0	9.1	-	-
≤72.5	23	6.0	5.9	3.9	9.4	2.9	0.0	13.6	-	-
≤79.6	35	9.1	11.8	9.7	10.8	2.9	0.0	27.3	-	-
≤87.4	65	16.8	23.5	20.4	15.8	12.9	2.9	36.4	13.24*	.034
Min	33.6		68.6	54.2	37.6	39.4	81.1	33.6		
Max	134.4		111.6	128.8	134.4	126.4	129.0	133.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 386. [b]n = 17. [c]n = 103. [d]n = 139. [e]n = 70. [f]n = 35. [g]n = 22.
*p < .05.

Tables 134 and 135 characterize FTT-C_M BR_{Fail} as a function of English language status in the overall and valid samples, respectively. Native speakers of English consistently demonstrated lower BR_{Fail} than ESL patients across FTT-C_M cutoffs in both samples. In the overall sample, chi-square and Fisher's exact tests were statistically significant at multiple FTT-C_M scores; likelihood ratios ranged from 0.19 to 0.69. Although chi-square tests and Fisher's exact tests did not reach statistical significance in the valid sample (likely due to the small sample sizes across cutoffs), a similar pattern was observed, with NSEs demonstrating lower BR_{Fail} than ESL patients. Likelihood ratios ranged from 0.26 to 0.91.

Table 134

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Overall Sample and as a Function of English Language Status

	Overall Sample[a]		English			
FTT-C_M	*f*	%	NSE[b]	ESL[c]	χ^2	LR
≤39.4	11	2.0	1.3	6.8	FET*	0.19
≤54.8	27	5.0	3.9	12.3	FET*	0.26
≤72.5	55	10.1	9.3	15.1	2.33	0.62
≤79.6	82	15.0	13.6	23.3	4.65*	0.58
≤87.4	136	25.0	23.5	34.2	3.85	0.69
Min	9.4		27.6	9.4		
Max	134.4		134.4	129.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of the average number of taps achieved with the dominant and non-dominant hands (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]$n = 545$. [b]$n = 463$. [c]$n = 73$.
*$p < .05$.

Table 135

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Combined Cutoffs in the Valid Sample and as a Function of English Language Status

	Valid Sample[a]		English			
FTT-C_M	*f*	%	NSE[b]	ESL[c]	χ^2	LR
≤39.4	3	0.8	0.6	2.3	FET	0.26
≤54.8	12	3.1	2.7	6.8	FET	0.40
≤72.5	23	6.0	5.6	6.8	FET	0.82
≤79.6	35	9.1	8.3	13.6	FET	0.61
≤87.4	65	16.8	16.6	18.2	0.07	0.91
Min	33.6		33.6	37.6		
Max	134.4		134.4	129.0		

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of the average number of taps achieved with the dominant and non-dominant hands (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-C_M score in the respective group; Max = maximum FTT-C_M score in the respective group.
[a]n = 386. [b]n = 337. [c]n = 44.
*p < .05.

FTT-DIFF$_M$. Table 136 shows FTT-DIFF$_M$ BR$_{Fail}$, at various cutoffs, as a function of WMT performance. Participants who passed the WMT demonstrated somewhat lower BR$_{Fail}$ than those who failed the WMT across FTT-DIFF$_M$ cutoffs. However, chi-square and Fisher's exact tests largely failed to reach statistical significance, likely due to the small sample sizes across cutoffs. Likelihood ratios ranged from 0.34 to 0.70.

Table 136

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Overall Sample and as a Function of Word Memory Test Performance

FTT-DIFF$_M$	Overall Sample[a]		WMT			
	f	%	Pass[b]	Fail[c]	χ^2	LR
≤-12.2	11	2.0	1.3	3.8	FET	0.34
≤-6.6	27	5.0	3.9	7.5	3.21	0.52
≤-3.2	55	10.1	8.5	13.8	3.47	0.62
≤-1.3	81	14.9	13.2	18.9	2.85	0.70
≤.50	136	25.0	19.7	37.7	19.59*	0.52
Min	-27.8		-27.8	-24.2		
Max	32.4		32.4	25.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); WMT = Word Memory Test (Green, 2003); Pass = participants passing the WMT at standard cutoffs; Fail = participants failing the WMT at standard cutoffs; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 545. [b]n = 386. [c]n = 159.
*p < .05.

Tables 137 and 138 characterize FTT-DIFF$_M$ BR$_{Fail}$ as a function of diagnosis in the overall and valid samples, respectively. In both samples, Fisher's exact tests failed to reach statistical significance at FTT-DIFF$_M$ ≤.50. Diagnosis accounted for 1.1% of the variance in BR$_{Fail}$ at this cutoff in the overall sample, and 2.3% of the variance in BR$_{Fail}$ at this cutoff in the valid sample (small effects). Trends in BR$_{Fail}$ at lower cutoffs were not interpreted due to the small sizes of multiple clinical groups.

Table 137

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Overall Sample and in Specific Clinical Groups

FTT-DIFF$_M$	Overall Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤-12.2	11	2.0	1.2	2.0	5.1	4.2	0.0	0.0	0.0	0.0	2.9	-	-
≤-6.6	27	5.0	3.2	4.0	6.8	8.3	0.0	25.0	11.1	0.0	7.1	-	-
≤-3.2	55	10.1	7.2	8.1	16.9	12.5	0.0	25.0	22.2	0.0	15.7	-	-
≤-1.3	81	14.9	11.2	13.1	22.0	16.7	0.0	33.3	22.2	0.0	24.3	-	-
≤.50	136	25.0	22.9	24.2	30.5	37.5	12.5	33.3	22.2	13.3	27.1	FET	.011
Min	-27.8		-21.5	-25.2	-27.8	-15.4	-0.8	-8.4	-9.2	-1.1	-23.0		
Max	32.4		19.6	25.4	32.4	12.4	17.0	10.2	8.4	13.0	13.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 545. [b]n = 249. [c]n = 99. [d]n = 59. [e]n = 24. [f]n = 8. [g]n = 12. [h]n = 9. [i]n = 15. [j]n = 70.
*p < .05.

Table 138

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Valid Sample and in Specific Clinical Groups

FTT-DIFF$_M$	Valid Sample[a]		TBI		Diagnostic Groups								
	f	%	Mild[b]	M-S[c]	NEU[d]	DEP[e]	ANX[f]	SMI[g]	CP/F[h]	ORT[i]	OTH[j]	χ^2	Φ^2
≤-12.2	5	1.3	0.0	1.4	4.4	6.3	0.0	0.0	0.0	0.0	1.8	-	-
≤-6.6	15	3.9	1.2	2.7	6.7	12.5	0.0	20.0	0.0	0.0	7.3	-	-
≤-3.2	33	8.5	5.6	5.4	15.6	12.5	0.0	20.0	20.0	0.0	14.5	-	-
≤-1.3	51	13.2	8.7	10.8	22.2	12.5	0.0	30.0	20.0	0.0	23.6	-	-
≤.50	76	19.7	16.1	16.2	31.1	25.0	14.3	30.0	20.0	7.7	25.5	FET	.023
Min	-27.8		-9.4	-25.2	-27.8	-15.4	-0.8	-7.6	-3.8	-1.1	-23.0		
Max	32.4		19.6	17.2	32.4	12.4	17.0	10.2	8.4	13.0	11.6		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); TBI = traumatic brain injury, Mild = mTBI; M-S = moderate/severe TBI; NEU = neurological; DEP = depression; ANX = anxiety; SMI = severe mental illness (psychosis/Bipolar Disorder); CP/F = chronic pain and fibromyalgia; ORT = orthopedic; OTH = other; FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 386. [b]n = 161. [c]n = 74. [d]n = 45. [e]n = 16. [f]n = 7. [g]n = 10. [h]n = 5. [i]n = 13. [j]n = 55.
*p < .05.

Tables 139 and 140 characterize FTT-$DIFF_M$ BR_{Fail} as a function of age in the overall and valid samples, respectively. Chi-square tests of independence failed to reach statistical significance where conducted in both samples. Age accounted for 0.5% to 1.3% of the variance in FTT-$DIFF_M$ BR_{Fail} in the overall sample (small effects), and for 0.9% to 2.3% of the variance in FTT-$DIFF_M$ BR_{Fail} in the valid sample (small effects). Visual inspection of cumulative percentages did not reveal any consistent trends, although sample sizes were small at most cutoffs, particularly in the valid sample.

Table 139

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Overall Sample and across Several Age Groups

FTT-$DIFF_M$	Overall Sample[a]		Age (years)						
	f	545	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤-12.2	11	2.0	0.0	0.8	3.8	2.5	3.3	-	-
≤-6.6	27	5.0	2.8	5.6	6.9	4.2	3.3	2.68	.005
≤-3.2	55	10.1	5.7	9.5	11.3	11.7	16.7	4.31	.008
≤-1.3	81	14.9	12.3	14.3	13.8	19.2	16.7	2.58	.005
≤.50	136	25.0	19.8	27.0	21.3	30.0	36.7	6.75	.013
Min	-27.8		-7.8	-25.2	-27.8	-21.2	-24.2		
Max	32.4		17.4	18.8	25.4	32.4	18.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); $DIFF_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); Min = minimum FTT-$DIFF_M$ score in the respective group; Max = maximum FTT-$DIFF_M$ score in the respective group.
[a]n = 545. [b]n = 106. [c]n = 126. [d]n = 160. [e]n = 120. [f]n = 30.
*p < .05.

Table 140

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Valid Sample and across Several Age Groups

FTT-DIFF$_M$	Valid Sample[a]		Age (years)						
	f	%	16-29[b]	30-39[c]	40-49[d]	50-59[e]	60-69[f]	χ^2	Φ^2
≤-12.2	5	1.3	0.0	1.1	3.6	0.0	0.0	-	-
≤-6.6	15	3.9	0.0	5.4	7.2	2.4	0.0	-	-
≤-3.2	33	8.5	3.7	9.8	9.9	11.0	5.6	3.68	.010
≤-1.3	51	13.2	9.9	14.1	12.6	18.3	5.6	3.64	.009
≤.50	76	19.7	12.3	23.9	17.1	28.0	11.1	8.69	.023
Min	-27.8		-4.4	-25.2	-27.8	-10.8	-3.8		
Max	32.4		17.4	18.8	19.6	32.4	18.4		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 386. [b]n = 81. [c]n = 92. [d]n = 111. [e]n = 82. [f]n = 18.
*p < .05.

Tables 141 and 142 characterize FTT-DIFF$_M$ BR$_{Fail}$ as a function of education level in the overall and valid samples, respectively. Chi-square and Fisher's exact tests did not reach statistical significance at any FTT-DIFF$_M$ cutoffs in either sample, suggesting that FTT-DIFF$_M$ BR$_{Fail}$ did not vary across education levels. Effect sizes indicated that education accounted for 0.2% to 1.1% of the variance in FTT-DIFF$_M$ BR$_{Fail}$ in the overall sample (small effects), and 0.4% to 1.3% of the variance in FTT-DIFF$_M$ BR$_{Fail}$ in the valid sample (small effects). Visual inspection of BR$_{Fail}$ did not reveal any consistent trends in either sample, although sample sizes were small at most cutoffs.

Table 141

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Overall Sample and across Several Education Levels

FTT-DIFF$_M$	Overall Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤-12.2	11	2.0	2.7	2.4	1.7	1.1	2.3	3.4	FET	.002
≤-6.6	27	5.0	2.7	6.0	3.4	6.5	4.5	6.9	FET	.004
≤-3.2	55	10.1	5.4	13.3	8.5	8.7	9.1	13.8	FET	.007
≤-1.3	81	14.9	8.1	18.7	11.9	16.3	11.4	20.7	FET	.011
≤.50	136	25.0	18.9	30.1	22.0	21.7	22.7	34.5	FET	.011
Min	-27.8		-27.8	-25.2	-23.0	-21.2	-12.2	-21.5		
Max	32.4		25.4	16.6	17.4	32.4	19.4	13.2		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 545. [b]n = 37. [c]n = 166. [d]n = 177. [e]n = 92. [f]n = 44. [g]n = 29.
*p < .05.

Table 142

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Valid Sample and across Several Education Levels

FTT-DIFF$_M$	Valid Sample[a]		Education (years)							
	f	%	≤8[b]	9-11[c]	12[d]	13-15[e]	16[f]	≥17[g]	χ^2	Φ^2
≤-12.2	5	1.3	5.9	1.0	2.2	0.0	0.0	0.0	-	-
≤-6.6	15	3.9	5.9	4.9	4.3	2.9	2.9	0.0	-	-
≤-3.2	33	8.5	5.9	10.7	8.6	5.7	8.6	9.1	FET	.004
≤-1.3	51	13.2	5.9	14.6	12.2	15.7	8.6	18.2	FET	.007
≤.50	76	19.7	5.9	22.3	18.0	20.0	17.1	31.8	4.95	.013
Min	-27.8		-27.8	-25.2	-23.0	-9.4	-10.8	-4.0		
Max	32.4		21.8	16.6	17.4	32.4	18.8	11.6		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 386. [b]n = 17. [c]n = 103. [d]n = 139. [e]n = 70. [f]n = 35. [g]n = 22.
*p < .05.

Tables 143 and 144 characterize FTT-$DIFF_M$ BR_{Fail} as a function of English language status in the overall and valid samples, respectively. Fisher's exact tests and chi-square tests of independence failed to reach statistical significance across cutoffs in either sample. However, sample sizes were quite small, particularly at more conservative cutoffs. Likelihood ratios ranged from 0.63 to 0.90 in the overall sample and from 0.96 to 1.98 in the valid sample.

Table 143

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Overall Sample and as a Function of English Language Status

FTT-$DIFF_M$	Overall Sample[a]		English			
	f	%	NSE[b]	ESL[c]	χ^2	LR
≤-12.2	11	2.0	1.7	2.7	FET	0.63
≤-6.6	27	5.0	4.5	6.8	FET	0.66
≤-3.2	55	10.1	9.9	11.0	0.07	0.90
≤-1.3	81	14.9	14.0	20.5	2.10	0.68
≤.50	136	25.0	24.2	30.1	1.19	0.80
Min	-27.8		-27.8	-21.5		
Max	32.4		32.4	21.8		

Note. FTT = Finger Tapping Test (Reitan, 1969); $DIFF_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-$DIFF_M$ score in the respective group; Max = maximum FTT-$DIFF_M$ score in the respective group.
[a]$n = 545$. [b]$n = 463$. [c]$n = 73$.
*$p < .05$.

Table 144

Cumulative Percentages of Male Participants Failing Various Finger Tapping Test-Difference Cutoffs in the Valid Sample and as a Function of English Language Status

FTT-DIFF$_M$	Valid Sample[a]		English			
	f	%	NSE[b]	ESL[c]	χ^2	LR
≤-12.2	5	1.3	1.2	0.0	-	-
≤-6.6	15	3.9	3.9	2.3	FET	1.70
≤-3.2	33	8.5	8.9	4.5	FET	1.98
≤-1.3	51	13.2	13.1	13.6	0.01	0.96
≤.50	76	19.7	19.6	20.5	0.02	0.96
Min	-27.8		-27.8	-7.2		
Max	32.4		32.4	21.8		

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); NSE = native speakers of English; ESL = English as a second language; LR = likelihood ratio; FET = Fisher's exact test; Min = minimum FTT-DIFF$_M$ score in the respective group; Max = maximum FTT-DIFF$_M$ score in the respective group.
[a]n = 386. [b]n = 337. [c]n = 44.
*p < .05.

Study 3

Classification accuracy statistics are reported below for all PVTs of interest. In all of the analyses, the WMT was used as the criterion measure. Although signal detection properties were investigated for multiple potential cutoffs on each PVT of interest, data are only presented for cutoffs that reached or exceeded the minimum acceptable level of specificity (.84; Larrabee, 2003). Furthermore, for each cutoff examined on a given PVT, positive and negative predictive values are reported at five hypothetical BR_{Fail} ranging from 10% to 50%. Given the 30% BR_{Fail} on the WMT in the overall sample, in the text, PPP and NPP values are interpreted at a hypothetical BR_{Fail} of 30% for all analyses. Where multiple cutoff scores were investigated on a PVT, PPP and NPP are interpreted at the first cutoff that demonstrated at least .90 specificity. As in Studies 1 and 2, classification accuracy data are presented for free-standing PVTs first, followed by embedded validity indicators.

MSVT. Table 145 shows signal detection properties for the MSVT. Standard cutoffs on the MSVT (Green, 2004) produced .52 sensitivity and .93 specificity. PPP and NPP values indicate that, in a setting with a 30% BR_{Fail}, there is a .77 probability that a profile is invalid when an examinee fails the MSVT at standard cutoffs, and a .82 probability that a profile is valid when the examinee passes the MSVT at standard cutoffs.

Table 145

Signal Detection Properties of the Standard Cutoff on the Medical Symptom Validity Test against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
MSVT	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
STN	462	21.3	.52	.93	PPP	.46	.66	.77	.84	.88
					NPP	.95	.89	.82	.75	.66

Note. MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs specified in the test manual; BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]n = 2165

NV-MSVT. Table 146 shows classification accuracy statistics for the NV-MSVT. Standard cutoffs on the NV-MSVT (Green, 2008) produced .93 specificity and .45 sensitivity. In settings with a 30% BR_{Fail}, there is a .72 probability that a profile is invalid when an examinee fails the NV-MSVT at standard cutoffs (PPP), and a .80 probability that a profile is valid when the examinee passes the NV-MSVT at standard cutoffs (NPP).

