Loma Linda University

TheScholarsRepository@LLU: Digital Archive of Research, Scholarship & Creative Works

Loma Linda University Electronic Theses, Dissertations & Projects

9-2021

A Review of Malingering Measures in Psychology

Dylan Faire

Follow this and additional works at: https://scholarsrepository.llu.edu/etd

 Part of the Psychology Commons

Recommended Citation

Faire, Dylan, "A Review of Malingering Measures in Psychology" (2021). *Loma Linda University Electronic Theses, Dissertations & Projects*. 1170.
https://scholarsrepository.llu.edu/etd/1170

This Doctoral Project is brought to you for free and open access by TheScholarsRepository@LLU: Digital Archive of Research, Scholarship & Creative Works. It has been accepted for inclusion in Loma Linda University Electronic Theses, Dissertations & Projects by an authorized administrator of TheScholarsRepository@LLU: Digital Archive of Research, Scholarship & Creative Works. For more information, please contact scholarsrepository@llu.edu.

LOMA LINDA UNIVERSITY
School of Behavioral Health
In conjunction with the
Department of Psychology

A Review of Malingering Measures in Psychology

by

Dylan Faire

A Project submitted in partial satisfaction of
the requirements for the degree
Doctor of Psychology

September 2021

© 2021

Dylan Faire
All Rights Reserved

Each person whose signature appears below certifies that this doctoral project in his/her opinion is adequate, in scope and quality, as a dissertation for the degree Doctor of Psychology.

_____, Chairperson

Grace J. Lee, Associate Professor of Psychology

_____

David Vermeersch, Professor of Psychology

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Lee who helped me immensely throughout his whole process. I would not have gotten this far without her guidance and encouragement, and she will always be considered a mentor to me.

I would also like to thank Dr. Vermeersch for his time and input on this project. I appreciated having him as a committee member for this project as his expertise in the subject matter clearly showed.

To my family and friends, your love and support through this long endeavor has been well received and I thank you for your patience with me. We did it, after six years, we did it.

CONTENT

ABBREVIATIONS

PVT	Perfromance Validity Test
TBI	Trauamtic Brain Injury
VSVT	Victoria Symptom Validity Test
APA	American Psychological Association
DSM-5	Diagnostic and Statistical Manual for Mental Disorders (5th Edition)
ADHD	Attention Deficit Hyperactivity Disorder
BPD	Borderline Personality Disorder
DID	Dissociative Identity Disorder
CAPS	Clinician Administered PTSD Scale
CVA	Cerebrovascular Accident
PTSD	Post Traumatic Stress Disorder
RDS	Reliable Digit Span
TOMM	Test of Memory Malingering
MMPI-2	Minnesota Multiphasic Personality Inventory – 2
SIRS	Structured Interview of Reported Symptoms
M-FAST	Miller Forensic Assessment of Symptoms Test
DCT	Dot Counting Test
RMT	Rey Memory Test
WMT	Word Memory Test

ABSTRACT OF THE DOCTORAL PROJECT

A Review of Malingering Measures in Psychology

By

Dylan Faire

Doctor of Psychology, Graduate Program in Psychology
Loma Linda University, September 2021
Dr. Grace Lee, Chairperson

Malingering describes the act of feigning physical or psychological symptoms for external gain, which may include exaggerating symptoms on psychological testing. Thus, standardized measures are needed to detect malingering and invalid responses for accurate interpretation of test results. This review examines existing literature on the most well-known cognitive and psychiatric malingering measures. Optimal contexts, in terms of population and setting, for each measure are examined, and directions for future research and recommendations discussed.

CHAPTER ONE

INTRODUCTION

Malingering occurs when an individual does not respond truthfully or accurately on psychological measures for the purpose of obtaining some external gain, such as monetary compensation or reprieve from legal consequences. According to the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition (DSM-5), an individual can be diagnosed with a malingering disorder if they "intentionally produce false or grossly exaggerated physical or psychological symptoms for personal gain." (American Psychiatric Association, 2013). There are also some cases in which malingering involves an individual attempting to "fake good" or present themselves in a more positive light. These instances can include situations such as custody evaluations, parole suitability hearings, and when wanting to be taken off medication (Gendreau, 1973). In contrast to appearing "ill" some individuals want to be perceived as being healthy for their personal gain. Malingering is often confused with other disorders such as factitious, conversion, or somatoform disorder, which all revolve around suboptimal effort being given, either intentionally or unintentionally, for different reasons. Furthermore, malingering is seen in many different settings and the prevalence rate can vary according to population, as well the setting in which the evaluation occurs.

When discussing the nature of malingering, the feigned symptoms can be either physical or psychological in nature. Feigned psychological symptoms can be further broken down into either cognitive or psychiatric malingering. Both types of malingering have different sets of symptoms, which comprise their respective categories. Psychiatric malingering refers to individuals suspected of feigning or exaggerating psychiatric

symptoms such as depression, personality disorders, anxiety, and psychotic disorders (Carlat, 1998). In contrast, cognitive malingering describes an attempt to feign or exaggerate a cognitive or intellectual impairment, such as difficulties with memory, attention, or completing activities of daily living (Walczyk et al., 2018). The differences between psychiatric and cognitive malingering will be explored further in this review.

Due to the inherently deceptive nature of its presentation, it can be difficult to accurately report the frequency with which malingering occurs in any given population (Sullivan, Lange, & Dawes, 2006). In the medical-legal system where individuals are involved in possible litigation for medical injuries, some studies suggest that approximately 8-21% of individuals will actively malinger (McDermott, Dualan, & Scott, 2013). This effectively means that up to one fifth of individuals involved in these civil cases may exaggerate their presentation in order to gain monetary compensation or some other advantage in a lawsuit. In criminal cases (i.e. assault, murder, robbery, etc.), the rate of malingering is estimated to be as high as 56% (Pollock et al., 1997; Norris et al., 1998). These results suggest that at least one in every two people in these settings will attempt to feign psychological symptoms.