Table 146

Signal Detection Properties of the Standard Cutoff on the Non-Verbal Medical Symptom Validity Test against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
NV-MSVT	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
STN	308	18.7	.45	.93	PPP	.41	.61	.72	.80	.81
					NPP	.94	.87	.80	.72	.72

Note. NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); STN = standard cutoffs specified in the test manual; BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 1645

TOMM Trial 1. Table 147 shows signal detection properties of various cutoffs on TOMM Trial 1. Trial 1 cutoffs between ≤33 and ≤44 produced specificity values between .84 and 1.00. Trial 1 ≤44 achieved .61 sensitivity with .87 specificity. Lowering the cutoff to Trial 1 ≤43 improved specificity (.90) with a negligible loss in sensitivity (.57). Perfect specificity was achieved at Trial 1 ≤33, at the expense of sensitivity (.20).

Although previously published TOMM Trial 1 cutoffs vary widely (i.e., between Trial 1 ≤35 and ≤45), depending on sample characteristics and the criteria used to define invalid performance, a cutoff of Trial 1 ≤40 was recommended in a review of Trial 1 findings (Denning, 2012). In the present sample, Trial 1 ≤40 demonstrated good classification accuracy with .96 specificity and .42 sensitivity.

In a setting with a 30% BR_{Fail}, the probability of an invalid profile in an examinee who fails Trial 1 ≤43 is .71. The probability of a valid profile in someone who passes Trial 1 ≤43 is .83.

Table 147

Signal Detection Properties of Various TOMM Trial 1 Cutoffs Achieving Acceptable Specificity against the Word Memory Test

TOMM Trial 1	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤33	125	7.1	.20	1.00	PPP	.84	.92	.95	.97	1.00
					NPP	.92	.83	.74	.65	.65
≤34	151	8.6	.24	.99	PPP	.78	.89	.93	.95	.94
					NPP	.92	.84	.75	.66	.66
≤35	168	9.6	.27	.99	PPP	.78	.89	.93	.95	.95
					NPP	.92	.84	.76	.67	.67
≤36	188	10.7	.29	.99	PPP	.70	.84	.90	.93	.95
					NPP	.93	.85	.76	.68	.68
≤37	207	11.8	.32	.98	PPP	.68	.83	.89	.93	.91
					NPP	.93	.85	.77	.68	.68
≤38	232	13.2	.35	.98	PPP	.64	.80	.87	.91	.92
					NPP	.93	.86	.78	.69	.69
≤39	258	14.7	.38	.97	PPP	.58	.75	.84	.89	.89
					NPP	.93	.86	.78	.70	.70
≤40	292	16.6	.42	.96	PPP	.55	.74	.83	.88	.88
					NPP	.94	.87	.79	.71	.71
≤41	333	18.9	.46	.95	PPP	.49	.68	.79	.85	.86
					NPP	.94	.87	.80	.72	.73
≤42	380	21.6	.51	.93	PPP	.45	.65	.76	.83	.83
					NPP	.94	.88	.81	.74	.74
≤43	449	25.5	.57	.90	PPP	.39	.59	.71	.79	.79
					NPP	.95	.89	.83	.76	.76
≤44	511	29.1	.61	.87	PPP	.34	.54	.67	.76	.76
					NPP	.95	.90	.84	.77	.77

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.

[a]n = 1759

TOMM Trial 2. Table 148 shows signal detection properties for TOMM Trial 2. TOMM Trial 2 cutoffs between ≤38 and ≤49 produced specificity values between .84 and 1.00. The highest conceivable cutoff (Trial 2 ≤49) achieved .54 sensitivity and .89 specificity. Lowering the cutoff to Trial 2 ≤48 improved specificity to .96 but reduced sensitivity to .44. Perfect specificity was attained at Trial 2 ≤38, although sensitivity was low (.16).

With respect to previously identified cutoffs, Trial 2 ≤44, the standard cutoff recommended in the TOMM manual (Tombaugh, 1996), was found to be overly conservative with .99 specificity and .29 sensitivity. Higher Trial 2 cutoffs were associated with improved specificity, while maintaining good specificity.

In a setting with a 30% BR_{Fail}, there is an .83 probability that a failure on Trial 2 ≤48 represents an invalid profile and a .80 probability that a pass on Trial 2 ≤48 denotes a valid profile.

Table 148

Signal Detection Properties of Various TOMM Trial 2 Cutoffs Achieving Acceptable Specificity against the Word Memory Test

TOMM	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
Trial 2	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤38	100	5.7	.16	1.00	PPP	.81	.90	.94	.96	1.00
					NPP	.91	.83	.73	.64	.54
≤39	118	6.7	.19	.99	PPP	.80	.90	.94	.96	.95
					NPP	.92	.83	.74	.65	.55
≤40	128	7.3	.20	.99	PPP	.77	.88	.93	.95	.95
					NPP	.92	.83	.74	.65	.55
≤41	144	8.2	.23	.99	PPP	.75	.87	.92	.95	.96
					NPP	.92	.84	.75	.66	.56
≤42	152	8.7	.24	.99	PPP	.76	.88	.92	.95	.96
					NPP	.92	.84	.75	.66	.57
≤43	167	9.5	.26	.99	PPP	.76	.88	.92	.95	.96
					NPP	.92	.84	.76	.67	.57
≤44	185	10.5	.29	.99	PPP	.73	.86	.91	.94	.97
					NPP	.93	.85	.76	.68	.58
≤45	199	11.3	.31	.99	PPP	.70	.84	.90	.93	.97
					NPP	.93	.85	.77	.68	.59
≤46	215	12.4	.34	.98	PPP	.70	.84	.90	.93	.94
					NPP	.93	.86	.78	.69	.60
≤47	250	14.2	.38	.98	PPP	.64	.80	.87	.91	.95
					NPP	.93	.86	.78	.70	.61
≤48	301	17.1	.44	.96	PPP	.56	.74	.83	.88	.92
					NPP	.94	.87	.80	.72	.63
≤49	448	25.5	.54	.89	PPP	.34	.54	.67	.76	.83
					NPP	.95	.88	.82	.74	.66

Note. TOMM = Test of Memory Malingering (Tombaugh, 1996); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.

[a]*n* = 1759

RDS. Table 149 shows signal detection properties for various RDS cutoffs. RDS cutoffs between ≤4 and ≤7 produced specificity values between .84 and 1.00. RDS ≤7 achieved .37 sensitivity and .86 specificity. Lowering the cutoff to RDS ≤6 improved specificity (.96) but resulted in a considerable loss in sensitivity (.17). RDS ≤5 resulted in near-perfect specificity (.99) and low sensitivity (.07). Perfect specificity was attained at RDS ≤4, although sensitivity was very low (.02). In a setting with a 30% BR_{Fail}, there is a .66 probability that a fail on RDS ≤6 represents an invalid profile and a .73 probability that a pass on RDS ≤6 reflects a valid profile.

Table 149

Signal Detection Properties of Various Reliable Digit Span Cutoffs Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
RDS	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤4	12	0.6	.02	1.00	PPP	.73	.86	.91	.94	1.00
					NPP	.90	.80	.70	.60	.51
≤5	53	2.6	.07	.99	PPP	.52	.71	.81	.87	.88
					NPP	.91	.81	.71	.62	.52
≤6	155	7.7	.17	.96	PPP	.34	.54	.66	.75	.81
					NPP	.91	.82	.73	.63	.54
≤7	425	21.2	.37	.86	PPP	.22	.39	.53	.63	.73
					NPP	.92	.84	.76	.67	.58

Note. RDS = Reliable digit span (Greiffenstein, Baker, & Gola, 1994); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 2003

WCST FMS. Table 150 shows signal detection properties for various cutoffs on WCST FMS. FMS cutoffs between ≥3 and ≥7 produced specificity values between .84 and 1.00. FMS ≥3 achieved .29 sensitivity and .86 specificity. Increasing the cutoff to FMS ≥4 improved specificity (.94) but produced lower sensitivity (.17). FMS ≥7 achieved perfect specificity with nearly zero sensitivity to invalid performance (.01).

With respect to previously identified cutoffs, FMS ≥3 demonstrated good classification accuracy. However, FMS ≥2 (Larrabee, 2003) produced unacceptably low specificity (.73).

PPP and NPP indicate that, in a setting with a 30% BR_{Fail}, there is a .55 probability that a failure on FMS ≥4 represents an invalid profile and a .73 probability that a pass on FMS ≥4 reflects a valid profile.

Table 150

Signal Detection Properties of Various Failure to Maintain Set Cutoffs Achieving Acceptable Specificity against the Word Memory Test

WCST FMS	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≥7	4	0.6	.01	1.00	PPP	.22	.39	.52	.63	.72
					NPP	.90	.80	.70	.60	.50
≥6	11	1.7	.02	.98	PPP	.14	.27	.38	.49	.50
					NPP	.90	.80	.70	.60	.50
≥5	26	4.1	.08	.98	PPP	.28	.46	.60	.70	.80
					NPP	.91	.81	.71	.62	.52
≥4	57	9.0	.17	.94	PPP	.24	.41	.55	.65	.74
					NPP	.91	.82	.73	.63	.53
≥3	117	18.5	.29	.86	PPP	.18	.34	.47	.58	.67
					NPP	.92	.83	.74	.64	.55

Note. WCST = Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); FMS = Failure to Maintain Set; BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]n = 634

CT-TE. Table 151 shows signal detection properties of various CT-TE cutoffs. CT-TE cutoffs between ≥90 and ≥135 produced specificity values between .84 and 1.00. CT-TE ≥90 achieved .34 sensitivity and .84 specificity. Raising the cutoff to CT-TE ≥100 improved specificity to .90 with a loss in sensitivity (.23). Perfect specificity was achieved at CT-TE ≥135, although sensitivity was near zero (.01).

With respect to previously identified cutoffs, CT-TE ≥90 and CT-TE ≥100 produced good classification accuracy. However, CT-TE ≥87 (Tenhula & Sweet, 1996; DiCarlo et al., 2000) was associated with unacceptably low specificity (.83).

PPP and NPP indicate that, in settings with a 30% BR_{Fail}, there is a .51 probability that the profile of a patient who fails CT-TE ≥100 is invalid, and a .73 probability that the profile of a patient who passes CT-TE ≥100 is valid.

Table 151

Signal Detection Properties of Various Category Test Total Errors Cutoffs Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
CT-TE	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≥135	8	0.4	.01	1.00	PPP	.33	.52	.65	.75	.81
					NPP	.90	.80	.70	.60	.50
≥130	16	0.7	.01	.99	PPP	.23	.40	.53	.64	.73
					NPP	.90	.80	.70	.60	.50
≥125	35	1.6	.03	.99	PPP	.22	.38	.52	.62	.71
					NPP	.90	.80	.70	.60	.50
≥120	62	2.8	.05	.98	PPP	.25	.43	.56	.67	.75
					NPP	.90	.81	.71	.61	.51
≥115	94	4.3	.08	.97	PPP	.23	.41	.54	.65	.73
					NPP	.90	.81	.71	.61	.51
≥110	143	6.5	.11	.95	PPP	.21	.37	.51	.61	.71
					NPP	.91	.81	.71	.62	.52
≥105	217	9.9	.17	.93	PPP	.21	.37	.51	.61	.70
					NPP	.91	.82	.72	.63	.53
≥100	294	13.4	.23	.90	PPP	.21	.38	.51	.62	.71
					NPP	.91	.82	.73	.64	.54
≥95	366	16.6	.28	.88	PPP	.20	.36	.49	.60	.69
					NPP	.92	.83	.74	.65	.55
≥90	459	20.9	.34	.84	PPP	.19	.35	.48	.59	.68
					NPP	.92	.84	.75	.66	.56

Note. CT-TE = Category Test Total Errors (DeFlippis & McCampbell, 1979, 1991, 1997); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 2199

TMT-A. Table 152 shows signal detection properties for various cutoffs on TMT-A. Cutoffs between ≥45 and ≥90 produced specificity values between .84 and 1.00. TMT-A ≥45 achieved .37 sensitivity and .85 specificity. Raising the cutoff to TMT-A ≥50 improved specificity to .91 with a loss in sensitivity (.30). A cutoff of TMT-A ≥90 was associated with perfect specificity, although sensitivity was low (.07).

With respect to previously identified cutoffs, TMT-A ≥63 was somewhat conservative, producing high specificity at the cost of low sensitivity (SPEC = .97, SENS = .16).

PPP and NPP values indicated that in a setting with a 30% BR_{Fail}, there is a .59 probability that a failure on TMT-A ≥50 represents an invalid profile, and a .75 probability that a pass on TMT-A ≥50 reflects a valid profile.

Table 152

Signal Detection Properties of Various Trail Making Test Part A Completion Time Cutoffs Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
TMT-A	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≥90	59	2.4	.07	1.00	PPP	.70	.84	.90	.93	.95
					NPP	.91	.81	.71	.62	.52
≥85	71	2.9	.08	.99	PPP	.59	.76	.85	.90	.93
					NPP	.91	.81	.72	.62	.52
≥80	91	3.7	.10	.99	PPP	.47	.66	.77	.84	.89
					NPP	.91	.81	.72	.62	.52
≥75	111	4.5	.12	.98	PPP	.46	.66	.77	.84	.88
					NPP	.91	.82	.72	.63	.53
≥70	136	5.5	.13	.98	PPP	.38	.57	.70	.78	.84
					NPP	.91	.82	.72	.63	.53
≥65	165	6.7	.15	.97	PPP	.36	.56	.68	.77	.83
					NPP	.91	.82	.73	.63	.53
≥60	209	8.5	.18	.96	PPP	.32	.51	.64	.74	.81
					NPP	.91	.82	.73	.64	.54
≥55	271	11.0	.22	.94	PPP	.29	.47	.61	.71	.78
					NPP	.92	.83	.74	.64	.55
≥50	372	15.1	.30	.91	PPP	.27	.46	.59	.69	.77
					NPP	.92	.84	.75	.66	.57
≥45	527	21.4	.37	.85	PPP	.22	.39	.52	.63	.72
					NPP	.92	.85	.76	.67	.58

Note. TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 2463

TMT-B. Table 153 shows signal detection properties of various TMT-B cutoffs. Cutoffs between ≥110 and ≥360 produced specificity values between .84 and 1.00. TMT-B ≥110 achieved .39 sensitivity and.87 specificity. Raising the cutoff to TMT-B ≥120 improved specificity to .90 with a minimal loss in sensitivity (.33). Further increases in the cutoff score produced steady losses in sensitivity and gradual gains in specificity. Thus, only classification accuracy data for cutoffs up to TMT-B ≥230 are shown in Table 149.

Additional analyses revealed that TMT-B ≥350 was associated with .99 specificity and .04 sensitivity (BR_{Fail} = 1.6%). Perfect specificity was attained using TMT-B ≥360, although sensitivity remained low (.04; BR_{Fail} = 1.5%).

With regard to previously identified cutoffs, TMT-B ≥120 demonstrated good classification accuracy in the present sample. TMT-B ≥200 was overly conservative, producing high specificity at the cost of low sensitivity (SPEC = .98, SENS = .13).

In a setting with a 30% BR_{Fail}, there is a .58 probability that a failure on TMT-B ≥120 represents an invalid profile and a .76 probability that a pass on TMT-B ≥120 represents a valid profile.

Table 153

Signal Detection Properties of Various Trail Making Test Part B Completion Time Cutoffs Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
TMT-B	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≥230	93	3.8	.10	.99	PPP	.45	.65	.76	.83	.88
					NPP	.91	.81	.72	.62	.52
≥220	105	4.3	.11	.98	PPP	.42	.62	.74	.81	.87
					NPP	.91	.82	.72	.62	.52
≥210	117	4.8	.12	.98	PPP	.43	.63	.75	.82	.87
					NPP	.91	.82	.72	.63	.53
≥200	129	5.3	.13	.98	PPP	.38	.58	.70	.79	.85
					NPP	.91	.82	.72	.63	.53
≥190	151	6.2	.14	.97	PPP	.36	.56	.68	.77	.84
					NPP	.91	.82	.73	.63	.53
≥180	171	7.0	.16	.97	PPP	.35	.55	.67	.76	.83
					NPP	.91	.82	.73	.63	.53
≥170	183	7.5	.17	.96	PPP	.34	.54	.66	.75	.82
					NPP	.91	.82	.73	.63	.54
≥160	208	8.5	.19	.96	PPP	.33	.53	.66	.75	.82
					NPP	.91	.83	.73	.64	.54
≥150	244	10.0	.22	.95	PPP	.32	.52	.65	.74	.81
					NPP	.92	.83	.74	.64	.55
≥140	298	12.2	.25	.93	PPP	.29	.48	.61	.71	.79
					NPP	.92	.83	.74	.65	.55
≥130	342	14.0	.28	.92	PPP	.27	.45	.59	.69	.77
					NPP	.92	.84	.75	.65	.56
≥120	422	17.3	.33	.90	PPP	.26	.44	.58	.68	.76
					NPP	.92	.84	.76	.67	.57
≥110	508	20.8	.39	.87	PPP	.25	.43	.56	.66	.75
					NPP	.93	.85	.77	.68	.59

Note. TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.

[a]*n* = 2442

TMT A+B. Table 154 shows signal detection properties for various cutoffs on TMT A+B. Cutoffs between TMT A+B ≥150 and ≥430 produced specificity values between .86 and 1.00. TMTA+B ≥150 achieved .41 sensitivity and .86 specificity. Raising the cutoff to TMT A+B ≥170 improved specificity to .91 but reduced sensitivity to .34. Further increases in cutoff scores produced small changes in sensitivity and specificity. Thus, only classification accuracy data for cutoffs up to TMT A+B ≥250 are shown in Table 154.

Additional analyses revealed that TMT A+B ≥300 was associated with .99 specificity and .10 sensitivity (BR_{Fail} = 4.0%). TMT A+B ≥430 produced perfect specificity and .04 sensitivity (BR_{Fail} = 1.6%).

With regard to previously identified cutoffs, TMT A+B ≥170 demonstrated good classification accuracy. TMT A+B ≥137 produced good sensitivity (.48) but unacceptably low specificity (.81).

Examination of PPP and NPP values revealed that in settings with a 30% BR_{Fail}, there is a .61 probability of an invalid profile in an examinee who fails TMT A+B ≥170 and a .76 probability of a valid profile in an examinee who passes TMT A+B ≥170.

Table 154

Signal Detection Properties of Various TMT A+B Cutoffs Achieving Acceptable Specificity against the Word Memory Test

TMT A+B	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≥250	147	6.0	.14	.97	PPP	.36	.56	.69	.77	.84
					NPP	.91	.82	.72	.63	.53
≥240	164	6.7	.15	.97	PPP	.36	.55	.68	.77	.83
					NPP	.91	.82	.73	.63	.53
≥230	183	7.5	.17	.97	PPP	.37	.56	.69	.78	.84
					NPP	.91	.82	.73	.64	.54
≥220	200	8.2	.18	.96	PPP	.34	.54	.66	.75	.82
					NPP	.91	.82	.73	.64	.54
≥210	217	8.9	.20	.96	PPP	.34	.54	.67	.76	.82
					NPP	.91	.83	.74	.64	.54
≥200	256	10.5	.23	.95	PPP	.34	.53	.66	.75	.82
					NPP	.92	.83	.74	.65	.55
≥190	293	12.0	.26	.94	PPP	.33	.52	.65	.74	.81
					NPP	.92	.84	.75	.66	.56
≥180	342	14.0	.30	.93	PPP	.31	.51	.64	.73	.81
					NPP	.92	.84	.76	.66	.57
≥170	400	16.4	.34	.91	PPP	.29	.48	.61	.71	.79
					NPP	.92	.85	.76	.67	.58
≥160	462	18.9	.37	.89	PPP	.27	.45	.59	.69	.77
					NPP	.93	.85	.77	.68	.59
≥150	540	22.1	.41	.86	PPP	.25	.43	.56	.66	.75
					NPP	.93	.85	.77	.69	.59

Note. TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]n = 2442

TMT-B/A. Table 155 shows signal detection properties for various TMT-B/A cutoffs. Cutoffs between <1.70 and <1.10 produced specificity values between .86 and 1.00. TMT-B/A <1.70 achieved .14 sensitivity and .86 specificity. Lowering the cutoff to TMT-B/A <1.50, which was been published previously in the literature, improved specificity to .92 but reduced sensitivity to .07. Although perfect specificity was achieved at TMT-B/A <1.10, there was zero sensitivity to invalid performance at that cutoff.

In a setting with a 30% BR_{Fail}, the probability of an invalid profile in an examinee who fails TMT-B/A <1.50 is .29. Conversely, the probability of a valid profile in an examinee who passes TMT-B/A <1.50 is .70.

Table 155

Signal Detection Properties of Various Trail Making Test Ratio Cutoffs Achieving Acceptable Specificity against the Word Memory Test

TMT-B/A	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
<1.10	10	0.4	.00	1.00	PPP	.10	.20	.30	.40	.50
					NPP	.90	.80	.70	.60	.50
<1.20	29	1.2	.01	.99	PPP	.07	.15	.24	.32	.42
					NPP	.90	.80	.70	.60	.50
<1.30	66	2.7	.02	.97	PPP	.08	.16	.24	.34	.43
					NPP	.90	.80	.70	.60	.50
<1.40	120	4.9	.05	.95	PPP	.09	.18	.28	.37	.47
					NPP	.90	.80	.70	.60	.50
<1.50	183	7.5	.07	.92	PPP	.10	.19	.29	.39	.49
					NPP	.90	.80	.70	.60	.50
<1.60	256	10.5	.10	.89	PPP	.09	.18	.28	.37	.47
					NPP	.90	.80	.70	.60	.50
<1.70	349	14.3	.14	.86	PPP	.10	.19	.29	.39	.49
					NPP	.90	.80	.70	.60	.50

Note. TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 2442

FTT-DH$_M$. Table 156 shows signal detection properties for various FTT-DH$_M$ cutoffs. Cutoffs between FTT-DH$_M$ ≤16.0 and ≤44.0 produced specificity values between .84 and 1.00. FTT-DH$_M$ ≤44.0 achieved .38 sensitivity and .86 specificity. Improved specificity was achieved at FTT-DH$_M$ ≤42.0 (.91) at the cost of lower sensitivity (.29). Lowering cutoffs further produced small increases in specificity and steady decreases in sensitivity. Perfect specificity was attained using FTT-DH$_M$ ≤16.0; however, sensitivity was low (SENS = .04).

With regard to previously identified cutoffs, FTT-DH$_M$ ≤35.0 demonstrated acceptable classification accuracy, achieving .95 specificity and .19 sensitivity. FTT-DH$_M$ ≤37.0 achieved .94 specificity and .20 sensitivity.

According to PPP and NPP, in settings with a 30% BR$_{Fail}$, there is a .58 probability that a failure on FTT-DH$_M$ ≤42.0 represents an invalid profile, and a .75 probability that a pass on FTT-DH$_M$ ≤42.0 represents a valid profile.

Table 156

Signal Detection Properties of Various Finger Tapping Test Dominant Hand Cutoffs (in Males) Achieving Acceptable Specificity against the Word Memory Test

FTT-DH_M	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤16.0	7	1.3	.04	1.00	PPP	1.00	1.00	1.00	1.00	1.00
					NPP	.90	.81	.71	.61	.51
≤18.0	11	2.0	.06	.99	PPP	.55	.73	.82	.88	.92
					NPP	.90	.81	.71	.61	.51
≤20.0	15	2.7	.07	.99	PPP	.42	.62	.74	.82	.87
					NPP	.91	.81	.71	.61	.52
≤22.0	17	3.1	.07	.99	PPP	.39	.59	.71	.79	.85
					NPP	.91	.81	.71	.62	.52
≤24.0	19	3.4	.07	.98	PPP	.31	.51	.64	.73	.81
					NPP	.91	.81	.71	.61	.51
≤26.0	23	4.2	.08	.97	PPP	.26	.44	.57	.68	.76
					NPP	.91	.81	.71	.61	.51
≤28.0	29	5.2	.10	.97	PPP	.28	.46	.59	.70	.77
					NPP	.91	.81	.72	.62	.52
≤30.0	32	5.8	.12	.97	PPP	.28	.47	.60	.70	.78
					NPP	.91	.81	.72	.62	.52
≤32.0	36	6.5	.14	.97	PPP	.32	.52	.65	.74	.81
					NPP	.91	.82	.72	.63	.53
≤34.0	44	8.0	.17	.96	PPP	.30	.49	.62	.72	.79
					NPP	.91	.82	.73	.63	.53
≤36.0	55	9.9	.20	.94	PPP	.27	.46	.59	.69	.77
					NPP	.91	.82	.73	.64	.54
≤38.0	59	10.7	.21	.94	PPP	.27	.45	.58	.69	.77
					NPP	.91	.83	.73	.64	.54
≤40.0	71	12.8	.26	.93	PPP	.28	.47	.60	.70	.78
					NPP	.92	.83	.74	.65	.56
≤42.0	82	14.8	.29	.91	PPP	.26	.45	.58	.68	.76
					NPP	.92	.84	.75	.66	.56
≤44.0	115	20.8	.38	.86	PPP	.23	.41	.54	.65	.73
					NPP	.93	.85	.76	.67	.58

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_M = average number of taps achieved with the dominant hand (males only); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003);

SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 553

FTT-DH_F. Table 157 shows signal detection properties of various FTT-DH_F cutoffs. Cutoffs between FTT-DH_F ≤12.0 and ≤34.0 produced specificity values between .86 and 1.00. FTT-DH_F ≤34.0 achieved .41 sensitivity and .86 specificity. FTT-DH_F ≤30.0 improved specificity to .91 with a reduction in sensitivity (.33). Lowering cutoffs further resulted in steady losses in sensitivity and small gains in specificity. Perfect specificity was achieved at FTT-DH_F ≤12.0, although sensitivity was very low (.03).

With regard to previously identified cutoffs, FTT-DH_F ≤28.0 demonstrated good classification accuracy, with .92 specificity and .32 sensitivity. FTT-DH_F ≤23.0 produced higher specificity (.98) and lower sensitivity (.19).

In settings with a 30% BR_{Fail}, the probability of an invalid profile given a failure on FTT-DH_F ≤30 is .61. The probability of a valid profile given a pass on FTT-DH_F ≤30 is .76.

Table 157

Signal Detection Properties of Various Finger Tapping Test Dominant Hand Cutoffs (in Females) Achieving Acceptable Specificity against the Word Memory Test

FTT-DH_F	Overall Sample[a]		Criterion PVT: WMT			BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤12.0	2	0.7	.03	1.00	PPP	1.00	1.00	1.00	1.00	1.00
					NPP	.90	.81	.71	.61	.51
≤14.0	4	1.5	.03	.99	PPP	.27	.45	.58	.69	.77
					NPP	.90	.80	.70	.61	.51
≤16.0	6	2.2	.06	.99	PPP	.42	.62	.74	.81	.87
					NPP	.90	.81	.71	.61	.51
≤18.0	9	3.3	.08	.98	PPP	.31	.51	.64	.73	.80
					NPP	.91	.81	.71	.62	.52
≤20.0	11	4.1	.11	.98	PPP	.39	.59	.71	.79	.85
					NPP	.91	.82	.72	.62	.52
≤22.0	14	5.2	.14	.98	PPP	.40	.60	.72	.80	.85
					NPP	.91	.82	.73	.63	.53
≤24.0	19	7.1	.21	.97	PPP	.44	.64	.75	.83	.88
					NPP	.92	.83	.74	.65	.55
≤26.0	26	9.7	.25	.95	PPP	.37	.57	.69	.78	.84
					NPP	.92	.84	.75	.66	.56
≤28.0	36	13.4	.32	.92	PPP	.31	.51	.64	.73	.80
					NPP	.92	.84	.76	.67	.57
≤30.0	44	16.4	.33	.91	PPP	.29	.47	.61	.71	.78
					NPP	.92	.84	.76	.67	.58
≤32.0	51	19.0	.40	.87	PPP	.26	.44	.57	.68	.76
					NPP	.93	.85	.77	.68	.59
≤34.0	55	20.4	.41	.86	PPP	.25	.42	.56	.66	.75
					NPP	.93	.85	.77	.69	.59

Note. FTT = Finger Tapping Test (Reitan, 1969); DH_F = average number of taps achieved with the dominant hand (females only); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 269

FTT-NDH$_M$. Table 158 shows signal detection properties of various FTT-NDH$_M$ cutoffs. Cutoffs between ≤16.0 and ≤40.0 produced specificity values between .85 and 1.00. FTT-NDH$_M$ ≤40.0 achieved .33 sensitivity and .85 specificity. FTT-NDH$_M$ ≤38.0 improved specificity to .91, with a reduction in sensitivity (.25). Lowering cutoffs further produced small gains in specificity and steady losses in sensitivity. Zero false positives were observed at FTT-NDH$_M$ ≤16.0; however, sensitivity to invalid performance was very low (.03).

With regard to previously identified cutoffs, Arnold and colleagues' (2005) cutoff of FTT-NDH$_M$ ≤30.0 achieved excellent specificity (.96) but low sensitivity (.11) in the present sample.

In settings with a 30% BR$_{Fail}$, there is a .53 probability that a fail on FTT-NDH$_M$ ≤38.0 represents an invalid profile (PPP), and a .74 probability that a pass on FTT-NDH$_M$ ≤38.0 represents a valid profile (NPP).

Table 158

Signal Detection Properties of Various Finger Tapping Test Non-Dominant Hand Cutoffs (in Males) Achieving Acceptable Specificity against the Word Memory Test

FTT-NDH$_M$	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR$_{Fail}$				
	f	BR$_{Fail}$	SENS	SPEC		10	20	30	40	50
≤16.0	6	1.1	.03	1.00	PPP	.57	.75	.84	.89	.92
					NPP	.90	.80	.71	.61	.51
≤18.0	10	1.8	.05	.99	PPP	.52	.71	.81	.87	.91
					NPP	.90	.81	.71	.61	.51
≤20.0	16	2.9	.06	.98	PPP	.31	.50	.63	.73	.80
					NPP	.90	.81	.71	.61	.51
≤22.0	19	3.5	.07	.98	PPP	.27	.45	.59	.69	.77
					NPP	.90	.81	.71	.61	.51
≤24.0	22	4.0	.08	.98	PPP	.28	.47	.60	.70	.78
					NPP	.91	.81	.71	.61	.52
≤26.0	23	4.2	.08	.97	PPP	.26	.44	.57	.68	.76
					NPP	.91	.81	.71	.61	.51
≤28.0	28	5.1	.10	.97	PPP	.26	.45	.58	.68	.76
					NPP	.91	.81	.72	.62	.52
≤30.0	33	6.1	.11	.96	PPP	.24	.42	.56	.66	.74
					NPP	.91	.81	.72	.62	.52
≤32.0	36	6.6	.13	.96	PPP	.25	.43	.57	.67	.75
					NPP	.91	.81	.72	.62	.52
≤34.0	47	8.6	.16	.95	PPP	.25	.43	.56	.67	.75
					NPP	.91	.82	.73	.63	.53
≤36.0	64	11.7	.23	.93	PPP	.26	.44	.57	.68	.76
					NPP	.92	.83	.74	.64	.55
≤38.0	75	13.8	.25	.91	PPP	.23	.40	.53	.64	.72
					NPP	.92	.83	.74	.64	.55
≤40.0	109	20.0	.33	.85	PPP	.20	.36	.49	.60	.69
					NPP	.92	.84	.75	.66	.56

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH$_M$ = average number of taps achieved with the non-dominant hand (males only); BR$_{Fail}$ = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR$_{Fail}$ = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 545

FTT-NDH$_F$. Table 159 shows signal detection properties for various FTT-NDH$_F$ cutoffs. Cutoffs between ≤16.0 and ≤32.0 produced specificity values between .84 and 1.00. FTT-NDH$_F$ ≤32.0 achieved .39 sensitivity and .84 specificity. Lowering the cutoff to FTT-NDH$_F$ ≤28.0 improved specificity to .91 with acceptable sensitivity (.28). Perfect specificity was attained at FTT-NDH$_F$ ≤16.0, although sensitivity was low (.05).

With regard to previously identified cutoffs, FTT-NDH$_F$ ≤25.0 (Arnold et al., 2005) achieved good specificity (.95) and low sensitivity (.11). FTT-NDH$_F$ ≤19.0 was overly conservative, achieving near-perfect (.99) specificity and low sensitivity (.10).

In settings with a 30% BR$_{Fail}$, the probability that a failure on FTT-NDH$_F$ ≤28.0 represents an invalid profile is .56. The probability that a pass on FTT-NDH$_F$ ≤28.0 represents a valid profile is .74.

Table 159

Signal Detection Properties of Various Finger Tapping Test Non-Dominant Hand Cutoffs (in Females) Achieving Acceptable Specificity against the Word Memory Test

FTT-NDH$_F$	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR$_{Fail}$				
	f	BR$_{Fail}$	SENS	SPEC		10	20	30	40	50
≤16.0	4	1.5	.05	1.00	PPP	.53	.72	.81	.87	.91
					NPP	.90	.81	.71	.61	.51
≤18.0	8	3.0	.10	.99	PPP	.53	.72	.81	.87	.91
					NPP	.91	.81	.72	.62	.52
≤20.0	13	4.9	.13	.98	PPP	.37	.57	.70	.78	.84
					NPP	.91	.82	.72	.63	.53
≤22.0	16	6.0	.16	.97	PPP	.38	.58	.71	.79	.85
					NPP	.91	.82	.73	.64	.54
≤24.0	18	6.8	.18	.97	PPP	.37	.57	.69	.78	.84
					NPP	.91	.82	.73	.64	.54
≤26.0	25	9.4	.20	.94	PPP	.26	.44	.57	.67	.76
					NPP	.91	.82	.73	.64	.54
≤28.0	34	12.8	.26	.91	PPP	.25	.43	.56	.67	.75
					NPP	.92	.83	.74	.65	.55
≤30.0	42	15.8	.31	.89	PPP	.24	.41	.54	.65	.74
					NPP	.92	.84	.75	.66	.56
≤32.0	57	21.4	.39	.84	PPP	.21	.38	.51	.62	.71
					NPP	.93	.85	.76	.67	.58

Note. FTT = Finger Tapping Test (Reitan, 1969); NDH$_F$ = average number of taps achieved with the non-dominant hand (females only); BR$_{Fail}$ = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR$_{Fail}$ = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 266

FTT-C_M. Table 160 shows signal detection properties for various FTT-C_M cutoffs. Cutoffs between ≤35.0 and ≤85.0 achieved specificity values between .84 and 1.00. FTT-C_M ≤85.0 produced .37 sensitivity and .87 specificity. FTT-C_M ≤80.0 reached .91 specificity with a loss in sensitivity (.30). Perfect specificity was achieved at FTT-C_M ≤35.0, although sensitivity was low (.03).

With regard to previously identified cutoffs, FTT-C_M ≤66.0 demonstrated excellent specificity (.96) and low sensitivity (.13) in the present sample.

In a setting with a 30% BR_{Fail}, the probability of an invalid profile is .58 in an examinee who fails FTT-C_M ≤80.0. Conversely, the probability of a valid profile is .75 in someone who passes FTT-C_M ≤80.0.