One study was conducted on the base rates of malingering among members of the American Board of Clinical Neuropsychology's practice demographics (Mittenberg, Patton, Canyock, & Condit, 2002). Among 33,351 cases of malingering and symptom exaggeration, approximately 29% of cases were related to personal injury, 30% for disability, 19% for criminal matters, and 8% for medical or psychiatric cases. A follow-up study was conducted to examine the rates of clinical diagnoses in cases of malingering. A survey of members of the Australian Psychological Society and College

of Clinical Neuropsychology yielded a sample of 1,818 cases of malingering in Australia (Mittenberg et al., 2002; Sullivan et al., 2006), out of which mild head injury was found to be the most prevalent diagnosis associated with malingering (23%). Following head injuries, pain or somatoform disorders accounted for 15% of the overall cases, moderate to severe head injuries accounted for 15% as well, and fibromyalgia or chronic fatigue syndrome also accounted for 15%. Given the prevalence of medical diagnoses in the cases reported, it is apparent that malingering is not just a problem for psychologists and lawyers, but a relevant issue for other medical professionals as well.

Detection of malingering is critical for valid interpretation of assessment results, diagnostic accuracy, and treatment recommendations and planning. The proposed project will review the concepts related to malingering as well as the foundational research in detecting malingering. More specifically, this project will review commonly used tests and measures used to detect malingering for a variety of psychological and neuropsychological disorders in various settings.

Aims of the Review:

- To gather a review of malingering tests used in psychology across different settings.
- To describe measures relevant to cognitive malingering and psychiatric malingering
- To provide a reference for clinicians looking to screen for potential malingering in their participants.
- To discuss future directions for research

CHAPTER TWO

METHODS

A comprehensive review of the literature was conducted on malingering and how malingering impacts the field of psychology utilizing several different databases. Specifically, the review utilized the following databases: Complementary Index, Academic Search Premier, Science Direct, PsycINFO, Social Sciences Citation Index, PsycARTICLES, and ERIC. Particular attention was paid to educational and scientific journals, including but not limited to: Archives of Clinical Neuropsychology, Journal of Personality Assessment, Psychological Assessment, Psychological Assessment: a Journal of Consulting and Clinical Psychology, Assessment, Behavioral Sciences and the Law, Journal of Clinical and Experimental Neuropsychology, Applied Neuropsychology, Applied Neuropsychology: Adult, Psychological Injury and Law, Journal of Clinical Psychology, and Journal of Forensic Psychiatry and Psychology. The reference sections of the articles used in this review was examined to uncover additional articles in relation to our main topics as well.

Tests to be reviewed will include:

- Psychiatric Malingering Tests
    - Structured Interview of Reported Symptoms (SIRS)
    - Minnesota Multiphasic Personality Inventory-2 (MMPI-2)
    - Miller Forensic Assessment of Symptoms (M-FAST)
- Cognitive Malingering Tests
    - Victoria Symptom Validity Test (VSVT)
    - Test of Memory Malingering (TOMM)

- Dot Counting Test
- The b Test
- Rey Fifteen Item Test (Rey-15)
- Word Memory Test (WMT)
- Reliable Digit Span (RDS)

A search for additional tests to be reviewed will be conducted using the following key words:

- Malingering / Memory Malingering / Psychiatric Malingering
- Symptom Validity
- Effort / Suboptimal Effort

Future directions in clinical recommendations will be made based upon available literature on our measures.

CHAPTER THREE

LITERATURE REVIEW

According to the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition (DSM-5), malingering disorder is diagnosed when individuals "intentionally produce false or grossly exaggerated physical or psychological problems" for external gain (American Psychiatric Association, 2013). Malingering is seen in many different settings, including both clinical and legal, and the prevalence rate can vary according to setting and population. Moreover, due to the deceptive nature of its presentation, it can be difficult to accurately detect the true prevalence of malingering in select populations (Sullivan, Lange, & Dawes, 2006). In the medical-legal system, studies suggest that approximately 8 to 21 percent of individuals will actively malinger (McDermott, Dualan, & Scott, 2013). This effectively means that up to one-fifth of patients in these circumstances may exaggerate their presentation in order to gain some benefit or advantage in a legal proceeding. Along these same lines, the likelihood of malingering tends to increase the more serious the matter/crime/situation is for the individual. In criminal cases, including charges of assault, murder, and robbery, the rate of malingering was estimated to be as high as 56%, with more incentives (i.e., access to medications, seeking to appear incompetent to stand trial, etc.) associated with higher rates (Pollock et al., 1997; Norris et al., 1998). However, malingering is also observed in clinical settings. A survey of American Board of Clinical Neuropsychology members reported that out of 33,000 clinical cases, 29% of personal injury cases, 30% of disability cases, 19% of criminal cases, and 8% of medical cases involved likely malingering and symptom exaggeration (Mittenburg et al., 2002). Additionally, 39% of mild hand injury cases, 27%

of neurotoxic cases, 31% of chronic pain cases, 35% of fibromyalgia/chronic fatigue cases, and 22% of electrical injury cases resulted in a diagnosis of likely malingering.

Given the potential legal and clinical consequences of misdiagnosing individuals who malinger, the use of validated, standardized measures to detect malingering is of critical importance. The purpose of this study is to review the most well-known and widely used malingering tests. This includes measures designed to detect false reporting or exaggeration of psychiatric symptoms, as well as neurocognitive symptoms. Psychiatric malingering tests to be reviewed include the Structured Interview of Reported Symptoms, the Minnesota Multiphasic Personality Inventory – 2, and the Miller Forensic Assessment of Symptoms. Cognitive malingering tests include the Victoria Symptom Validity Test, Reliable Digit Span, Test of Memory Malingering, The *b* Test, Dot Counting Test, Rey-15 Item Memory Test, and the Word Memory Test.

Psychiatric Malingering Measures

Psychiatric malingering tests are designed to detect feigning or exaggeration of psychiatric symptoms and disorders including mood disorders, personality disorders, and psychotic disorders. Psychiatric malingering is often observed in forensic populations, where criminal defendants are presenting as incompetent to stand trial, in individuals involved in civil litigation, veteran populations aiming to establish or increase their service connection, and other individuals who may have other incentive to present with a psychiatric disorder. Malingering may also manifest itself in clinical populations in an attempt to obtain psychiatric or pain medications, disability benefits, or some other form of external benefit. Due to such situations, psychological measures have been developed

to help clinicians assess for the probability of malingering and symptom exaggeration in various populations. The list of tests below is not exhaustive but represents some of the most widely used tests of psychiatric malingering in the field of psychology with a substantial amount of research evidence to support their validity. It is important to examine these measures and their utility in various settings and populations to determine which measures are the most effective at differentiating between genuine presentations of a psychiatric illness and suspected cases of malingering and can ultimately aid in the most accurate assessment of an individual's psychiatric functioning.