Table 160

Signal Detection Properties of Various Finger Tapping Test-Combined Cutoffs (in Males) Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
FTT-C_M	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤35.0	6	1.1	.03	1.00	PPP	.57	.75	.84	.89	.92
					NPP	.90	.80	.71	.61	.51
≤40.0	11	2.0	.05	.99	PPP	.42	.62	.74	.81	.87
					NPP	.90	.81	.71	.61	.51
≤45.0	17	3.1	.08	.99	PPP	.47	.66	.77	.84	.89
					NPP	.91	.81	.72	.62	.52
≤50.0	18	3.3	.08	.99	PPP	.41	.61	.73	.81	.86
					NPP	.91	.81	.71	.62	.52
≤55.0	27	5.0	.09	.97	PPP	.25	.43	.57	.67	.75
					NPP	.91	.81	.71	.62	.52
≤60.0	27	5.0	.09	.97	PPP	.25	.43	.57	.67	.75
					NPP	.91	.81	.71	.62	.52
≤65.0	34	6.2	.12	.96	PPP	.25	.43	.57	.67	.75
					NPP	.91	.81	.72	.62	.52
≤70.0	47	8.6	.17	.95	PPP	.27	.45	.58	.69	.77
					NPP	.91	.82	.73	.63	.53
≤75.0	62	11.4	.22	.93	PPP	.26	.44	.57	.68	.76
					NPP	.91	.83	.74	.64	.54
≤80.0	83	15.2	.30	.91	PPP	.26	.44	.58	.68	.76
					NPP	.92	.84	.75	.66	.56
≤85.0	111	20.4	.37	.87	PPP	.23	.41	.54	.65	.73
					NPP	.93	.85	.76	.67	.58

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (males only); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 545

FTT-C_F. Table 161 shows signal detection properties for various FTT-C_F cutoffs. Cutoffs between ≤30.0 and ≤65.0 produced specificity values between .84 and 1.00. FTT-C_F ≤65.0 achieved .38 sensitivity and .85 specificity. Excellent specificity (.96) was attained by lowering the cutoff to FTT-C_F ≤55.0, with a loss in sensitivity (.26). FTT-C_F ≤30.0 produced perfect specificity with low sensitivity (.03).

With regard to previously identified cutoffs, FTT-C_F ≤58.0 achieved acceptable specificity (.91) and sensitivity (.26).

In a setting with a 30% BR_{Fail}, there is a .62 probability that a failure on FTT-C_F ≤55.0 represents an invalid profile, and a .75 probability that a pass on FTT-C_F ≤55.0 reflects a valid profile.

Table 161

Signal Detection Properties of Various Finger Tapping Test-Combined Cutoffs (in Females) Achieving Acceptable Specificity against the Word Memory Test

	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
FTT-C_F	*f*	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤30.0	3	1.1	.03	1.00	PPP	.43	.63	.74	.82	.87
					NPP	.90	.80	.71	.61	.51
≤35.0	5	1.9	.05	.99	PPP	.36	.56	.68	.77	.83
					NPP	.90	.81	.71	.61	.51
≤40.0	9	3.4	.11	.99	PPP	.57	.75	.83	.89	.92
					NPP	.91	.82	.72	.63	.53
≤45.0	13	4.9	.16	.99	PPP	.55	.74	.83	.88	.92
					NPP	.91	.83	.73	.64	.54
≤50.0	19	7.1	.21	.97	PPP	.45	.65	.76	.83	.88
					NPP	.92	.83	.74	.65	.55
≤55.0	30	11.3	.26	.96	PPP	.30	.49	.62	.72	.79
					NPP	.92	.83	.75	.65	.56
≤60.0	42	15.8	.28	.88	PPP	.20	.36	.49	.60	.70
					NPP	.92	.83	.74	.65	.55
≤65.0	53	19.9	.38	.85	PPP	.22	.39	.52	.63	.72
					NPP	.92	.85	.76	.67	.58

Note. FTT = Finger Tapping Test (Reitan, 1969); C_M = sum of average number of taps achieved with the dominant and non-dominant hands (females only); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power.
[a]*n* = 266

FTT-DIFF$_M$. Table 162 shows signal detection properties for various FTT-DIFF$_M$ cutoffs. Cutoffs between FTT-DIFF$_M$ ≤-27.0 and ≤-1.0 produced specificity values between .84 and 1.00. FTT-DIFF$_M$ ≤-1.0 achieved .24 sensitivity and .85 specificity. Improved specificity (.91) was observed FTT-DIFF$_M$ ≤-3.0, with a loss in sensitivity (.14). Lowering the cutoffs further produced small gains in specificity and small losses in sensitivity. Thus, only classification accuracy data for cutoffs up to FTT-DIFF$_M$ ≤-12.0 are shown in Table 158. Additional analyses revealed that FTT-DIFF$_M$ ≤-27.0 was associated with perfect specificity, although there was zero sensitivity to invalid performance at this cutoff (BR_{Fail} = 0.2%).

With regard to previously identified cutoffs, FTT-DIFF$_M$ ≤-2.0 achieved .88 specificity and .17 sensitivity in the present sample.

PPP and NPP indicate that in a setting with a 30% BR_{Fail}, there is a .41 probability of an invalid profile given a failure on FTT-DIFF$_M$ ≤-3.0, and a .71 probability of a valid profile given a pass on FTT-DIFF$_M$ ≤-3.0.

Table 162

Signal Detection Properties of Various Finger Tapping Test-Difference cutoffs (in Males) Achieving Acceptable Specificity against the Word Memory Test

FTT-$DIFF_M$	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR_{Fail}				
	f	BR_{Fail}	SENS	SPEC		10	20	30	40	50
≤-12.0	11	2.0	.04	.99	PPP	.24	.42	.56	.66	.74
					NPP	.90	.80	.71	.61	.51
≤-11.0	12	2.2	.04	.98	PPP	.21	.38	.51	.62	.71
					NPP	.90	.80	.70	.61	.51
≤-10.0	13	2.4	.04	.98	PPP	.19	.34	.47	.58	.68
					NPP	.90	.80	.70	.60	.51
≤-9.0	16	2.9	.04	.98	PPP	.17	.32	.45	.56	.65
					NPP	.90	.80	.70	.61	.51
≤-8.0	19	3.5	.05	.97	PPP	.16	.31	.43	.54	.64
					NPP	.90	.80	.70	.61	.51
≤-7.0	24	4.4	.06	.96	PPP	.16	.30	.43	.54	.63
					NPP	.90	.80	.71	.61	.51
≤-6.0	31	5.7	.08	.95	PPP	.16	.30	.43	.54	.64
					NPP	.90	.81	.71	.61	.51
≤-5.0	34	6.2	.09	.95	PPP	.18	.32	.45	.56	.66
					NPP	.90	.81	.71	.61	.51
≤-4.0	44	8.1	.11	.93	PPP	.16	.30	.42	.53	.63
					NPP	.90	.81	.71	.61	.51
≤-3.0	58	10.6	.14	.91	PPP	.15	.29	.41	.52	.61
					NPP	.91	.81	.71	.61	.52
≤-2.0	73	13.4	.17	.88	PPP	.14	.26	.38	.49	.59
					NPP	.91	.81	.71	.61	.51
≤-1.0	97	17.8	.24	.85	PPP	.15	.28	.40	.51	.61
					NPP	.91	.82	.72	.63	.53

Note. FTT = Finger Tapping Test (Reitan, 1969); $DIFF_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands (males only); BR_{Fail} = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR_{Fail} = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 545

FTT-DIFF$_F$. Table 163 shows signal detection properties for various FTT-DIFF$_F$ cutoffs. Cutoffs between ≤-21.0 and ≤-1.0 produced specificity values between .85 and 1.00. FTT-DIFF$_F$ ≤-1.0 achieved .28 sensitivity and .85 specificity. Improved specificity (.90) was attained using FTT-DIFF$_F$ ≤-3.0, with a loss in sensitivity (.20). Lowering cutoffs further resulted in steady losses in sensitivity and small gains in specificity. Thus, only classification accuracy data for cutoffs up to FTT-DIFF$_M$ ≤-11.0 are shown in Table 159.

Additional analyses revealed that FTT-DIFF$_M$ ≤-16.0 produced .99 specificity and .07 sensitivity (BR_{Fail} = 2.6%). Perfect specificity was achieved using FTT-DIFF$_M$ ≤-21.0, although sensitivity to invalid performance was very low (.03; BR_{Fail} = 0.8%).

With regard to previously identified cutoffs, FTT-DIFF$_F$ ≤-5.0 demonstrated excellent specificity (.96) and acceptable sensitivity (.18) in the present sample.

In a setting with a 30% BR_{Fail}, there is a .48 probability of an invalid profile in an examinee who fails FTT-DIFF$_F$ ≤-3.0, and a .73 probability of a valid profile in an examinee who passes FTT-DIFF$_F$ ≤-3.0.

Table 163

Signal Detection Properties of Various Finger Tapping Test-Difference Cutoffs (in Females) Achieving Acceptable Specificity against the Word Memory Test

FTT-DIFF$_F$	Overall Sample[a]		Criterion PVT: WMT			Hypothetical BR$_{Fail}$				
	f	BR$_{Fail}$	SENS	SPEC		10	20	30	40	50
≤-11.0	10	3.8	.08	.98	PPP	.27	.46	.59	.69	.77
					NPP	.91	.81	.71	.61	.52
≤-10.0	12	4.5	.10	.97	PPP	.27	.46	.59	.69	.77
					NPP	.91	.81	.72	.62	.52
≤-9.0	13	4.9	.11	.97	PPP	.30	.49	.63	.72	.80
					NPP	.91	.81	.72	.62	.52
≤-8.0	16	6.0	.16	.97	PPP	.38	.58	.71	.79	.85
					NPP	.91	.82	.73	.64	.54
≤-7.0	17	6.4	.16	.97	PPP	.35	.55	.67	.76	.83
					NPP	.91	.82	.73	.63	.54
≤-6.0	19	7.1	.16	.96	PPP	.29	.48	.62	.71	.79
					NPP	.91	.82	.73	.63	.53
≤-5.0	20	7.5	.18	.96	PPP	.31	.51	.64	.73	.80
					NPP	.91	.82	.73	.64	.54
≤-4.0	27	10.2	.21	.93	PPP	.26	.44	.57	.68	.76
					NPP	.91	.83	.73	.64	.54
≤-3.0	33	12.4	.20	.90	PPP	.20	.35	.48	.59	.69
					NPP	.91	.82	.73	.63	.53
≤-2.0	39	14.7	.25	.88	PPP	.19	.34	.47	.58	.68
					NPP	.91	.82	.73	.64	.54
≤-1.0	47	17.7	.28	.85	PPP	.17	.32	.45	.56	.66
					NPP	.91	.83	.73	.64	.54

Note. FTT = Finger Tapping Test (Reitan, 1969); DIFF$_F$ = difference between the average number of taps achieved with the dominant and non-dominant hands (females only); BR$_{Fail}$ = base rate of failure (percent of the sample that failed the validity cutoff); PVT = performance validity test; WMT = Word Memory Test (Green, 2003); SENS = sensitivity; SPEC = specificity; Hypothetical BR$_{Fail}$ = prevalence of invalid performance in the assessment setting (in percent); PPP = positive predictive power; NPP = negative predictive power. Boldface font denotes previously published cutoffs.
[a]*n* = 266

Relative classification accuracies of PVTs. To determine the relative diagnostic power of the PVTs examined in Study 3, the most liberal cutoff achieving at least 90% specificity was identified on each PVT, and sensitivity values corresponding to those cutoffs were compared between PVTs. Results from these analyses are summarized in Table 164, which shows sensitivity rates for all of the PVTs examined, in order from highest to lowest. Specifically, free-standing PVTs demonstrated the highest sensitivity to invalid performance, although they only detected about half of the cases identified as invalid by the WMT (SENS =.44-.57). Embedded validity indicators produced lower sensitivity compared to free-standing PVTs; sensitivity rates varied considerably, ranging from .35 (using TMT A+B ≥165) to .09 (using TMT-B/A <1.58).

Table 164

Classification Accuracy of All PVTs at the Most Liberal Cutoff to achieve a ≤10% False Positive Error Rate

PVT	Cutoff[a]	SENS	SPEC
TOMM Trial 1	≤43	.57	.90
MSVT	STN	.52	.93
NV-MSVT	STN	.45	.93
TOMM Trial 2	≤48	.44	.96
TMT A+B	≥165	.35	.90
TMT-B	≥120	.33	.90
FTT-DH_F	≤29.6	.33	.90
FTT-DH_M	≤42.8	.33	.90
TMT-A	≥49	.31	.90
FTT-C_M	≤81.0	.31	.90
FTT-NDH_F	≤29.0	.28	.90
FTT-C_F	≤59.0	.26	.90
FTT-NDH_M	≤38.0	.25	.91
CT-TE	≥100	.23	.90
FTT-$DIFF_F$	≤-3.0	.20	.90
WCST FMS	≥4	.17	.94
RDS	≤6	.17	.96
FTT-$DIFF_M$	≤-2.6	.15	.90
TMT-B/A	<1.58	.09	.90

Note. PVT = performance validity test; SENS = sensitivity to invalid performance; TOMM = Test of Memory Malingering (Tombaugh, 1996); MSVT = Medical Symptom Validity Test (Green, 2004); STN = standard cutoffs identified in the test manual; NV-MSVT = Non-Verbal Medical Symptom Validity Test (Green, 2008); TMT A+B = Sum of completion times (in seconds) on Trail Making Test Part A and Part B (Reitan & Wolfson, 1985); TMT-B = Trail Making Test (Reitan & Wolfson, 1985) Part B completion time; FTT-DH_F = average number of taps achieved with the dominant hand on the Finger Tapping Test (Reitan, 1969; females only); TMT-A = Trail Making Test (Reitan & Wolfson, 1985) Part A completion time in seconds; FTT-C_M = sum of average number of taps achieved with the dominant and non-dominant hands on the Finger Tapping Test (Reitan, 1969; males only); FTT-DH_M = average number of taps achieved with the dominant hand on the Finger Tapping Test (Reitan, 1969; males only); FTT-NDH_F = average number of taps achieved with the non-dominant hand on the Finger Tapping Test (Reitan, 1969; females only); FTT-C_F = sum of average number of taps achieved with the dominant and non-dominant hands on the Finger Tapping Test (Reitan, 1969; females only); FTT-NDH_M = average number of taps achieved with the non-dominant hand on the Finger Tapping Test (Reitan, 1969; males only); CT-TE = Total Errors on the Category Test (DeFlippis & McCampbell, 1979, 1991, 1997); FTT- $DIFF_F$ = difference between the average number of taps achieved with the dominant and non-dominant hands on the Finger Tapping Test (Reitan, 1969; females only); WCST FMS = Wisconsin Card Sorting Test, Failure to Maintain Set (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); RDS = Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994); FTT- $DIFF_M$ = difference between the average number of taps achieved with the dominant and non-dominant hands on the Finger Tapping Test (Reitan, 1969; males only); TMT-B/A = Trail Making Test ratio (Part B completion time in seconds/Part A completion time in seconds; Reitan & Wolfson, 1985).

[a]Most liberal cutoff on the respective PVT that achieved a specificity of at least .90 in the present sample

CHAPTER 4

Discussion

This dissertation aimed to characterize performance on 14 PVTs of interest (three free-standing PVTs and 11 embedded validity indicators) in a predominantly medical-legal sample. BR_{Fail} were reported as a function of various patient variables (i.e., diagnosis, age, education, gender, and English language status) in the overall sample (Study 1) and in a subsample of patients who passed the WMT (Green, 2003), a well-validated criterion PVT, at standard cutoffs (valid sample; Study 2). Classification accuracy statistics were also computed against the WMT for various cutoffs on each PVT of interest to determine the range of cutoffs achieving acceptable specificity (Study 3). The major findings from this dissertation, concerning (1) the criterion PVT, (2) Studies 1 and 2, and (3) Study 3, are discussed below in sequence.

The Criterion PVT

In the present dissertation, the WMT was used as a criterion PVT to demonstrate the global effect of invalid performance across PVTs of interest. The overall BR_{Fail} on the WMT was 30.2%, which is consistent with previous estimates for settings with a mixture of clinical and forensic referrals (i.e., 30%; Mittenberg et al., 2002) and slightly lower than estimates for exclusively medical-legal settings (i.e., 40%; Larrabee, 2003). Examinees who passed the WMT had significantly lower BR_{Fail} than those who failed the WMT on nearly all PVTs of interest including the MSVT, NV-MSVT, TOMM Trials 1 and 2, RDS, WCST FMS, CT-TE, TMT-A, TMT-B, TMT A+B, FTT-DH_M, FTT-NDH_M, and FTT-C_M. Although not statistically significant, a similar trend was observed on FTT-$DIFF_M$. These results indicate that BR_{Fail} on a wide range of free-standing PVTs and embedded validity indicators are related to invalid performance as

measured by the WMT. In doing so, they provide support for the construct validity of the aforementioned PVTs of interest.

Interestingly, those who passed the WMT and those who failed the WMT demonstrated largely comparable BR_{Fail} on TMT-B/A. The fact that TMT-B/A BR_{Fail} were not related to WMT outcome may be due to these measures' different approaches to performance validity assessment (Erdodi, 2017). On the WMT, the decision about performance validity is based on level of accuracy and consistency of responses (Green, 2003), while TMT-B/A outcomes are based whether examinees demonstrate the expected differences in performance on an easier versus more difficult task (Iverson et al., 2002).

Predictably, as a whole, the valid sample demonstrated lower BR_{Fail} than the overall sample across cutoffs on nearly all PVTs of interest, with the exception of TMT-B/A.

Studies 1 and 2

Studies 1 and 2 yield large-scale normative data for a wide range of PVTs in a large overall medical-legal sample and in a subsample screened for valid performance using an independent criterion. By reporting BR_{Fail} as a function of relevant patient variables, these data help to provide an empirical basis for the interpretation of individual PVT failures by clinicians. For example, using these data, a clinician can compare an examinee's PVT score to one or more reference groups (i.e., overall sample, valid sample, as well as various diagnostic and/or demographic groups) in order to better understand how commonly such a score is observed in these groups and, ultimately, decide how the examinee's score should interpreted.

Specific findings concerning the relationships between patient variables and BR_{Fail} on PVTs of interest are discussed under the relevant sub-headings below. For each patient variable,

findings involving free-standing PVTs are discussed first, followed by those involving embedded validity indicators. FTT-based validity indicators are discussed under a separate heading.

The effect of diagnosis on PVT failure rates. With regard to free-standing PVTs of interest, in the overall sample, diagnosis was related to BR_{Fail} on the WMT, MSVT, NV-MSVT, and TOMM Trials 1 and 2, although the effects were small in magnitude: patients with mTBI, chronic pain/fibromyalgia, and orthopedic injuries tended to demonstrate the highest BR_{Fail}, while those with moderate-severe TBI had low BR_{Fail}. These results indicate that, in the overall sample, groups expected to have genuine cognitive impairment secondary to known and/or persisting structural brain damage (e.g., those with moderate-severe TBI) demonstrate lower BR_{Fail} than groups that are not expected to have such impairment (i.e., chronic pain/fibromyalgia, mTBI). Although paradoxical, this finding is consistent with the existing literature. In TBI, for example, higher BR_{Fail} are frequently observed among those with mTBIs as compared to moderate-severe TBIs, particularly in medical-legal or forensic settings (Carone, 2008; Erdodi & Rai, 2017; Grote et al., 2000; Sweet, Goldman, & Guidotti Breting, 2013). Given that full neuropsychological recovery is expected to occur within three months following mTBI (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005), the higher BR_{Fail} among forensic mTBI cases have been attributed to various external factors including diagnosis threat (e.g., Suhr & Gunstad, 2002) and incentive to appear impaired (e.g., Binder & Rohling, 1996) rather than genuine and persisting cognitive impairment. Similarly, patients with chronic pain/fibromyalgia and orthopedic injuries frequently complain of cognitive difficulties (Bennett, 1996; Glass & Park, 2001; Grace, Neilson, Hopkins, & Berg, 1999; Iverson & McCracken, 1997; Schnurr & MacDonald, 1995) and have been found to demonstrate substantial rates of PVT failure in a number of studies (Gervais, Rohling, Green, & Ford, 2004; Gervais, Russell, Green, et al., 2001;

Greve, Ord, Bianchini, & Curtis, 2009; Meyers & Diep, 2000; Meyers & Volbrecht, 2003; Schmand et al., 1998; Suhr, 2003). As in TBI, various emotional and motivational factors (e.g., comorbid anxiety or depression, negative expectations/attributions) may contribute to poor neuropsychological performance among those with chronic pain or fibromyalgia (Binder, 2012; Greve, Bianchini, & Ord, 2012). However, financial incentives appear to be particularly influential (e.g., Gervais et al., 2004; Mittenberg et al., 2002). Consistent with this, the rate of PVT and SVT failures among pain patients has been found to be associated in a dose-response fashion with increasing evidence of malingering and not with injury characteristics (Greve, Bianchini et al., 2009; Greve, Etherton et al., 2009; Greve, Ord et al., 2009).

Results showing higher BR_{Fail} among groups expected to have no known and/or persisting brain dysfunction (e.g., chronic pain/fibromyalgia, mTBI) relative to groups with neurologically-based cognitive impairment (e.g., moderate-severe TBI, neurological conditions) are also consistent with the findings of Mittenberg and colleagues' (2002) survey of neuropsychologists. In that study, the highest base rates of malingering/symptom exaggeration were reported among litigating or compensation-seeking patients with mTBI, fibromyalgia/chronic fatigue, and pain or somatoform disorders, while some of the lowest base rates of malingering/symptom exaggeration were reported among those with moderate or severe head injury and vascular dementia. Notably, however, survey respondents in the Mittenberg et al. (2002) study used multiple sources of evidence to identify malingering/symptom exaggeration including: PVT failure; severity or pattern of cognitive impairment that was inconsistent with the condition; discrepancies among records, self-report, and observed behaviour; implausible self-reported symptoms in interview; implausible changes in test scores across repeated examinations; validity scales on objective personality tests.

Interestingly, in the valid sample, diagnosis was unrelated to BR_{Fail} on the MSVT and TOMM Trial 1. These findings suggest that, when used with credibly-performing examinees, the MSVT and TOMM Trial 1 are not associated with an elevated risk of false positives in any particular diagnostic groups relative to others. Unfortunately, the relationship between diagnosis and BR_{Fail} on TOMM Trial 2 or the NV-MSVT in the valid sample remains unclear as no chi-square or Fisher's exact tests were conducted in this sample.

With regard to embedded validity indicators, BR_{Fail} varied significantly as a function of diagnosis on CT-TE, TMT-A, TMT-B, and TMT-A+B in both the overall and valid samples, although effect sizes were small. Groups expected to have genuine cognitive impairment (i.e., those with moderate-severe TBI, neurological conditions, and severe mental illness) were among those with the highest BR_{Fail} on several of these validity indicators. The fact that BR_{Fail} continued to vary significantly as a function of diagnosis when cases failing the WMT were removed (i.e., in the valid sample) suggests a risk for increased false positives when using these indicators to assess performance validity in cognitively impaired groups. Given that the CT and TMT were originally designed to detect cognitive dysfunction rather than performance validity (Choca & Morris, 1992; DeFlippis & McCampbell, 1979, 1991, 1997; Reitan & Wolfson, 1985), these results are not surprising. Consistent with this, the potential sensitivity of embedded validity indicators to genuine cognitive impairment (e.g., secondary to neurological and neurodevelopmental conditions) has been identified as a risk associated with their use over free-standing PVTs (Boone, 2007). It has been recommended that such indicators be supplemented with one or more free-standing PVTs in practice.

On RDS, patients expected to be cognitively impaired generally demonstrated lower BR_{Fail} compared to groups that are unlikely to have such impairment, suggesting that this

indicator may be more resistant to the impact of true cognitive dysfunction than the embedded validity indicators discussed above. At RDS ≤6, for example, the highest BR_{Fail} were observed among the orthopedic injury, chronic pain/fibromyalgia, and mTBI groups in the overall sample, and among the orthopedic injury and mTBI groups in the valid sample. Patients with moderate-severe TBI demonstrated the lowest BR_{Fail} in both samples. This pattern of results suggests that RDS is largely unaffected by true cognitive dysfunction at a cutoff of ≤6 and therefore likely represents an effective validity indicator at that level. Notably, however, elevated BR_{Fail} were observed among neurological patients at RDS ≤7 in both the overall and valid samples, suggesting an increased risk of false positives in this group at this cutoff. The latter finding is consistent with past research showing that a cutoff of RDS ≤7 produces unacceptably high false positives in several clinical groups including patients with cerebrovascular accidents (Heinly et al., 2005).

BR_{Fail} did not vary significantly as a function of diagnosis on WCST FMS or TMT-B/A in either the overall or valid samples. These results suggest that WCST FMS and TMT-B/A are appropriate for use with a variety of clinical groups and do not carry an increased risk of false-positive errors in one or more diagnostic groups relative to others.

The effect of age on PVT failure rates. In the overall sample, BR_{Fail} were significantly associated with age on all of the free-standing PVTs (i.e., WMT, MSVT, NV-MSVT, TOMM Trials 1 and 2), although effects were small in magnitude. On each PVT, BR_{Fail} tended to increase with older age. Given the prevalence of invalid performance in the overall sample, however, it remains unclear whether this finding reflects (a) higher rates of invalid performance among older examinees, (b) higher false positive rates among older examinees, or (c) both.

In the valid sample, age was unrelated to BR_{Fail} on TOMM Trials 1 and 2, with the exception of Trial 2 ≤49, where the oldest group demonstrated almost three times the BR_{Fail} observed among the other age groups. MSVT and NV-MSVT BR_{Fail} also varied significantly as a function of age in the valid sample, with comparable BR_{Fail} among those aged 16 to 49 years and higher BR_{Fail} in older examinees. Taken together, these results suggest an elevated risk for false positive errors among older examinees when using TOMM Trial 2 ≤49 and the MSVT and NV-MSVT (at standard cutoffs).

The fact that BR_{Fail} on the MSVT, NV-MSVT, and TOMM Trial 2 ≤49 were related to age in the valid sample is somewhat surprising as free-standing PVTs are often thought to be resistant to the effects of older age. BR_{Fail} on the MSVT and NV-MSVT were likely related to age in the present dissertation because only the standard cutoffs were used to classify patients as valid versus invalid on these instruments; the GMIP criteria were not used. As discussed above, the GMIPs distinguish severe cognitive impairment from non-credible performance in cases where standard cutoffs on the MSVT and NV-MSVT are failed (Green, 2004; Green, 2008) and have been shown to reduce false positives among older adults with memory impairment and dementia in a number of studies (Henry et al., 2009; Howe et al., 2007; Howe & Loring, 2009; Green, 2011; Singhal et al., 2009). Considering this, it is possible that the elevated BR_{Fail} among older examinees in the valid sample included false positive cases that may have met GMIP criteria and ultimately been classified as valid.

With regard to TOMM Trial 2, previous research found the standard cutoff (Trial 2 ≤44) to produce adequate specificity in older adults with mild cognitive impairment (Teichner & Wagner, 2004; Walter et al., 2014). The current findings extend past research by showing that, although age is not meaningfully related to TOMM Trial 2 BR_{Fail} up to a cutoff of Trial 2 ≤48,

the most liberal conceivable cutoff (i.e., Trial 2 ≤49) is associated with an increased risk for false positive errors among older examinees.

With regard to embedded validity indicators, age was associated with BR_{Fail} on RDS, CT-TE, TMT-A, TMT-B, and TMT A+B in the overall sample with small to medium effect sizes. Similar results were observed in the valid sample, although RDS BR_{Fail} only varied significantly as a function of age at ≤7. In both samples, and across PVTs, the youngest age group tended to have the lowest BR_{Fail}, and BR_{Fail} increased with age. Although previous research has not investigated the effects of age on these scores as indicators of performance validity specifically, there is strong evidence for age effects in the tests' normative samples (Heaton, Miller, Taylor, & Grant, 2004; Wechsler, 2008). Consistent with this, the current findings indicate that higher BR_{Fail} are predictably observed in older examinees compared to younger examinees when these tests are co-opted as PVTs. The fact that the relationship between age and BR_{Fail} on these validity indicators persisted in the valid sample suggests an increased risk for false positive errors when using these indicators to assess performance validity among older examinees. Notably, given that RDS BR_{Fail} were only associated with age at ≤7 in the valid sample, using a more conservative cutoff on this indicator (e.g., RDS ≤6) is likely to protect against false positive errors.

Follow-up analyses suggested that the results observed across the aforementioned PVTs with regard to age may have been confounded by education level to some degree. Specifically, mean education level was approximately one year lower among the oldest age group (60-69 years) compared to the other groups. Therefore, the elevated BR_{Fail}, which were observed in the oldest group across many PVTs of interest, likely reflected a combination of older age and slightly lower education level.

On WCST FMS and TMT-B/A, BR_{Fail} did not vary as a function of age in either sample. These findings suggest that these indicators can be used to assess performance validity across examinees aged 16 to 69 years, without elevated risks of false positives in some age groups relative to others.

The effect of education level on PVT failure rates. With regard to free-standing PVTs, education was associated with BR_{Fail} on the WMT, MSVT, NV-MSVT, and TOMM Trial 1 in the overall sample, although the effects were small in magnitude. On each PVT, the least educated group tended to have the highest BR_{Fail}, and BR_{Fail} decreased with higher levels of education. Again, given the prevalence of invalid performance in the overall sample, it remains unclear whether these findings reflect (a) higher rates of invalid performance among less educated examinees, (b) higher false positive rates among less educated examinees, or (c) both. TOMM Trial 2 BR_{Fail} were not related to education level in the overall sample.

In the valid sample, BR_{Fail} were not meaningfully related to education level on any of the free-standing PVTs of interest. These results suggest that the MSVT, NV-MSVT, and TOMM are appropriate for use with examinees with a wide range of educational levels, and lower levels of education are not associated with an increased risk of false positives on these PVTs among credibly-performing examinees.

With regard to embedded validity indicators, education level was related to BR_{Fail} on RDS, CT-TE, TMT-A, TMT-B, and TMT A+B in both samples such that the highest BR_{Fail} tended to be observed in the least educated groups. The fact that BR_{Fail} continued to vary as a function of education level in the valid sample suggests an increased risk for false positive errors when using these indicators to assess performance validity among examinees with low levels of education. Notably, given that TMT-A BR_{Fail} were only meaningfully associated with education

level at the most liberal cutoff examined (≥43) in the valid sample, using more conservative TMT-A cutoffs may help protect against false positive errors in examinees with low levels of education.

Consistent with the above findings, examination of demographically-corrected normative data for the Category Test (Heaton et al., 2004), Trail Making Test (Heaton et al., 2004), and WAIS Digit Span (Pearson, 2009) indicates that more educated examinees perform better on these measures than less educated examinees: on each of these tests, the same raw score corresponds to lower standardized scores (reflecting worse performance) among examinees with more years of education and higher standardized scores (reflecting better performance) among examinees with fewer years of education.

The relationship between WCST FMS BR_{Fail} and education level remains somewhat unclear due to the relatively small sample sizes across cutoffs in both samples and, in particular, among the least educated group (≤8 years). Nonetheless, the pattern of BR_{Fail} across examinees with 9 to ≥17 years of education suggests that lower levels of education (i.e., ≤12 years) are associated with higher BR_{Fail} and an increased risk for false positive errors when using WCST FMS as a validity indicator.

While the results of Studies 1 and 2 suggest that WCST FMS may vary as a function of education level, FMS performance does not appear to be related to education level in the WCST normative sample. Based on information reported in the test manual (Heaton et al., 1993), the relationship between FMS raw scores and education was not investigated in the normative groups. However, age-adjusted FMS scores were not related to education level. Furthermore, due to low sample sizes in some age-by-education cells, normative data for FMS were generated using only age. Thus, FMS raw scores correspond to the exact same percentile equivalents

(standardized scores) for all examinees within a given age bracket, regardless of their education levels. It is possible that FMS performance was not related to education level in the WCST normative sample because the authors used age-adjusted FMS scores rather than raw scores. Given the findings of Studies 1 and 2, however, the relationship between WCST FMS raw scores and education level requires further investigation in future studies.

While BR_{Fail} generally decreased with higher levels of education on the aforementioned PVTs, unexpectedly high BR_{Fail} were observed among one or both of the most highly educated groups (16 years and ≥17 years) on a number of PVTs including the MSVT, TOMM Trials 1 and 2, CT-TE, TMT-A, TMT-B, and TMT A+B. This pattern was observed in both the overall and valid samples and, across most PVTs, the elevated BR_{Fail} were more pronounced in those with ≥17 years of education than those with 16 years of education.

To further investigate these unexpected findings, follow-up analyses were conducted, which identified age and diagnosis as potentially significant confounds across education groups. Specifically, the average age of examinees with ≤8, 16, and ≥17 years of education was approximately five years older than the average age of participants in the remaining groups. Additionally, the groups with 16 and ≥17 years of education contained larger proportions of patients with neurological conditions compared to other education groups and therefore may have been more cognitively impaired. Taken together, these results suggest that the BR_{Fail} for the least educated group reflected a combination of lower education level and older age, while the BR_{Fail} for the two most educated groups were likely impacted by older age and more severe cognitive impairment.

On TMT-B/A, BR_{Fail} were positively associated with education level in both samples: the *lowest* BR_{Fail} were observed in patients with the lowest levels of education, and BR_{Fail} *increased*

with higher levels of education. Although paradoxical, this finding indicates that patients with higher levels of education complete TMT-B much more quickly, relative to TMT-A, than those with lower levels of education. Thus, a higher education level may confer a greater advantage to examinees on TMT-B as compared to TMT-A. Consistent with this hypothesis, a stronger negative correlation was reported between education level and TMT-B completion time as compared to education level and TMT-A completion time among healthy, community-dwelling adults (Tombaugh, 2004). Taken together, these findings indicate that although TMT-B/A is appropriate for use with examinees with lower levels of education, there may be an increased risk of false positive errors when using this validity indicator with more educated examinees.

The effect of gender on PVT failure rates. In the overall sample, BR_{Fail} on the WMT, MSVT, and NV-MSVT were related to gender, with males demonstrating slightly higher BR_{Fail} than females, while BR_{Fail} on TOMM Trials 1 and 2 were not. Among credibly-performing examinees (i.e., in the valid sample), there was no relationship between gender and BR_{Fail} on any of the free-standing PVTs of interest.

With regard to embedded validity indicators, there was some evidence for a gender effect on WCST FMS in the overall sample: although BR_{Fail} were not associated with gender at WCST FMS ≥2, males were more likely than females to fail WCST FMS ≥3. Gender was not related to WCST FMS BR_{Fail} at either cutoff in the valid sample. Additionally, there were no significant relationships between gender and BR_{Fail} on RDS, CT-TE, TMT-A, TMT-B, TMT A+B, or TMT-B/A in either sample.

Taken together, the results indicate that, among credibly-performing examinees, there are no meaningful relationships between gender and BR_{Fail} across any of the PVTs of interest

examined. As such, the free-standing PVTs and embedded validity indicators examined in this dissertation are appropriate for assessing performance validity among both males and females.

The effect of English language status on PVT failure rates. In the overall sample, NSEs demonstrated significantly lower BR_{Fail} than ESL patients on all free-standing PVTs, including the WMT. As suggested above, this finding may reflect higher rates of invalid performance among ESL examinees and/or higher rates of false positives among ESL examinees.