Structured Interview of Reported Symptoms (SIRS)

The SIRS is a tool, which was first developed in 1992 to aid in the assessment of malingering and feigning psychiatric symptoms in clinical and forensic populations. The most recent version was updated in 2010 to the SIRS-2. It is a structured interview composed of 172 items that make up 8 primary scales and 5 supplementary scales. The primary scales include Rare Symptoms (RS), Symptom Combinations (SC), Improbable and Absurd Symptoms (IA), Blatant Symptoms (BL), Subtle Symptoms (SU), Selectivity of Symptoms (SEL), Severity of Symptoms (SEV), and Reported versus Observed Symptoms (RO). These scales help to distinguish between genuine psychopathology and intentional malingerers (Rogers et al., 1992). The supplementary scales include Direct Appraisal of Honesty (DA), Defensive Symptoms (DS), Improbable Failure (IF), Overly Specified Symptoms (OS), and Inconsistency of Symptoms (INC). The items assess for a variety of symptoms, some of which are highly atypical for select disorders or for any psychiatric disorder; thus, the SIRS-2 can identify examinees who indiscriminately

endorse symptoms that are atypical or inconsistent with their reported condition. Total scores are rated into 4 categories: honest, intermediate, probable feigning, or definite feigning. Total scores are used when criterion for feigning based on scale elevations is not met in which multiple scales may be in the probable range of validity. Psychometric properties of this test include a sensitivity of .80 and a specificity of .975; it maintains positive and negative predictive powers of more than .90 (Rogers et al., 2010). This validation data was gathered from a combination of clinical, clinical-forensic, correctional, and community samples (Rogers et al., 2010).

A meta-analysis was conducted to examine the efficacy of the original SIRS in different populations, including criminal offenders, forensic outpatients, community dwelling adults, and inpatients. Malingerers reported a significantly higher number of symptoms on the SIRS compared to genuine responders, with very large effect sizes for the total scores (d = 2.02) and average primary scales (d = 1.53) that exceed standards for assessing effect sizes in malingering research (Green et al., 2011; Rogers 2010). This measure has been widely researched in many populations and has even been found to accurately categorize up to 94.2% of examinees as malingerers reporting PTSD, depression, and anxiety (Green et al., 2012). In a survey of forensic psychologists, it was reported that 86% used the SIRS in every evaluation, and 36% of them used it in at least half of their evaluations (Archer et al., 2006). The SIRS/SIRS-2 is also administered in a wide variety of forensic settings and is often regarded as a gold standard of detecting (Rogers, 2018) feigning of psychiatric symptoms, however, there are some limitations for utility of the SIRS-2.

On the other hand, some researchers have found the SIRS-2's validity and

reliability to be quite questionable. One researcher examined the validity data from the original SIRS and the SIRS-2 and found inconsistent and incomplete data involved in the process of validating the instrument (DeClue, 2011). For one, no demographic information was found for the validity sample used for the original SIRS; thus, it is unclear how generalizable the findings are to other groups and whether demographic factors might affect the interpretation of test scores. Furthermore, the classifications of "genuine, intermediate, and feigned" on the SIRS-2 appeared to be formulated and based on different groups of subjects in different studies which have never been subject to peer review. Several of the validation studies used procedures and methods which either cannot be replicated, or are not consistent with original validation studies for the SIRS-2 (Rogers et al., 2020). This brings into question the level of standardization and scientific rigor in how the data were collected and analyzed, and whether the same classifications can be validly used in other populations.

Another point of criticism was that over half of the "genuine" non-malingering patients in the original sample of the classification were diagnosed with dissociative identity disorder (DID), which is not commonly present in the general population. Furthermore, in a separate study investigating the SIRS among 20 participants diagnosed with DID and 43 "well-informed" and "motivated" DID simulators, both groups were found to endorse a high number of symptoms which categorized them as malingerers (Brand et al., 2006). Lastly, a recent study found the SIRS-2 to have severe limitations when compared to the original SIRS, mainly that the sensitivity was significantly reduced using the SIRS-2 classification rules (Tarescavage et al., 2016). The SIRS-2 was specifically designed to address the issue of false-positives (being classified as a

malingerer) especially in very impaired populations (Brand et al., 2006). However, it appears to have sacrificed sensitivity in order to increase specificity. These limitations may harm the utility of the SIRS-2 in situations where the validity of the research comes into question.

Although the normative data for the SIRS comes from an English-speaking sample (Rogers 1992), translated versions of the SIRS have also been investigated. A Chinese translation of the SIRS-2 was evaluated in two different samples. One was a simulation design utilizing 80 undergraduate subjects (split evenly into a group of honest and feigning responders), and 40 inpatients with mental illness in Xiangya Hospital in Changsha, China. The second sample used a known-groups comparison design with 100 psychiatric outpatients who were split into an honest group (n = 80) and a malingering group (n = 20) based on the Chinese version of the MMPI-2 along with forensic review of police reports and medical records. After administration of the SIRS-2, the researchers found good overall internal consistency between the primary scales in both groups examined (Liu et al., 2013). The sensitivity rate for detecting malingering among the university students was lower than the patient group (60% vs 85% respectively), which the researchers hypothesized as a result of the students being unfamiliar with the nature of mental illness symptoms. In addition, these students were not compensated for their time, which may have led to their decreased performance on the measures (Graue et al., 2007). Overall, the study showed the SIRS-2 demonstrated good validity and reliability in detecting malingering and demonstrated good discriminant validity between the study participants in the first and second sample (d = 1.80).

The validity of a Spanish version was also examined and was found to accurately

distinguish between coached feigners and healthy controls, with large effect sizes observed between groups (Md = 2.0, range of 1.38-2.47) (Correa et al., 2010). The study found an overall higher effect size than in the original English SIRS (*Md* = 2.00 Spanish, and Md = 1.74 English) for the detection of feigned mental disorders, and the alpha coefficients and interrater reliabilities were high across all the primary scales of the Spanish SIRS. This study did, however, use healthy outpatient controls as compared to a forensic population in original validation studies. Nonetheless, the Spanish version appears to be an effective comparison to the English version of the measure.