When cases failing the WMT were removed (i.e., in the valid sample), MSVT BR_{Fail} did not differ meaningfully as a function of English language status: NSEs and ESL patients demonstrated generally comparable BR_{Fail}. The fact that BR_{Fail} on a verbal task like the MSVT were not related to English language status is unexpected. However, this finding may be due to the use of the WMT as the criterion to establish the valid sample. Both the WMT and MSVT require a grade 3 reading level (Green, 2003; Green, 2004). Considering this, ESL examinees who were included in the valid sample possessed, at the very least, this level of reading ability and therefore were able to pass the MSVT at comparable rates to their NSE counterparts.

Interestingly, NSEs did demonstrate significantly lower BR_{Fail} as compared to ESL patients on non-verbal free-standing PVTs (i.e., the NV-MSVT and TOMM Trials 1 and 2) in the valid sample. These results suggest a risk for false-positive errors when using these instruments to assess performance validity among ESL patients.

With regard to embedded validity indicators, NSEs demonstrated significantly lower BR_{Fail} than ESL patients on RDS, CT-TE, TMT-A, TMT-B, and TMT A+B, in both the overall and valid samples. The fact that BR_{Fail} continued to vary as a function of English language status when cases failing the WMT were excluded (i.e., in the valid sample) suggests a risk for false

positive errors when using these indicators to assess performance validity among ESL patients. The finding involving higher RDS BR_{Fail} among ESL examinees is consistent with previous research in which various RDS cutoffs achieved lower levels of specificity among ESL examinees as compared to Caucasians (Salazar et al., 2007). It is also consistent with past studies showing that ESL status affects performance on various neuropsychological tasks involving high verbal mediation (e.g., Boone et al., 2007).

Although ESL patients are often expected to perform more poorly than NSEs on verbal tasks specifically, the results of Studies 1 and 2 indicate that (1) ESL examinees also fail PVTs involving non-verbal or visuospatial stimuli at higher rates than NSEs (i.e., NV-MSVT, TOMM, CT-TE, TMT), and (2) these differences in BR_{Fail} continue to be observed after examinees demonstrating non-credible performance on an external criterion PVT are excluded from the analyses. The elevated BR_{Fail} observed among ESL patients on visuospatial and non-verbal PVTs, particularly in the valid sample, may be the result of cultural differences. Although neuropsychologists have long believed that the use of visuospatial and non-verbal tests ameliorates the impact of culture on testing, it has been argued that such instruments are not "culture free" (Rosselli & Ardila, 2003). Indeed, skills that are considered to be universal among neurotypical adults (e.g., those involving writing or drawing, or remembering abstract information, for example) may be less emphasized or entirely absent in some cultures (Ardila & Moreno, 2001; Berry, Poortinga, & Segall, 1992; Irvine & Berry, 1988). Furthermore, other cultures may not value "fast performance" in the same way that the North American culture does, which would affect performance on tests such as TMT (e.g., Leon-Carrion, 1989). In addition to cultural differences, factors such as educational experience, educational level, level of acculturation, and socioeconomic status may contribute to the lower neuropsychological test

scores (and higher PVT BR_{Fail}) observed among ESL patients relative to NSEs (Manly, Byrd, Touradji, & Stern, 2004; Mehta et al., 2004; Perez-Arce & Puente, 1996).

BR_{Fail} on WCST FMS and TMT-B/A were unrelated to English language status in either sample, suggesting that these indicators are equally appropriate for use with NSEs and ESL patients.

FTT-based validity indicators. BR_{Fail} on FTT-based validity indicators among males were reported as part of Studies 1 and 2. Because the FTT was not as widely administered in the current data set as other measures used in this dissertation, sample sizes were small at most cutoffs in the overall sample and all cutoffs in the valid sample. Not surprisingly, further stratification of the data as a function of various patient variables (e.g., diagnosis, age, education) resulted in several small cell sizes across analyses. In light of this, some estimates of BR_{Fail} reported for FTT-based validity indicators in Studies 1 and 2 may be unreliable. Additionally, the relationships between BR_{Fail} and patient variables remain somewhat unclear as null-hypothesis significance tests were not conducted at many of the cutoffs examined and, where such tests were conducted, they were likely underpowered. Considering these issues, further investigation of the relationships between FTT-based validity indicators and patient variables is warranted with larger samples. Nonetheless, the findings from Studies 1 and 2 are discussed below.

The relationship between BR_{Fail} on FTT-based validity indicators and diagnosis remains unclear due to the small sizes of several clinical groups, although effect sizes suggested small effects where computed (i.e., at one cutoff each on FTT-DH_M and FTT-C_M).

BR_{Fail} on FTT-DH_M, FTT-NDH_M, and FTT-C_M appeared to be related to age in both samples, with small to medium effect sizes. The youngest group tended to have the lowest BR_{Fail}, and BR_{Fail} increased with older age. The fact that FTT-DH_M and FTT-NDH_M scores decline with

older age is supported by normative data for the FTT. In the widely used Heaton et al. (2004) normative data, for example, the same raw score on FTT-DH or FTT-NDH corresponds to a higher standardized score (indicating better performance) among older men and to a low standardized score (indicating worse performance) among younger men, suggesting that older examinees are expected to obtain lower raw scores on FTT, while younger examinees are expected to obtain higher raw scores. FTT-$DIFF_M$ BR_{Fail} did not appear to be associated with age in either sample and, therefore, this indicator is likely equally appropriate for use with examinees between the ages of 16 and 69 years.

BR_{Fail} on FTT-DH_M, FTT-NDH_M, and FTT-C_M appeared to be related to education level in both samples, although effect sizes were small. On each of these validity indicators, BR_{Fail} tended to decrease with higher levels of education. This finding is consistent with normative data for FTT (Heaton et al., 2004), in which the same raw score on FTT-DH or FTT-NDH corresponds to a higher standardized score in men with low levels of education and to a lower standardized score in men with high levels of education. However, in the present dissertation, the most educated group (≥17 years of education) demonstrated unexpectedly high BR_{Fail} relative to other education groups in both the overall and valid samples. As noted above, follow-up analyses suggested that age and diagnosis may represent confounds in the most educated group. While it is possible that the older age of examinees in this group may have accounted for their higher BR_{Fail} on FTT-DH_M, FTT-NDH_M, and FTT-C_M, normative data for the FTT (Heaton et al., 2004) suggest that older age alone is unlikely to fully account for this finding: although older age would be expected to result in higher BR_{Fail}, the high level of education among the most educated group would likely buffer the effects of age, at least to some degree. Considering this, the unexpectedly high BR_{Fail} observed among the most educated group is likely driven by a larger

proportion of patients with motor impairment secondary to neurological conditions. BR_{Fail} on $FTT\text{-}DIFF_M$ did not appear to be related to education level in either sample, suggesting that this indicator may be equally appropriate for use with individuals of varying education levels.

With respect to English language status, ESL patients tended to demonstrate higher BR_{Fail} on $FTT\text{-}DH_M$, $FTT\text{-}NDH_M$, and $FTT\text{-}C_M$ relative to NSEs in both the overall and valid samples. Given that BR_{Fail} continued to vary as a function of English language status when examinees who failed the WMT were excluded from the analyses, the higher BR_{Fail} among ESL patients, particularly in the valid sample, may reflect cultural differences, where motor speed may not be valued and/or emphasized in the ESL patients' cultures to the same degree as it is in the North American culture, for example (Rosselli & Ardila, 2003). BR_{Fail} on $FTT\text{-}DIFF_M$ were largely comparable among ESL patients and NSEs, suggesting that this validity indicator may not be affected by English language status.

Taken together, results for FTT-based validity indicators suggest that there may be an increased risk for false positive errors on $FTT\text{-}DH_M$, $FTT\text{-}NDH_M$, and $FTT\text{-}C_M$ among older, less educated, and ESL examinees. $FTT\text{-}DIFF_M$ BR_{Fail}, however, appear to be unrelated to the patient variables examined.

Study 3

Classification accuracy of individual PVTs and validity indicators. The results of Study 3 are largely consistent with past research in that many of the previously published cutoffs on PVTs of interest achieved acceptable (.84-1.00) specificity in the current sample. These include previously published cutoffs for the MSVT and NV-MSVT (Armistead-Jehle & Gervais, 2012; Green, 2007; Green, 2011), TOMM Trials 1 and Trial 2 (Bauer et al., 2007; Denning, 2012; Erdodi & Rai, 2017; Fazio, 2016; Greve et al., 2006a; Greve et al., 2006b; Greve, Etherton

et al., 2009; Horner et al., 2006; Jones, 2013; Schroeder et al., 2013; Tombaugh, 1996; Wisdom et al., 2012), RDS (Babikian et al., 2006; Duncan & Ausborn, 2005; Greiffenstein et al., 1994; Heinly et al., 2005; Larrabee, 2003; Mathias et al., 2002; Meyers & Volbrecht, 1998), WCST FMS (Greve, Heinly et al., 2009), CT-TE (Greve et al., 2007), TMT-based indicators (Busse & Whiteside, 2012; Iverson et al., 2002), and FTT-based indicators (Arnold et al., 2005; Axelrod et al., 2014; Shura et al., 2016; Smith et al., 2014).

Notably, there were a few cutoffs that demonstrated acceptable classification accuracy in previous studies but were associated with unacceptably low specificity in the current sample. For example, in a study by Larrabee (2003), WCST FMS ≥2 achieved acceptable specificity in moderate-to-severe closed head injury patients and good sensitivity in cases demonstrating worse-than-chance performance on a free-standing PVT (PDRT; Binder, 1993). However, the same cutoff produced a 27% false positive rate in Study 3 of the present dissertation. Similarly, TMT A+B ≥137", which was associated with .90 specificity and .21 sensitivity in a sample of post-deployment veterans (Shura et al., 2016), resulted in a 19% false positive error rate in Study 3.

These discrepancies in classification accuracy results for WCST FMS ≥2 and TMT A+B ≥137" are likely due to different sample characteristics. The valid group in Larrabee's (2003) study consisted of patients with moderate-to-severe closed head injury, while Study 3 involved a more heterogeneous sample containing patients with neurological, psychiatric, and medical/orthopedic diagnoses. Base rate data reported in Studies 1 and 2 indicate that, in addition to patients with TBI, those with neurological conditions, severe mental illness, orthopedic injuries, and *other* diagnosis demonstrate elevated BR_{Fail} on WCST FMS ≥2. Considering this,

the unacceptably high false-positive rate associated with WCST FMS ≥2 in the present dissertation is not surprising.

In the case of TMT A+B ≥137", although more than half of the post-deployment veterans included in the Shura et al. (2016) study met criteria for a psychiatric disorder at the time of the assessment and approximately 60% had a lifetime history of mTBI, individuals with moderate or severe TBI were excluded. In contrast, the current sample contained patients with a variety of conditions that would be expected to produce genuine and significant cognitive impairment (e.g., moderate to severe TBI, neurological conditions, severe mental illness). Thus, the mixed clinical sample in the present study was likely more cognitively impaired than the sample of post-deployment veterans used in the Shura et al. (2016) study and, therefore, required a more conservative cutoff to maintain an acceptable level of specificity.

In addition to WCST FMS ≥2 and TMT A+B ≥137", CT-TE ≥87 produced an unacceptably high false positive rate (17%) in the present dissertation. The latter finding is likely due to the fact that this cutoff originated from studies using analogue-simulation designs (Tenhula & Sweet, DiCarlo et al., 2000). As discussed earlier, analogue-simulation designs tend to overestimate classification accuracy (Babikian & Boone, 2007; Sollman & Berry, 2011).

While the majority of previously published cutoffs produced acceptable classification accuracy in Study 3, some previously published cutoffs were found to be overly conservative, producing high specificity at the cost of low sensitivity. This was particularly true for lower (i.e., more conservative) cutoffs on TOMM Trials 1 and 2, TMT-A ≥63, TMT-B ≥200, and cutoffs on several FTT-based indicators. Study 3 results indicate that more liberal cutoffs on these indicators can be used to assess performance validity in medical-legal samples, while maintaining good (≥.90) specificity.

Relative classification accuracy of PVTs of interest. Comparison of classification accuracy statistics across PVTs of interest revealed that free-standing PVTs achieved higher levels of sensitivity than did embedded validity indicators. Although these findings are consistent with previous literature showing that free-standing PVTs achieve greater sensitivity to invalid performance than embedded validity indicators when used alone (Iverson & Binder, 2000; Miele et al., 2012; Sweet & Nelson, 2007), it is possible that instrumentation artifacts influenced these findings to some degree (Erdodi, 2017). Specifically, all of the free-standing PVTs of interest in the present dissertation matched the criterion PVT in terms of testing paradigm and cognitive domain (i.e., the WMT and all free-standing PVTs were forced-choice recognition memory tests). On the other hand, none of the embedded validity indicators were derived from memory tests and only one of them (i.e., CT-TE) employed a forced-choice recognition paradigm. As a result of these differences, sensitivity and specificity may have been overestimated for the free-standing PVTs and underestimated for the embedded validity indicators.

General Discussion

Free-standing PVTs versus embedded validity indicators. Taken together, the results of this dissertation suggest that free-standing PVTs are not only more sensitive to invalid performance than embedded validity indicators, but they also tend to be more resistant to the effects of some patient variables (e.g., diagnosis, education). The stronger association between diagnosis and BR_{Fail} on embedded validity indicators (as compared to free-standing PVTs) is likely due to the fact that these indicators are derived from standard tests of neuropsychological ability. Given that neuropsychological tests are designed to detect true cognitive (or in the case of FTT, motor) impairment, the "after-market" validity cutoffs are less effective than free-standing PVTs at separating genuine impairment from non-credible responding.

The fact that BR_{Fail} on embedded validity indicators tended to be more strongly related to age and education level than BR_{Fail} on free-standing PVTs is likely due in large part to the fact that nearly all of the embedded validity indicators examined (with the exception of RDS) represented raw rather than standardized scores, either alone in combination. It is well established that age and education affect raw scores on neuropsychological tests (Strauss et al., 2006). As a result, these variables are routinely accounted for in the normative data used to interpret patients' neuropsychological test scores (e.g., Heaton et al., 2004; Pearson, 2009; Strauss et al., 2006). Considering this, the practice of co-opting raw scores as validity indicators is particularly problematic as it carries the risk for increased false positives in older and less educated examinees.

To reduce the impact of patient variables such as age and education on PVT scores, more recent studies have examined the utility of age- and demographically-corrected test scores as validity indicators. Examples of such indicators include the age-corrected scaled score on the Digit Span subtest of the WAIS (DS-ACSS; Jasinski et al., 2011), age-corrected scores on TMT-A and TMT-B (Ashendorf, Clark, & Sugarman, 2017), and demographically-corrected *T* scores on the Controlled Oral Word Association Test (Sugarman & Axelrod, 2015). The existing literature suggests that these indicators not only account for confounding demographic factors but they also demonstrate good classification accuracy. Thus, standardized scores may represent more effective embedded validity indicators than raw scores and should continue to be investigated further in future research.

Subtypes of embedded validity indicators. Compared to the other embedded validity indicators examined in this dissertation, WCST FMS, TMT-B/A and FTT-DIFF differed in that they were largely unrelated to the patient variables examined in Studies 1 and 2, and they also

demonstrated among the lowest sensitivity to invalid performance (as defined by the WMT) in Study 3. These findings may be due to the fact that WCST FMS, TMT-B/A and FTT-DIFF assess performance validity in different ways than do many of the other PVTs of interest (Erdodi, 2017).

Most of the PVTs of interest examined in this dissertation evaluate performance validity based on the examinee's absolute level of performance on the task: better performance on the task is classified as valid and worse performance is classified as invalid. WCST FMS, however, captures an atypical aspect of performance rather than poor performance on the task more globally (Larrabee, 2003). An FMS error is recorded when an examinee matches a card incorrectly after matching five or more consecutive cards correctly but before successfully completing the category. Data presented in the WCST manual (Heaton et al., 1993) indicate that multiple FMS errors are observed infrequently across examinees, irrespective of age, diagnosis, or level of impairment. Given the rarity of such errors, when used as an embedded validity indicator, WCST FMS provides qualitatively different information from most other performance validity assessment tools and, therefore, may be conceptualized as a unique subtype of embedded validity indicator.

TMT-B/A and FTT-DIFF also differ from most other embedded validity indicators in their approach to performance validity assessment. Specifically, rather than considering the absolute performance on a single task, these indicators evaluate examinees' relative performances on two tasks of varying difficulty levels (i.e., TMT-A and TMT-B, FTT-DH and FTT-NDH) and flag as invalid those examinees who perform comparably or better on the more difficult task as compared to the easier task (Arnold et al., 2005; Iverson et al., 2002). The inner logic behind the use of TMT-B/A and FTT-DIFF as validity indicators may explain why these

indicators were found to be unrelated to patient variables in Studies 1 and 2. Specifically, although performance on each task is affected by diagnosis and demographic variables, the act of comparing an examinee's performance on one task to their own performance on another task offsets the effects of these variables. Thus, the resulting difference- or ratio-score is not associated with diagnosis and demographic variables in the same way as the scores on each individual task. Taken together, the findings from this dissertation suggest that, like WCST FMS, TMT-B/A and FTT-DIFF may also constitute a unique subtype of embedded validity indicators.

When compared to free-standing PVTs and other embedded validity indicators, WCST FMS, TMT-B/A and FTT-DIFF were among the least sensitive to invalid performance as defined by the WMT. While these findings can be interpreted as suggesting that these indicators are less effective as PVTs, their low sensitivities may simply reflect a mismatch in assessment modality/detection mechanism between the predictor PVTs and the criterion PVT (Erdodi, 2017): indeed, an examinee who fails the WMT may not necessarily violate the normative difficulty gradients on tasks like the TMT and FTT or make multiple atypical FMS errors on the WCST. Consistent with this hypothesis, recent studies have found that the match between criterion and predictor variables (i.e., in terms of sensory modality, cognitive domain, and testing paradigm) influences the outcome of signal detection analyses and may represent a significant confound in performance validity research more broadly (Erdodi, 2017; Erdodi, Abeare, et al., 2017; Erdodi, Tyson, et al., 2017). Given the major research and clinical implications of these findings, further investigation of Erdodi's (2017) *domain specificity hypothesis* is required to better understand how various characteristics of predictor and criteria PVTs, and the match or mismatch between them, impact classification accuracy.