Although widely used in forensic settings, there are special circumstances in which the SIRS/SIRS-2 should be used with caution. In particular, the presence of an intellectual disability comorbid with psychiatric disorders can affect the validity of the SIRS/SIRS-2 in detecting cases of malingering. Intellectual disabilities complicate testing procedures, particularly when individuals may have difficulty understanding constitutional rights or may be more suggestible in police interrogations (Goldman, 2001). Furthermore, individuals with intellectual disability may not acknowledge functional or cognitive impairments; this can cause an issue with underreporting. Additionally, these individuals may have difficulty understanding test items, which could lead to inaccurate responding to symptom-related questions (Appelbaum, 1994). In such cases, measures of cognitive functioning in combination with psychiatric symptoms should be employed (Boone et al., 2002; Nelson et al., 2003; Weinborn et al., 2003). In an investigation of false positive rates on the SIRS and SIRS-2 among 43 examinees with intellectual disabilities who were not involved in litigation and had no motivations of external gain, 23.3% of examinees were misclassified as malingerers on the original

SIRS, and 7% were misclassified as malingerers on the SIRS-2 (Weiss et al., 2011). Moreover, the rate of false positives was almost three times higher than reported in the SIRS-2 manual (Rogers et al., 2010). When examining the Direct Appraisal of Honesty scale (DA), approximately 54.5% of respondents were classified as true honest responders on the SIRS and 73% were classified as true honest responders on the SIRS-2; this demonstrates the SIRS-2's improvement over the SIRS in terms of classification of feigned psychiatric symptoms in individuals with intellectual disability; however, the SIRS-2 still resulted in up to 27% who may have been misclassified as malingerers. Lastly, the researchers found the scales SU and SEV to also be affected by a comorbid intellectual disability, misclassifying 9.3% and 11.6% of respondents as malingerers, respectively. The findings in this study highlight the importance of the need to conduct a comprehensive evaluation that incorporates other possibly confounding factors such as intellectual ability, rather than relying solely on a single measure of malingering. Particular caution and attention may need to be paid to participants with intellectual disabilities as comprehension, suggestibility, and insight act as factors to an individual's overall presentation.

The SIRS-2 may also have limited validity in detecting malingering of attention deficit hyperactivity disorder. One study examined the validity of the SIRS-2 in individuals instructed to feign ADHD symptoms (simulators), individuals with genuine ADHD (assessed via DSM-V diagnostic criteria), and neurotypical non-ADHD individuals matched to the ADHD group by age, gender, and intellectual functions (Becke et al., 2019). Although specificity of the SIRS-2 was relatively high, ranging from .89 to .93, sensitivity was as low as .19 if indeterminate cases were counted towards

genuine classifications and .25 if the indeterminate cases were excluded before calculating sensitivity. Simulators' responses closely resembled those of the genuine ADHD group and 59.1% of simulators were incorrectly identified as honest responders, while approximately 30% of ADHD patients were incorrectly identified as feigning. Thus, the SIRS-2 does not appear to have adequate sensitivity for the detection of feigned ADHD.

The SIRS should also be used with caution in patients reporting post-traumatic stress disorder (PTSD). The developer of the SIRS conducted research with prison inmates with several psychiatric conditions including PTSD and found that the 8 primary scales of the SIRS were able to discriminate between simulators and genuine presentations of PTSD, demonstrating significant differences between groups of simulators versus true patients ($eta^2 \geq .30$) (Rogers et al., 1992; Hall et al., 2007). However, in another study that examined the correlation between SIRS scores and PTSD symptom severity in military combat veterans, approximately 53% of respondents were classified as exhibiting clear symptom exaggeration on the SIRS (Freeman et al., 2008). In addition, the total severity of PTSD symptoms was positively correlated with SIRS scores ($r = .447$, $P < .001$). Thus, patients with more severe PTSD symptoms may be more likely to be misidentified as malingering. Based on these studies, the SIRS may benefit from further research into its utility with veterans diagnosed with combat-related PTSD.

Overall, the SIRS is effective at evaluating malingering and feigned symptoms in many populations including mood disorders and schizophrenia (Rogers, 1992), but the results must still be interpreted with caution. Additional research is needed in the area of

PTSD and how accurately the SIRS can detect malingering among participants with PTSD. There are varying theories regarding base rates of malingering among veterans seeking service connection, so this may play a factor in the SIRS ability to truly detect malingering. DID is an additional area of research for the SIRS, and at the same time, this is a complex diagnosis with multiple ways of presenting itself in patients. Lastly, the detection of ADHD needs additional research by the SIRS. As mentioned earlier, the most accurate classification of malingering comes from additional data and information, which is not obtained just by malingering measures.

Minnesota Multiphasic Personality Inventory-2 (MMPI-2)

The MMPI-2 is a tool used for the assessment of psychopathology and personality structure. The first version of the MMPI was developed in 1940 and the second version (MMPI-2) was published in 1989. It is one of the most widely used adult measures of personality in the world at the moment (Drayton 2009), and is used in many settings including forensics, medicine, mental health, and medical legal evaluations. It has been validated utilizing populations with PTSD, depression, schizophrenia, mood disorders, and somatic symptom disorders. It is a 567-item self-report questionnaire with responses rated as True or False. The MMPI-2 has 10 primary clinical scales that assess for psychiatric symptoms and personality characteristics include Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychopathic Deviant (Pd), Masculinity-Femininity (Mf), Paranoia (Pa), Psychasthenia (Pt), Schizophrenia (Sc), Hypomania (Ma), and Social Introversion (Si). In addition to clinical scales, several validity scales assess for responses that suggest lying, minimization and/or exaggeration of symptoms, as well as

inconsistent or random response patterns. The validity scales include the Cannot Say (CNS), Variable Response Inconsistency (VRIN), True Response Inconsistency (TRIN), Infrequency (F), F Back (FB), F-Psychopathology (FP), Symptom Validity (FBS), Lie (L), Correction (K), and Superlative Self-Presentation (S). An individual's response pattern on these validity scales not only provide valuable information regarding the accuracy and validity of their responses on other clinical scales, but also provide insight into their motivations and psychological status, and a context within which to interpret their other scores and resulting psychological profile.