Implications for clinical and medical-legal practice. Based on research indicating that a large proportion of invalid cases go undetected when using a single PVT, experts in the field have repeatedly recommended using multiple, non-redundant PVTs, interspersed throughout the test battery, to monitor test-taking effort continuously during neuropsychological assessments (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2012c; Orey et al., 2000; Sweet & Nelson, 2007; Vickery et al., 2004). Despite the fact that such recommendations were first made nearly two decades ago, however, there is currently little consensus on which PVTs and/or embedded validity indicators should be used in different assessment settings, the order in which they should be administered, and whether liberal or conservative cutoffs should be applied. Additionally, although there is reasonable agreement among experts that evidence of invalid performance on two independent measures renders the overall neuropsychological profile invalid (Bilder, Sugar, & Hellemann, 2014; Boone, 2013; Davis & Millis, 2014; Larrabee, 2003), the question about how many PVTs to administer continues to be hotly debated (Davis & Millis, 2014; Larrabee, 2014a; Silk-Eglit, Stenclik, Miele, Lynch, & McCaffrey, 2015). Larrabee (2014a), for example, demonstrated that administering multiple PVTs, each with .90 specificity (i.e., a 10% per-test false-positive rate), results in a small increase in the *overall* false-positive rate as the number of PVTs increases. In such cases, he recommended reducing false positives by adjusting cutoffs in light of patient characteristics and/or increasing the number of PVT "failures" required before a neuropsychological profile is deemed invalid (Larrabee, 2014b).

Results from this dissertation can help to address some of the challenges identified with the use of PVTs in practice, particularly with respect to the selection of PVTs and cutoff scores. By reporting classification accuracy for multiple cutoffs on a wide range of PVTs, the results of Study 3 allow clinicians to compare sensitivity and specificity both between and within PVTs

(i.e., how the classification accuracy of one PVT compares to another, how different cutoffs on a given PVT compare to eachother) and select PVTs and cutoff scores according to the needs of their practices. Once the PVTs and cutoff scores are selected, the base rate data reported as part of Study 2 can be used as an empirical basis to adjust cutoffs for particular patient groups. If, for example, credibly-performing older examinees demonstrated an elevated BR_{Fail} at a particular cutoff on a given PVT, clinicians using that PVT with older patients may wish to use a more conservative cutoff to protect against false-positive errors.

Limitations. The results of this dissertation should be interpreted in light of its limitations. First, the data used for this dissertation were collected in a primarily medical-legal setting and in a relatively restricted geographical region (i.e., a midwestern Canadian city). Therefore, these findings may not generalize to other assessment settings (e.g., where neuropsychological assessments are conducted purely for clinical purposes) or geographical regions. Second, participants were assigned to diagnostic groups based on their *primary* diagnoses and no information was available about comorbid conditions. Given that patients assessed in clinical/medical-legal practice often have multiple diagnoses, the diagnostic groups included in the present dissertation were not mutually exclusive. Third, there was large variability in the sizes of various diagnostic and demographic groups, and some patient variables were confounded with others. For example, the oldest age group was slightly less educated than the younger groups; the least educated group and the two most educated groups were older in age than the remaining education groups; and the two most educated groups contained a larger proportion of patients with neurological conditions than the other education groups. Considering these findings, in many cases, the reported BR_{Fail} likely reflect the interaction of multiple patient variables as opposed to any single variable. Fourth, due to the retrospective nature of this

dissertation and a flexible-battery approach to test selection at the practices from which the data were obtained, there were many tests that were not administered to all participants and some tests (e.g., FTT) were not administered widely enough to draw meaningful conclusions about trends in BR_{Fail} across diagnostic/demographic groups in Studies 1 and 2. Fifth, the WMT (Green, 2003) was used as the criterion PVT, in part because it was the most widely administered PVT. It is possible that the results of the present dissertation would have been different if another criterion (e.g., MND criteria, a criterion based on a different free-standing PVT or multiple PVTs) had been used. Finally, while very interesting findings related to English language status emerged in Studies 1 and 2, unfortunately, the ESL group in the present dissertation was poorly defined and quite heterogeneous. Participants were included in the ESL group if they identified as non-native speakers of English during the clinical interview, regardless of their country of origin (i.e., Canada or elsewhere in the world). Considering this, the ESL group likely contained examinees who varied widely in their true levels of English proficiency. Unfortunately, however, additional information that would have been important for quantifying the participants' levels of English proficiency (e.g., age at which conversational English was learned, years of education completed in Canada or in the English language, number of years spent living in Canada) was not available.

Future Directions. The results of the present dissertation suggest a number of important directions for future research. First and foremost, further research is needed on individual PVTs (both free-standing and embedded) to better understand how they operate in different clinical and demographic groups, including and especially among ESL populations, and in different settings (e.g., medical-legal versus clinical). Ideally, future research would lead to the development of population-specific cutoffs (i.e., cutoffs unique to different clinical and/or demographic groups)

on each PVT so that cutoffs established with one clinical group (e.g., mTBI) do not have to be extended to new groups in the absence of any evidence to support this practice.

The results of the present dissertation also provide some evidence for differentiating embedded validity indicators into subtypes based on assessment modality/detection mechanism. However, further research is required to better understand the classification accuracy and clinical utility of validity indicators that differ in their detection mechanisms (e.g., whether certain detection mechanisms are better suited for particular patient groups and/or specific assessment contexts than other detection mechanisms). Importantly, given recent studies suggesting that the classification accuracy of predictor PVTs may vary depending on the extent to which they match criteria PVT (i.e., in terms of sensory modality, cognitive domain, or testing paradigm; Erdodi, 2017; Erdodi, Abeare, et al., 2017; Erdodi, Tyson, et al., 2017), PVTs of interest should be evaluated against a variety of criteria (e.g., MND criteria, different free-standing PVTs, and/or aggregrate measures based on multiple PVTs) in order to obtain more accurate estimates of their true classification accuracies.

Finally, where possible, multiple PVTs should be evaluated in the same samples so as to facilitate direct comparisons between individual PVTs. Such data would help to determine whether and to what extent different PVTs provide non-redundant information with respect to performance validity and, ultimately, inform the selection of PVTs for multivariate models of performance validity assessment.

CHAPTER 5

Conclusions

The results of this dissertation converge in three major conclusions. First, the assessment of performance validity using psychometric instruments is necessary in order to establish the veracity of neuropsychological test data, particularly in medical-legal settings where patients often have substantial external incentives to appear impaired. Second, given the relationships between patient characteristics and BR_{Fail}, particularly on embedded validity indicators, future research investigating population-specific cutoffs would greatly enhance the clinical utility of PVTs. In the meantime, individual PVT scores should be interpreted in the context of patient variables, in an empirically-informed manner, rather than by rigid adherence to published cutoffs. Finally, although free-standing PVTs tend to be more sensitive to invalid performance than embedded validity indicators when used individually, no PVT has perfect classification accuracy. Therefore, rather than using a single PVT to assess performance validity, practitioners should use multiple, non-redundant PVTs that are interspersed throughout the test battery.

REFERENCES

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB'97 manual for the computerized assessment of response bias*. Durham: CogniSyst.

Ardila, A., & Moreno, S. (2001). Neuropsychological evaluation in Aruaco Indians: An exploratory study. *Journal of the International Neuropsychological Society, 7,* 510-515.

Armistead-Jehle, P., & Denney, R. L. (2014). The detection of feigned impairment using the WMT, MSVT, and NV-MSVT. *Applied Neuropsychology: Adult, 22,* 147-155.

Armistead-Jehle, P., & Gervais, R. O. (2011). Sensitivity of the Test of Memory Malingering and the Nonverbal Medical Symptom Validity Test: A Replication Study. *Applied Neuropsychology, 18,* 284-290.

Arnett, P. A., & Franzen, M. D. (1997). Performance of substance abusers with memory deficits on measures of malingering. *Archives of Clinical Neuropsychology, 12,* 513-518.

Arnold, G., & Boone, K. B. (2007). Use of motor and sensory tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 178-209). New York, NY: The Guilford Press.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S. & McPherson (2005). Sensitivity and specificity of finger tapping test scores for the detection of suspect effort. *The Clinical Neuropsychologist, 19,* 105-120.

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.

Ashendorf, L., Constantinou, M., & McCaffrey, R. J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology, 19*(1), 125-130.

Axelrod, B. N., Meyers, J. E., & Davis, J. J. (2014). Finger tapping test performance as a measure of performance validity. *The Clinical Neuropsychologist, 28*(5), 876-888.

Babikian, T., & Boone, K. B. (2007). Intelligence tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 103-127). New York, NY: The Guilford Press.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Psychologist, 20,* 145-159.

Backhaus, S. L., Fichtenberg, N. L., & Hanks, R. A. (2004). Detection of suboptimal performance using a floor-effect strategy in patients with traumatic brain injury. *The Clinical Neuropsychologist, 18*(4), 591-603.

Baldessarini, R. J., Finklestein, S. & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40,* 569-573.

Batt, K. Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society, 14,* 1074-1080.

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering Trial 1 and Word Memory Test Immediate Recognition as screening tools for insufficient effort. *Assessment, 14*(3), 215-222.

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: a meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215-227.

Bennett, R. M. (1996). Fibromyalgia and the disability dilemma. *Arthritis and Rheumatism, 39*(10), 1627-1634.

Berry, J. W., Poortinga, Y. P., & Segall, M. G. H. (1992). *Cross-cultural psychology*. Cambridge: Cambridge University Press.

Bianchini, K. J, Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*(1), 19-45.

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

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*(13-14), 1623-1638.

Bilder, R. M., Sugar, C. A., & Hellemann, G. S. (2014). Cumulative false positive rates given multiple performance validity tests: Commentary on Davis and Millis (2014) and Larrabee (2014). *The Clinical Neuropsychologist, 28*(8), 1212-1223.

Binder, L. M. (1993a). *Portland digit recognition test manual* (2nd ed.). Portland, OR, USA: Private Publication.

Binder, L. M. (1993b). Assessment of malingering after mild head trauma with the Portland digit recognition test. *Journal of Clinical and Experimental Neuropsychology, 15*, 170-182.

Binder, L. M. (2012). Forensic assessment of medically unexplained symptoms. In G. J. Larrabee (Ed.), *Forensic Neuropsychology: A Scientific Approach* (pp.336-364). New York, NY: Oxford University Press.

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effect of financial incentives on recovery after closed head injury. *American Journal of Psychiatry, 153,* 7-10.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*(2), 175-181.

Boone (2007). A reconsideration of the Slick et al. (1999) criteria for Malingered Neurocognitive Dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29-49). New York, NY: The Guilford Press.

Boone, K. B. (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY: The Guilford Press.

Boone, K. B. (2011). Clarification or confusion? A review of Rogers, Bender, and Johnson's A critical analysis of the MND criteria for feigned cognitive impairment: Implications for forensic practice and research. *Psychol. Inj. and Law 4,* 157-162.

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

Boone, K. B., Lu, P., & Herzberg, D. (2002a). *The b Test Manual*. Los Angeles: Western Psychological Services.

Boone, K. B., Lu, P., & Herzberg, D. (2002b). *Rey Dot Counting Test: A handbook*. Los Angeles: Western Psychological Services.

Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Ponton, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology, 22,* 355-365.

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

Busse, M., & Whiteside, D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Conner's Continuous Performance Test-II, Brief Test of Attention, and Trail Making Test. *The Clinical Neuropsychologist, 26*(4), 675-687.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22,* 960-971.

Carone, D. A. (2014). Young children with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: Implications for mTBI. *The Clinical Neuropsychologist, 28*(1): 146-162.

Cato, M. A., Brewster, J., Ryan, T., & Guiliano, A. J. (2003). Coaching and the ability to simulate mild traumatic brain injury symptoms. [Erratum.] *The Clinical Neuropsychologist, 17,* 285-286.

Choca, J., & Morris, J. (1992). Administering the Category Test by computer: Equivalence of results. *The Clinical Neuropsychologist, 6,* 9-15.

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

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 22,* 705-722.

DeFlippis, N. A., & McCampbell, E. (1979, 1991, 1997). *Manual for the Booklet Category Test*. Odessa, FL: Psychological Assessment Resources.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test (2nd ed., Adult Version): Manual*. Bloomington, MN: NCS Pearson.

Demakis, G. J. (2004). Application of clinically-derived malingering cutoffs on the California Verbal Learning Test and the Wechsler Adult Intelligence Test-Revised to an analog malingering study. *Applied Neuropsychology, 11*(4), 222-228.

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,* 417-432.

DiCarlo, M. A., Gfeller, J. D., & Oliveri, M. V. (2000). Effects of coaching on detecting feigned cognitive impairment with the Category Test. *Archives of Clinical Neuropsychology, 15*(5), 399-413.

Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the Test of Memory Malingering (TOMM). *Assessment, 12*(2), 123-129.

Duncan, S. A., & Ausborn, D. L. (2002). The use of reliable digits to detect malingering in a criminal forensic pretrial population. *Assessment, 9*(1), 56-61.

Dwyer, C. A. (1996). Cutoffs and testing: Statistics, judgment, truth, and error. *Psychological Assessment, 8,* 360-362.

Ekman, P., O'Sullivan, M., & Frank, M. (1999). A few can catch a liar. *Psychological Science, 10,* 236-266.

Erdodi, L. A. (2017). Aggregating validity indicators: The salience of domain specificity and the indeterminate range in multivariate models of performance validity assessment. *Applied Neuropsychology: Adult*. Advance online publication. doi: 10.1080/23279095.2017.1384925

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

Erdodi, L. A., 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. doi: 10.1037/pas0000319

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.

Erdodi, L. A., Nussbaum, S., Sagar, S., Abeare, C. A., & Schwartz, E. S. (2017). Limited English proficiency increases BRFail on performance validity tests with high verbal mediation. *Psychological Injury and Law, 10*(1), 96-103.

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.

Etherton, J. L., Bianchini, K. J., Greve, K. W. & Heinly, M. T. (2005). Sensitivity and specificity of reliable digit span in malingered pain-related disability. *Assessment, 12*(2), 130-136.

Faust, D. (1995). The detection of deception. *Neurologic Clinics, 13,* 255-265.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19,* 508-515.

Frederick, R. I., Sarfaty, S. D., Johnston, D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology, 8,* 118-125.

Gavett, B. E., O'Bryant, S. E., Fisher, J. M., & McCaffrey, R. J. (2005). Hit rates of adequate performance based on the Test of Memory Malingering (TOMM) Trial 1. *Applied Neuropsychology, 12*(1), 1-4.

Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM BRFail in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19,* 475-487.

Gervais, R. O., Russell, A. S., Green, P., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology, 28,* 1892-1899.

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology, 44,* 1013-1023.

Glass, J. M., & Park, D. C. (2001). Cognitive dysfunction in fibromyalgia. *Current Rheumatology Reports, 3*(2), 123-127.

Glassmire, D. M., Ross, P. T., Kinney, D. I., & Nitch, S. R. (2016). Derivation and cross-validation of cutoff scores for patients with schizophrenia spectrum disorders on WAIS-IV Digit Span-based performance validity measures. *Assessment*, *23*(3), 292-306.

Goebel, R. A. (1983). Detection of faking on the Halstead-Reitan neuropsychological test battery. *Journal of Clinical Psychology, 39,* 731-742.

Goldberg, H. E., Back-Madruga, C., & Boone, K. B. (2007). The impact of psychiatric disorders on cognitive symptom validity test scores. In K. B. Boone (Ed.), *Assessment of feigned*

cognitive impairment: A neuropsychological perspective (pp. 281-309). New York, NY: The Guilford Press.

Grace, G. M., Nielson, W. R., Hopkins, M., & Berg, M. A. (1999). Concentration and memory deficits in patients with fibromyalgia syndrome. *Journal of Clinical and Experimental Neuropsychology, 21*(4), 477-487.

Green, P. (2003). Green's Word Memory Test User's Manual. Edmonton, Canada: Green's Publishing.

Green, P. (2004). Greens' medical Symptom Validity Test User's Manual. Edmonton, Canada: Green's Publishing.

Green, P. (2008). Green's Non-verbal Medical Symptom Validity Test User's Manual. Edmonton, Canada: Green's Publishing.

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

Green, P., & Flaro, L. (2003). Word memory test performance in children. *Child Neuropsychology, 9*(3), 189-207.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23*(9), 741-750.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2*, 97-124.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology: Adult, 18*(2), 86-94.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*(12), 1045-1060.

Greiffenstein, M. F. (2008). Basic of forensic neuropsychology. In J. E. Morgan & J. H. Ricker (Eds.). *Textbook of Clinical Neuropsychology* (pp. 905-942). New York, NY: Taylor & Francis.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224.

Greiffenstein, M. F., Greve, K. W., Bianchini, K. J., & Baker, W. J. (2008). Test of memory malingering and word memory test: a new comparison of concordance rates. *Archives of Clinical Neuropsychology, 23*(7-8), 801-807.

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology, 19,* 533-541.

Greve, K. W., & Bianchini, K. J. (2006). Should the retention trial of the Test of Memory Malingering be optional?. *Archives of Clinical Neuropsychology, 21,* 117-119.

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology, 28,* 1176-1190.

Greve, K. W., Bianchini, K. J., & Ord, J. S. (2012). The psychological assessment of persons with chronic pain. In G. J. Larrabee (Ed.), *Forensic Neuropsychology: A Scientific Approach* (pp. 302-363). New York, NY: Oxford University Press.