Prior research has focused on the F, Fb, and the Fp scales of the MMPI-2 as indicators of feigning and malingering. These 3 scales help to determine if a participant is exaggerating or overreporting psychological symptoms. The F minus K (F – K) index is frequently used in addition to the other validity scales and was originally developed by observing individuals asked to simulate psychoneurosis on the original MMPI (Gough 1947). The overall score distribution cutoffs at the highest 2.5% and lowest 2.5% of the sample (which represent a ± 2 standard deviation in a normal distribution of scores) has been suggested to reflect fake bad and fake good profiles, respectively, on the F-K index (Rothke et al., 1994). However, different clinical populations exhibited significant variability in mean scores on the F-K index, including psychiatric patients (M_{Male} = -1.46, M_{Female} = -2.91) (Docktor et al., 1993), head injury patients (M_{Male} = -7.10, M_{Female} = -9.48), disability benefits patients (M_{Male} = +14.5, M_{Female} = +6.2) and job applicants for police positions (M_{Male} = -19.0, M_{Female} = -19.33)(Rothke et al., 1994). Differences in group means resulted in select groups having a higher percentage of patients being classified as malingerers. Thus, interpretation of the F-K index should take such group

differences into consideration.

In addition to the F-K index, researchers have investigated the F, Fp, and F-K altogether. Several researchers (Arbisi 2006; Arbisi et al., 1995; Butcher et al., 2001) have suggested reading the F and Fp scale sequentially. First, an elevated F scale would suggest an individual either has genuine psychiatric symptoms, is malingering, or is demonstrating invalid responding. The Fp is then examined for an elevated score, which further suggests an individual may be malingering. The MMPI-2 manual suggests a cut score of F ≥ 100 *T* and Fp ≥ 100 *T* for establishing a high probability of malingering (Butcher et al., 2001). If one scale is high and the other one is low, caution should be taken for interpreting test results; it is possible to have a valid administration with one scale high and one scale low. Other scholars have suggested that the Fp scale should be used as a primary scale for detecting feigning (Rogers et al., 2003). In addition, these authors used the dissimulation scale (Ds) which focuses on endorsement of neurotic symptoms. The Ds scale combined with the Fp scale was suggested to ensure the most accuracy when classifying malingerers. A raw score of > 8 on Fp and a raw score of > 35 for Ds resulted in a high likelihood of malingering. In addition to the above models, researchers have spent significant amounts of time examining the F, Fp, F - K which has been discussed earlier; examining these scales seems to work well within correctional settings. It has been suggested that malingering is present if any one of the following scores are obtained: F ≥ 115*T*, Fp ≥ 100*T*, and F – K ≥ 16 raw^2 (Megargee, 2005).

A group of researchers compared all three score combinations (F and Fp; Fp and Ds; F, Fp, and F-K) to find the best model for detecting malingering among the MMPI-2 using inmate simulators and inmates hospitalized for psychiatric treatments (Steffan et

al., 2010). After comparing all three models, the combination of F and Fp demonstrated the best classification statistics and contributed "incrementally to the predictive power of the remaining models in a series of sequential logistic regression analyses." Of course, these results were generated on an inmate population, so generalization to clinical populations may not apply. Inmates with severe psychopathology and possibly significant malingerers may have the best applicability for the study.

The FBS scale detects overstatement of somatic symptoms and over-exaggeration of subjective disability and has been examined in the context of medical impairments. In an investigation of 345 medical inpatients participating in sleep studies, results showed no unique associations between the FBS and sleep/medical variables, and the false positive rate for patients being identified as malingerers was less than 20%, using cut scores of 80-T and 85-T. Male inpatients were more likely to be classified as malingerers than female inpatients (Greiffenstein 2010). The study showed that medical impairments, including obesity, pathological sleepiness, hypoxemia, diabetes, myocardial infarction, and body mass index, did not have a significant influence on the FBS scores. In addition to the FBS, additional scales have been examined to rule out respondents exaggerating symptoms of chronic pain. An MMPI-2 was administered utilizing a group of known malingerers, coached malingerers, and actual chronic pain patients (Bianchini et al., 2018). Results from the testing showed the MMPI-2's validity scales had a "high degree of accuracy" when differentiating malingerers versus non-malingerers. Increased specificity was also observed in most scales besides the L scale. This study was able to demonstrate the reliability of the MMPI-2 to differentiate between exaggerated chronic pain and real chronic pain symptoms in the medico-legal setting.

One group of researchers created a compilation of the MMPI-2 scales to construct a malingering index to be used in the detection of exaggerated symptoms. The malingering discrimination function index (M-DFI) was created by "replicating a mathematical algorithm consisting of weighted clinical, content, and content component scales" (Bacchiochi et al., 2006). The M-DFI was able to distinguish psychiatric patients from participants instructed to feign mental illness. This validation and development study further demonstrates good overall accuracy for this index as a whole. However, other researchers began to compare this index to the MMPI-2's original F scale and found the M-DFI to be overall less effective than the original F scale in predictive utility (Toomey et al., 2008). While the F scale demonstrated acceptable sensitivity and specificity in a criminal forensic population, the M-DFI yielded poor sensitivity. The authors suggest utilizing the original F scales to obtain the best quality measures to assess for malingering.

The validity of the MMPI-2 is compromised when an examinee has previous knowledge about its workings (Bagby et al., 2002). Respondents can also invalidate the use of these scales when taught to identify the specifics of malingering (Storm et al., 2000). When comparing research participants who were coached on how to feign symptoms, non-coached participants who were instructed to feign symptoms, and actual psychiatric patients, the F scales, including F and Fp, were effective in discriminating non-coached malingering participants from psychiatric patients, but were less effective in discriminating between coached malingerers and psychiatric patients (Bury et al., 2002; Storm et al., 2000). Thus, malingerers may receive instruction on how to respond to the MMPI-2 so as to avoid detection, and clinicians should always be mindful of the

possibility of respondents having coaching or prior knowledge of the instruments being administered, particularly in forensic settings. Lastly, an older study examined the effects of coaching for the MMPI-2 And found approximately 50% of lawyers and law students reported feeling a responsibility to warn their clients of the MMPI 2's validity scales (Wetter et al., 1995). Although this was a rough estimate, it is also possible lawyers may not have enough information about the validity scales to effectively coach their clients. However, even small amounts of coaching can skew the overall test results.

One interesting component of malingering is assessing how an individual can become successful at malingering. There are thought to be various factors involved in avoiding detection for psychological measures. One proposed factor for avoiding detection has been general intelligence. It may be reasonably expected that an individual who exhibits a high level of intelligence may have the ability to acquire information on malingering and how the MMPI-2 is designed, which in turn may provide a better chance of avoiding detection when compared to those individuals who have no such knowledge (Kroger et al., 1975). One researcher examined the role of intelligence and knowledge as predictors of avoiding detection on the MMPI-2 (Pelfry 2004). He examined the F and the F through K index of the MMPI-2, the MMPI-2 knowledge questionnaire, as well as the Wonderlic Personnel Test to measure intelligence. Overall, he found individuals with more intelligence and knowledge of the MMPI-2 demonstrated an ability to avoid detection as malingerers over individuals who had no existing knowledge and lower intelligence. Again, this study goes to demonstrate that public domain knowledge of these tests could affect their validity. A detailed review of an individual’s abilities is needed to best understand and rule in the presence of malingering.

Lastly, examining PTSD in the context of the MMPI-2 and malingering is warranted due to veteran presentations for establishing or increasing service connection disabilities. PTSD is a complex diagnosis with differing hypothesized base rates for malingering; this remains a source of controversy today, but an approximate of 30% of compensation seeking veterans have demonstrated malingering (Tolin et al., 2010). Several researchers have found Fp and F - K as the best malingering predictors among civilians diagnosed with PTSD when compared to college students instructed to malinger (Elhai et al., 2001). A group of researchers examined 377 male veterans who were seen at the Behavioral Science Division of the National Center for PTSD. Participants included veterans seeking compensation (n = 290) and those who were not seeking compensation (n = 87); 73.8% of the compensation group and 59.8% of the non-compensation group received a diagnosis of PTSD, and all participants in both groups were assumed to be nonmalingerers. After base rates of malingering in PTSD were examined from the literature and the MMPI-2 was administered, they found the validity scales (Fp, F, FBS, and F-K) had adequate sensitivity, specificity, and efficiency (> .61) under most of the base-rate estimations received from other literature (Tolin et al., 2010). Additionally, compensation-seeking veterans were observed to score significantly higher on all clinical scales of the MMPI-2 besides the masculinity/femininity scale. The study also provided "adequate" cutoff scores, specificity, and sensitivity, for all the validity scales when comparing base-rate estimations. Results also helped debunk an assumption that most compensation-seeking veterans are exaggerating. Overall, the MMPI-2 is a useful tool in many different clinical settings, and still must be interpreted with caution when considering the label of malingering. There are countless validation studies conducted on

the MMPI-2 validity scales to this day attempting to find the most efficient methods for ruling out feigning and malingering.

Miller Forensic Assessment of Symptoms Test (M-FAST)

The M-FAST is a 25-item structured interview designed to assess for potential malingering. It was developed based on several response styles characteristic of psychiatric malingering including reporting of extreme symptoms, rare combinations, unusual hallucinations, unusual symptom course, negative image, suggestibility, and reported versus observed symptoms. It takes approximately 5 to 10 minutes to administer, 10 minutes to score and is widely utilized in different clinical populations and settings. Scores on range from 0 to 15, and the manual recommended cutoff score is 6, which was reported to provide "the highest level of positive predictive power without decreasing negative predictive power to any considerable extent," (Miller, 2001). The test was validated on forensic patients incompetent to stand trial and outpatient civilians being evaluated for a mental illness and disability. These validation studies have also been extended to other populations such as inmates, psychiatric inpatients, and more. It has also been found to be effective in identifying individuals attempting to malinger disorders such as depression, bipolar, psychotic disorders, and posttraumatic stress disorder (Miller, 2005). It has a high test-retest reliability (r = .92, p <.01) and has been found to have similar validity in comparison to the MMPI-2 and SIRS (Rogers et al, 1992).

As described using previous instruments in this review, the measure has also been validated in correctional settings. Utilizing a sample of 50 incarcerated males who applied for mental health services in a prison, these inmates were administered the M-

FAST and SIRS (Guy et al., 2004). Participants who were identified as malingerers based on the SIRS also had significantly higher total and scale scores on the M-FAST, and a cutoff score of 6 resulted in the most accurate classification of malingerers, with a positive predictive power of .78 and a negative predictive power of .89. Importantly, results were similar across race, including African American Hispanic and Caucasian inmates, suggesting adequate generalizability of the M-FAST.

More recently the efficacy of the M-FAST has been examined by utilizing multiple studies with different comparison groups to find a general cutoff score for screening potential malingerers across settings. Using a random-effects model, a group of researchers conducted a meta-analysis of 21 different research reports to examine effect sizes when comparing malingerers (identified by other criterion or simulation groups) to non-malingerers (Detullio et al., 2019). Populations examined included forensic adult samples with severe psychiatric disorders, clinical patients with depression and anxiety, as well as non-clinical, community dwelling participants. The overall sample size was 2120 participants in which the effect size for the difference between the overreporting and comparison groups was observed to be “very large”, for both clinical and nonclinical samples. Using Miller's cutoff score of 6 appeared to provide adequate sensitivity and specificity (sensitivity = .83, specificity = .85). Moreover, the results were similar across studies both with and without coaching, suggesting that the M-FAST may be more robust to the effects of coaching.

Overall, a cutoff score of ≥ 6 in all these studies provided adequate evidence to detect potential malingerers. The score of 6 has been used to detect the malingering of PTSD for combat veterans (Ahmadi et al., 2013), forensic patients (Vitacco et al., 2008),

and among diverse backgrounds (Miller, 2005), while some researchers have found utilizing a score of ≥ 5 is more effective at capturing malingering. After validating the M-FAST with a Spanish speaking population, a group of researchers found a cutoff score of ≥ 5 as having higher sensitivity and specificity (*n* = 102, sensitivity = .93; *n* = 102, specificity = .97) for their population when compared with the cutoff score of ≥ 6 (Montes et al., 2014). Using the cutoff score of 5 correctly classified 94% of the sample (honest, coached, and uncoached) and demonstrated a positive predicted power of .98 and the negative predictive power of .87. These results suggest the necessity for adjusted cutoff scores for non-English speaking populations. It may be up to the clinician to investigate their client's background and culture before using a generalized cutoff score to detect malingering.

After examination of these three commonly used psychiatric malingering measures, several conclusions could be drawn. The SIRS is effective at examining mood disorders as well as schizophrenia and has proven valid in these populations and reliable. It is estimated that over 86% of forensic psychologists use the SIRS in every evaluation due to the prevalence of mood disorders (Archer et al., 2006). The M-FAST is the most time-efficient, taking only 5 to 10 minutes to administer; it can be used as a quick screener in situations to detect the possibility of psychiatric malingering and is shown to be valid and reliable with cutoff scores of 5 and 6 depending on the context. It has also been validated across different racial groups and may be more resistant to coaching. Lastly, the MMPI-2 is effective in many clinical populations and has many scales to detect a wide variety of presenting problems including severe psychopathology. It is effective at assessing mood disorders, PTSD, head injuries, and more. As described, each

test has its own utility and usefulness and must be determined at the end of the day by the clinician.