Greve, K. W., Bianchini, K. J., & Roberson, T. (2007). The Booklet Category Test and malingering in traumatic brain injury: Classification accuracy in known groups. *The Clinical Neuropsychologist, 21,* 318-337.

Greve, K. W., Bianchini, K. J., Etherton, J. L., Ord, J. S., & Curtis, K. L. (2009). Detecting malingered pain-related disability: Classification accuracy of the Portland Digit Recognition Test. *The Clinical Neuropsychologist, 23*(5), 850-869.

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*, 1250-1271.

Greve, K. W., Heinly, M. T., Bianchini, K. J., & Love, J. M. (2009). Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *The Clinical Neuropsychologist, 23,* 343-362.

Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation, 90*(7), 1117-1126.

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22,* 896-918.

Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology, 6,* 709-719.

Haggerty, K. A., Frazier, T. W., Busch, R. M., & Naugle, R. I. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist, 21,* 917-928.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology, 17,* 709-714.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test manual - revised and expanded*. Odessa, FL: Psychological Assessment Resources, Inc.

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 – Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Smith, H. H. Jr., 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.

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

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS Digit Span-based indicators of malingered neurocognitive dysfunction. *Assessment, 12*(4), 429-444.

Henry, M., Merten, H., Wolf, S. A., & Harth, S. (2009). Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and neurology patients. *Journal of Clinical and Experimental Neuropsychology, 32*(1), 19-27.

Hilsabeck, R. C., Gordon, S. N., Hietpas-Wilson, T., & Zartman, A. I. (2011). Use of trial 1 of the test of memory malingering (TOMM) as a screening measure of effort: Suggested discontinuation rules. *The Clinical Neuropsychologist, 25,* 1228-1238.

Hilsabeck., R. C., Thompson, M. D., Irby, J. W., Adams, R. L., Scott, J. G., & Gouvier, W. D. (2003). Partial cross-validation of the Wechsler Memory Scale-Revised (WMS-R) general memory-attention/concentration malingering index in a nonlitigating sample. *Archives of Clinical Neuropsychology,* 18(1), 71-79.

Howe, L. L. S., & Loring, D. W. (2009). Classification accuracy and predictive ability of the Medical Symptom Validity Test;s dementia profile and general memory impairment profile. *The Clinical Neuropsychologist, 23,* 329-342.

Howe, L. L. S., Anderson, A. M., Kaufman, D. A. S., Sachs, B. C., & Loring, D. W. (2007). Characterization of the Medical Symptom Validity Test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology, 22*, 753-761.

Hubbard, K. L. (2008). Feigning cognitive deficits among psychiatric inpatients: Validation of three measures of cognitive malingering. *Dissertation Abstracts International, Section B: The Sciences and Engineering, 70,* 3782.

Hurtubise, J., Scavone, A., Sagar, S. & Erdodi, L. A. (2017). Psychometric markers of genuine and feigned neurodevelopmental disorder in the context of applying for academic accommodations. *Psychological Injury and Law, 10*(2), 121-137.

Irvine, S. H., & Berry, J. W. (Eds.). (1988). *Human abilities in cultural context*. New York: Cambridge University Press.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma and Rehabilitation, 15*(2), 829-858.

Iverson, G. L., King, R. J., Scott, J. G., & Adams, R. L. (2001). Cognitive complaints in litigating patients with head injuries or chronic pain. *Journal of Forensic Neuropsychology, 2*(1), 19-30.

Iverson, G. L., & Lange, R. T. (2011). Moderate and severe traumatic brain injury. In M. R. Schoenberg & J. G. Scott (Eds.), *The Little Black Book of Neuropsychology: A Syndrome-Based Approach* (pp. 663-696). New York, NY: Springer.

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. D. (2002). Detecting exaggeration and malingering with the Trail Making Test. *The Clinical Neuropsychologist, 16*(3), 398-406.

Iverson, G. L., & McCracken, L. M. (1997). 'Postconcussive' symptoms in persons with chronic pain. *Brain Injury, 11*(11), 783-790.

Ivnik, R. J., Smith, G. E., Cerhan, J. H., Boeve, B. F. Tangalos, E. G., & Peterson, R. C. (2001). Understanding the diagnostic capabilities of cognitive tests. *The Clinical Neuropsychologist, 15,* 114-124.

Jasinski, L. J., Berry, D. T. R., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 300-314.

Johnson, Z., & van den Brock, M. D. (2001). Letter to the editor. *Brain Injury, 15,* 187-188.

Kiewel, N. A., Wisdom, N. M., Bradshaw, M. R., Pastorek, N. J., & Strutt, A. M. (2012). A retrospective review of Digit Span-related effort indicators in probably Alzhiemer's disease patients. *The Clinical Neuropsychologist, 26*(6), 965-974.

Lamberty, G. J., Putnam, S. H., Chatel, D. M., Bieliauskas, L. A., & Adams, K. M. (1994). Derived Trail Making Test indices. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 7,* 230-234.

Larrabee, G. J. (1992). Interpreting strategies for evaluation of neuropsychological data in legal settings. *Forensic Reports, 5,* 257-264.

Larrabee, G. J. (2000). Forensic neuropsychological assessment. In R. D. Vanderploeg, *Clinician's guide to neuropsychological assessment* (pp. 301-335). Mahwah, NJ: Lawrence Erlbaum.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*(3), 410-425.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22,* 666-79.

Larrabee, G. J. (2012a). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 625-631.

Larrabee, G. J. (2012b). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed.) (pp. 116-159). New York, NY: Oxford University Press, Inc.

Larrabee, G. J. (2012c). A scientific approach to forensic neuropsychology. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed.) (pp. 3-22). New York, NY: Oxford University Press, Inc.

Larrabee, G. J. (2014a). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29,* 364-373.

Larrabee, G. J. (2014b). Minimizing false positive error with multiple performance validity tests: Response to Bilder, Sugar, and Hellemann (2014). *The Clinical Neuropsychologist, 28*(8), 1230-1242.

Larrabee, G. J., Greiffenstein, M. F., Greve, L., W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334-371). New York: Oxford University Press.

Leon-Carrion, J. (1989). Trail making test scores for normal children: Normative data from Spain. *Perceptual and Motor Skills, 68,* 627-630.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press, Inc.

Lezak, M., Howieson, M., & Loring, D. (2004). *Neuropsychological assessment* (4th ed.). New York, NY: Oxford University Press, Inc.

Lindstrom, W. A., Lindstrom, J. H., Coleman, C., Nelson, J., & Gregg, N. (2009). The diagnostic accuracy of symptom validity tests when used with postsecondary students with learning disabilities: a preliminary investigation. *Archives of Clinical Neuropsychology, 24,* 659-669.

Loring, D. W., Goldstein, F. C., Chen, C., Drane, D. L., Lah, J. J., Zhao, L., Larrabee, G. J., & Alzheimer's Disease Neuroimaging Initiative. False-positive error rates for reliable digit

span and auditory verbal learning test performance validity measures in amnestic mild cognitive impairment and early Alzheimer's disease. *Archives of Clinical Neuropsychology, 31*(4), 313-331.

Lu, P. H., Rogers, S. A., & Boone, K. B. (2007). Use of standard memory tests to detect suspect effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 128-151). New York, NY: The Guilford Press.

Macciocchi, S. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: implications for test interpretation. *Archives of Clinical Neuropsychology, 21*(5), 395-404.

Manly, J. J., Byrd, D. A., Touradji, P., & Stern, Y. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders. *Applied Neuropsychology,* 1137-1146.

Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction using the reliable digit span in traumatic brain injury. *Assessment, 9*(3), 301-308.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194-216.

Mehta, K., Simonsick, E. M., Rooks, R., Newman, A. B., Pope, S. K., Rubin, S. M., et al. (2004). Black and white differences in cognitive function scores: What explains the difference? *American Geriatrics Society, 52,* 2120-2127.

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,* 719-726.

Meyers, J. E., & Diep, A. (2000). Assessment of malingering in chronic pain patients using neuropsychological tests. *Applied Neuropsychology, 7*(3), 133-139.

Meyers, J. E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment, 5*(3), 303-307.

Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18*(3), 261-276.

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology, 27*, 10-22.

Milanovich, J. R., Axelrod, B. N., & Millis, S. R. (1996). Validation of the Simulation Index-Revised with a mixed clinical population. *Archives of Clinical Neuropsychology, 11*(1), 53-59.

Millis, S. R. (2008). Assessment of incomplete effort and malingering in the neuropsychological examination. In J. E. Morgan & J. H. Ricker (Eds.). *Textbook of Clinical Neuropsychology* (pp. 891-904). New York, NY: Taylor & Francis.

Millis, S. R. (2009). What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 21-37). New York: Psychology Press.

Millis, S. R., & Kler, S. (1995). Limitations of the Rey Fifteen-Item Test in the detection of malingering. *The Clinical Neuropsychologist, 9*(3), 241-244.

Millis, S. R., & Putnam, S. H. (1996). Detection of malingering in postconcussive syndrome. In M. Rizzo & D. Tranel (Eds.), *Head injury and postconcussive syndrome* (pp. 481-498). New York: Churchill Livingstone.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. New York, NY: Oxford University Press.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*98), 1094-1102.

Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 3*(1/2), 227-240.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *Clinical Neuropsychologist, 17,* 263-272.

Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13,* 440-449.

Nitch, S. R., & Glassmire, D. M. (2007). Non-forced-choice measures to detect noncredible cognitive performance. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 78-102). New York, NY: The Guilford Press.

Norman, M., Evans, J., Miller, W., & Heaton, R. (2000). Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 22,* 80-94.

O'Bryant, S. E., Engel, L. R., Kleiner, J. S., Vasterling, J. J., & Black, F. W. (2007). Test of memory malingering (TOMM) trial 1 as a screening measure for insufficient effort. *Clinical Neuropsychologist, 21*(3), 511-521.

O'Bryant, S. E., Gavett, B. E., McCaffrey, R. J., O'Jile, J. R., Huerkamp, J. K., Smitherman, T. A., & Humphreys, J. D. (2008). Clinical utility of Trial 1 of the Test of Memory Malingering (TOMM). *Applied Neuropsychology, 15,* 113-116.

Paniak, C., Reyolds, S., Toller-Lobe, G., Melnyk, A., Nagy, J., & Schmidt, D. (2002b). A longitudinal study of the relationship between financial compensation and symptoms after treated mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 24,* 187-194.

Pankratz, L., & Erickson, R. C. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4,* 379-389.

Patton, D., Duff, K., Schoenberg, M., Mold, J., Scott, J., & Adams, R. (2003). Performance of cognitively normal African Americans on the RBANS in community dwelling older adults. *The Clinical Neuropsychologist, 17,* 515-530.

Pearson (2009). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV – Technical Manual*. San Antonio, TX: Author.

Perez-Arce, P., & Puente, A. (1996). Neuropsychological assessment of ethnic minorities. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological tests* (pp. 283-300). Delray Beach: GR Press.

Reitan, R. M. (1969). *Manual for administration of neuropsychological test batteries for adults and children*. Indianapolis, Ind.

Reitan, R. M. (2001). Differentiating between peripheral and central lateralized neuropsychological deficits. *Journal of Forensic Neuropsychology, 2,* 21-27.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and interpretation*. Tucson, AZ: Neuropsychology Press.

Reslan, S., & Axelrod, B. N. (2017). Evaluating the Medical Symptom Validity Test (MSVT) in a sample of veterans between the ages of 18 to 64. *Applied Neuropsychology: Adult, 24*(2), 132-139.

Reynolds, C. R. (Ed.). (1998). *Detection of malingering durng head injury litigation*. New York: Plenum Press.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 398-426). New York: Guilford Press.

Rogers, R. (2008). Researching response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.). New York: Guilford Press.

Rogers, R., Harrel, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychological Review, 13,* 255-274.

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition, 52,* 326-333.

Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated

psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist, 22,* 547-564.

Salazar, X. F., Lu, P. H., Wen, J., & Boone, K. B. (2007). The use of effort tests in ethnic minorities and in non-English speaking and English as a second language populations. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405-427). New York, NY: The Guilford Press.

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). *Journal of Neurology, Neurosurgery, & Psychiatry, 64,* 339-343.

Schnurr, R. F., & MacDonald, M. R. (1995). Memory complaints in chronic pain. *The Clinical Journal of Pain, 11*(2), 103-111.

Schroeder, R. W., Buddin, W. H. Jr., Hargrave, D. D., VonDran, E. J., Campbell, E. B., Brockman, C. J., Heinrichs, R. J., & Baade, L. E. (2013). Efficacy of Test of Memory Malingering Trial 1, Trial 2, the Retention Trial, and the Albany Consistency Index in a criterion group forensic neuropsychological sample. *Archives of Clinical Neuropsychology, 28*, 21-29.

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment, 19*(1), 21-30.

Seusse, M., Wong, V. W. C., Stamper, L. L., Carpenter, K. N., & Scott, R. B. (2015). Evaluating the clinical utility of the Medical Symptom Validity Test (MSVT): A clinical series. *The Clinical Neuropsychologist, 29*(2), 214-231.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practice with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213-223.

Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gantz, R. E., & Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, *23*, 94-103.

Silk-Eglit, G. M., Stenclik, J. H., Miele, A. S., Lynch, J. K., & McCaffrey, R. J. (2015). Rates of false-positive classification resulting from the analyses of additional embedded performance validity mesures. *Applied Neuropsychology: Adult, 22*(5), 335-347.

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology, 24,* 721-728.

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,* 545-561.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *VSVT: Victoria Symptom Validity Test (Version 1.0)*. Odessa, Florida: Psychological Assessment Resources.

Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. *Journal of Clinical and Experimental Neuropsychology, 16,* 472-481.

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.

Sollman, M. J., & Berry, D. T. R. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26,* 774-789.

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.

Strauss, E., Sherman, E. S., & Spreen, O. (2006). *A compendium neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press, Inc.

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

Suhr, J. A. (2003). Neuropsychological impairment in fibromyalgia: Relation to depression, fatigue, and pain. *Journal of Psychosomatic Research, 55*(4), 321-329.

Suhr, J. A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology, 21*(5), 701-708.

Suhr, J. A., & Gunstad, J. (2002). "Diagnosis threat": The effect of negative expectations on cognitive performance in head injury. *Journal of Clinical and Experimental Neuropsychology, 24*(4), 448-457.

Sweet, J. J. (Ed.) (1999). *Forensic neuropsychology: Fundamentals and practice*. Exton, PA: Swets & Zeitlinger.

Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 255-285). New York: Psychology Press.

Sweet, J. J., Goldman, D. J., & Guidotti Breting, L. M. (2013). Traumatic brain injury: Guidance in a forensic context from outcome, dose-response, and response bias research. *Behavioral Sciences and the Law, 31,* 756-778.

Sweet, J. J., & Nelson, N. W. (2007). Validity indicators within executive function measures: Use and limits in detection of malingering. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 152-177). New York, NY: The Guilford Press.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *Clinical Neuropsychologist, 16*(4), 495-505.

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology, 19,* 455-464.

Tenhula, W. N., & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist, 10*(1), 104-116.

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

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9,* 260-268.

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19,* 203-214.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology, 15,* 578-590.

Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist, 27*(8), 1234-1246.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16,* 45-73.

Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19,* 37-48.

Victor, T. L., Boone, K. B., Serpa, G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23,* 297-313.

Walter, J., Morris, J., Swier-Vosnos, A., & Pliskin, N. (2014). Effects of magnitude of dementia on a symptom validity measure. *The Clinical Neuropsychologist, 28*(7), 1197-1208.

Warrington, E. K. (1984). Recognition Memory Test. Windsor, UK: The NFER-NELSON Publishing Company Ltd.

Wechsler, D. A. (1997). *Wechsler Adult Intelligence Scale (3rd ed.)*. New York, NY: Psychological Corporation.

Wechsler, D. A. (2008). *Wechsler Adult Intelligence Scale (4th ed.)*. San Antonio, TX: Psychological Corporation.

Whiteside, D., Dunbar-Mayer, P., & Waters, D. P. (2009). Relationship between TOMM performance and PAI validity scales in a mixed clinical sample. *The Clinical Neuropsychologist, 23,* 523-533.

Whitfield, K. E., Fillenbaum, G. G., Pieper, C., Albert, M. S., Berkman, L. F., Blazer, D. G., et al. (2000). The effect of race and health-related factors on naming and memory. *Journal of Aging and Health, 12,* 69-89.

Wisdom, N. M., Brown, W. L., Chen, D. K., & Collins, R. L. (2012). The use of all three Test of Memory Malingering trials in establishing the level of effort. *Archives of Clinical Neuropsychology, 27,* 208-212.

Woods, S. P., Weinborn, M., & Lovejoy, D. W. (2003). Are classification accuracy statistics underused in neuropsychological research? *Journal of Clinical and Experimental Neuropsychology, 25*(3), 431-439.

Young, J. C., Sawyer, R. J., Roper, B. L., & Baughman, B. C. (2012). Expansion and re-examination of Digit Span effort indices on the WAIS-IV. *The Clinical Neuropsychologist, 26*(1), 147-159.

Zenisek, R., Millis, S. R., Banks, S. J., & Miller, J. B. (2016). Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Archives of Clinical Neuropsychology, 31,* 426-433.

VITA AUCTORIS

NAME:	Jaspreet Kaur Rai
PLACE OF BIRTH:	Edmonton, Alberta, Canada
YEAR OF BIRTH:	1989
EDUCATION:	University of Alberta, Edmonton, AB 2011, B.Sc. (Specialization) Psychology University of Windsor, Windsor, ON 2014, M.A. Clinical Neuropsychology University of Windsor, Windsor, ON 2018, Ph.D. Clinical Neuropsychology