Cognitive Malingering Measures

Cognitive malingering refers to the feigning of symptoms related to a cognitive disorder such as a learning disability, ADHD, injury-related cognitive impairment, dementia or general cognitive dysfunction. Cognitive malingering is typically seen in the context of neuropsychological testing and is regularly assessed for in these settings (Pace et al., 2019). Cognitive malingering can be assessed by comparing an individual's performance with the degree and type of impairment that would typically be expected of someone with the reported condition, based on clinical research, and considering the context of one's clinical history and their level of functional abilities. Some tests also employ the use of items that appear to be more difficult than they are. The best performance validity tests (PVTs) are basic (not complex/little if any psychological jargon), understood by multiple populations (i.e., grade-school level reading skills), and difficult to fail (Heilbronner et al., 2009).

Victoria Symptom Validity Test

The Victoria Symptom Validity Test is a freestanding performance validity test (PVT) developed for use in multiple clinical settings. It is one of the most widely used and recognized measures used to detect the exaggeration or feigning of cognitive impairments (Sharland et al., 2007). It is a computer-administered test in which an examinee is given a series of two-item forced choice recognition questions involving a

string of five numbers. Stimuli are presented as being “Easy” or “Hard” based on the complexity of the string being presented, and items are scored based on latency of response and correct identification of a digit string in each trial. Cutoff scores for different clinical populations have been used with varying degrees of accuracy. According to the manual, scores ≥ 16 on Hard or Easy items are interpreted as reflecting “likely valid” performance, 8-15 items correct reflect “questionably valid” performance, and ≤ 7 items are interpreted as “invalid” performance (Slick et al., 1997). Sensitivity for these scores was 81% while specificity was 100%. Additionally, research on non-litigating patients with memory impairment reported 82% sensitivity and 94% specificity of the VSVT using the following cutoff scores on Easy or Hard items: “Compliant” from 21 to 24, “questionably compliant” from 16 to 20, and “probably not compliant” from 8 to 15 (Strauss et al., 2006).

However, research in various clinical populations has found that different cutoff scores may be warranted for different settings. For example, in VA hospitals, veterans can be seen for a compensation and pension exam to assess service connection eligibility, which is tied to the amount of compensation a veteran receives for a given disability. The authors determined that a cutoff score of ≤ 19 for Hard items yielded excellent sensitivity (.91) and specificity (.93) for accurately discriminating between non-malingerers and probable-to-definite malingerers (Jones, 2013). For civilian populations with mild TBI involved in litigation, a cutoff of <18 Hard items correct was found to be most effective at identifying malingering, yielding .68 sensitivity and .90 specificity (Silk et al., 2014). This study also found that using a cutoff of < 23 Easy items correct with the same population yielded a lower sensitivity of .35, but good specificity of .95. In a study

comparing healthy controls, individuals asked to simulate ADHD, and individuals asked to simulate a reading disorder, a cutoff score of < 20 on Hard items was observed to be effective at detecting simulators, with 80% sensitivity and 100% specificity for ADHD and 89.3% sensitivity and 100% specificity for reading disorder simulators (Frazier et al., 2008). Easy items were not included due to them generally being less effective at detecting malingering in the study and showing weaker group discrimination than the hard items.

The presence of complex mental health issues can affect cognitive functions and further complicate interpretation of performance on PVTs. Some mental illnesses, such as borderline personality disorder (BPD) and bipolar disorder, are associated with cognitive dysfunction that can affect one's overall clinical profile (Erdodi et al., 2017). This effect is referred to as psychogenic interference. Some of the most complex mental illnesses revolve around personality disorders and trauma, including borderline personality disorder. BPD is characterized by an unstable sense of self and chaotic interpersonal relationships that cause significant distress and impairment in daily life. When VSVT scores of examinees with BPD were compared to those of healthy controls and non-compensation seeking examinees with a brain dysfunction, results showed BPD patients had slower response times and more errors compared to the other groups (Ruocco, A.C., 2016). Within the BPD group, 2% of examinees were classified as probably not complaint scores, 10% as questionably compliant, and 88% compliant utilizing the Strauss cutoff scores. These results suggest that individuals with BPD or other personality disorders may be more prone to being identified as malingerers and highlight the necessity to consider personality disorders in the assessment of malingering and

possibly allow extra time/consideration to examinees.

As illustrated in the examples above, there is no general consensus on a cutoff score that can be used universally for every population and doing so can result in misclassification of examinees as malingerers despite the presence of real cognitive impairments. Additionally, this points to the need for screening of examinees for co-occurring substance use disorders, active suicidal/homicidal ideation, personality disorders and other conditions that may be associated with cognitive impairment. These factors, along with co-morbid complex medical conditions, can influence how an individual participant scores on the VSVT. In a sample of 120 epilepsy patients undergoing a neuropsychological evaluation, 34 exhibited "invalid" or "questionably valid" scores utilizing a cutoff of ≤ 20 for questionably valid responses and ≤ 18 for invalid (Loring et al., 2005, Loring et al., 2007). Additionally, a significant group effect for age was observed in which older patients typically performed worse on the VSVT. These cutoff scores were based on an earlier study with a sample of 30 epilepsy patients that found all patients scored at least 18/24 correct (Grote et al., 2000). These results further the argument that different cutoff scores are needed for each population, and the manualized cutoff scores have limitations in their validity. Lastly, these results also show that older individuals may be at risk of being classified as exhibiting suboptimal effort on the VSVT.

A more recent systematic review of the VSVT was conducted utilizing 17 peer reviewed studies which included 7 simulation designs, 7 differential prevalence designs, and 3 known groups designs, and an independent cross validation sample of 200 mixed clinical neuropsychiatric patients referred for psychological evaluation (Resch et al.,

2021). Clinical diagnoses included bipolar disorder, anxiety, ADHD, depression, substance use disorders, and more. All participants completed the VSVT and additional PVTs including the TOMM, Digit Span, WMT, Rey 15-Item, Dot Counting Test, and the *b* test; participants with no criterion failures (i.e., no test scores that fell in the invalid range) were classified into a valid group (n = 153) while participants with ≥ 2 criterion failures were classified into the invalid group (n = 47). Overall, the study found a cutoff score of ≤ 22 on Easy items (sensitivity = .53, specificity = .92), ≤ 16 on Difficult items (sensitivity = .57, specificity = .90), and ≤ 40 for Total items (sensitivity = .62, specificity = .88) to be the most effective cutoff scores for this sample. Likewise, when examining response latency (an uncommon and less researched scoring method), researchers found ≥ 2.50 on easy items (sensitivity = .49, specificity = .92), ≥ 4.15 on difficult items (sensitivity = .36, specificity = .89, and ≥ 3.07 on total response latency (sensitivity = .51, specificity = .90) to be the most effective cut off times in terms of response latency. As shown in the numbers, response latency indices produced a generally lower sensitivity rate, which brings into question the utility of response latency indices to supplement the detection of malingering. Overall, the VSVT has been shown to be effective in detecting malingering in different clinical populations with the use of cutoff scores specifically designed for various diagnostic groups.

Reliable Digit Span (RDS)

The Reliable Digit Span (RDS) is a statistically validated performance validity measure for assessing malingering of cognitive impairment in many different populations (Meyers et al., 1998). This embedded measure is taken directly from the Digit Span (DS)

subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) wherein a participant is read a series of numerical digits and asked to recall them in either forward, backward, or sequential order. The RDS score is derived from the longest sequence completed correctly on the DS Forward Trial, summed with the longest sequence completed correctly on the DS Backwards Trial. Unlike the VSVT, the RDS is not a standalone freestanding measure, but is rather an "embedded" measure that is derived from performance on a test (Digit Span) that is designed to assess another construct (attention and working memory).

The "gold standard" in terms of the optimal RDS cutoff score for detecting malingering in a clinical population has been widely researched. After several independent studies, an RDS score of 7 has emerged as the "gold standard" cutoff score for detecting probable malingering in several clinical populations (Axelrod et al., 2006; Zenisek et al., 2016). Another published study attempted to examine different scores and markers used for detecting malingering and negative response bias (detailed below) in a clinical sample of 2,400 individuals presenting with a variety of problems such as traumatic brain injury, alcohol abuse, and Alzheimer's disease (Iverson et al., 2003). In this sample, a score of 4 or lower on the RDS was rarely observed in any of the diagnostic groups examined in the study (2%). Conclusions drawn from the study were aimed at providing a framework for detecting probable negative response bias, which refers to systematically poor performance that is not consistent with genuine neurocognitive compromise and can arise due to fatigue, anxiety, or the presence of psychiatric disturbances (Denney, 2008). The following cutoff scores were reported to indicate suspected negative response bias: an overall score on the RDS of five or less; a

longest digit span forward (LDSF) of 4 or less for those younger than 55, and an LDSF of 3 or less for those 55 and older; and a longest digit span backwards (LDSB) of 2 or less, regardless of age. The traumatic brain injury group was noted as having the lowest scores among all the clinical problems. The cutoff scores found in this study further help to elaborate on the point that scoring abnormally low on the RDS measure can be a great first step in the detection of malingering and negative response bias. Additionally, lower cutoff scores may need to be employed for individuals with comorbid psychiatric conditions and cognitive dysfunction. Lastly, low scores may be indicative of low effort and negative response bias rather than intentional malingering.

As mentioned above, TBI patients have exhibited lower scores on the RDS compared to other clinical groups. In a study investigating the validity of the RDS among patients with mild TBI, those who were involved in litigation were compared to mild TBI patients who were not involved in litigation; both groups experienced less than one hour loss of consciousness from a TBI (Meyers et al., 1998). The participants were administered various neuropsychological tests, including Digit Span. A subset of participants also completed another PVT, the forced choice task (FC), which served as the criterion for malingering. The TBI group involved in litigation had a mean RDS score of 8.02, while the non-litigation group had a mean RDS score of 9.59, and the litigation group generally scored lower across all neuropsychological tests. Using the standard cutoff score of 7 or less on the RDS, 4.1% of participants in the non-litigation group were identified as malingering while 48.9% of participants in the litigation group were identified as malingerers. Notably, none of the non-litigation participants failed the FC; however, 9 of the 40 (22.5%) litigating participants failed the FC, and 7 of these 9

(77.8%) participants also failed the RDS (< 7). Lastly, 16 participants in Group 1 were classified as malingerers by the RDS who were not classified as malingerers by the FC. Thus, the RDS classified more participants as malingerers than the FC in both the litigating and non-litigating groups, suggesting greater sensitivity of the RDS, but may also reflect a higher rate of false positives in this clinical population.

The use of RDS in dementia populations, however, should be used with caution. In a sample of 579 neuropsychological examinees (M_{age} = 72.8), individuals who were diagnosed with Vascular dementia (n = 8), Dementia with Lewy Bodies (n = 27), Alzheimer's disease (n = 133), and Frontotemporal dementia (n = 15) performed significantly worse than other examinees who were diagnosed with Parkinsonian syndromes, Mild Cognitive Impairment, and other unspecified cognitive disorders (Zenisek et al., 2016). The standard cutoff of RDS ≤ 7 resulted in a high rate of false positives (29.7%). Even with the use of lower cutoff scores, 12.8% of all participants scored ≤ 6, and 4.3% scored ≤ 5. Additional studies have examined the RDS classification accuracy for dementia and found that specificity was low. One study found only 30% of individuals with Alzheimer's disease obtaining a passing score using a cutoff score of ≤ 7 (Merten et al., 2007). An additional study examined the RDS in 228 participants with severe memory disorders and found the cutoff score of ≤ 7 to yield a specificity of 68% which is below acceptable levels (Heinly et al., 2005).

Despite its limitations with regards to dementia populations, the RDS does have other advantages. For instance, reliable generalization of the RDS to other cultures has been established by various studies with cutoff scores varying by plus or minus one point (Jasinski et al., 2011). Scores on the RDS in the U.S. were compared with scores in other

countries such as China, Spain, and Japan and it was found that the sensitivity and specificity rates of the RDS were similar across regions. However, the need for correct translation and standardization when used in foreign language speaking countries was noted. The RDS may be effective and valid cross-culturally which adds to its overall efficacy for clinicians.

The RDS has also been theorized to be immune to the Flynn effect, which refers to the phenomenon that intelligence test scores have been observed to steadily rise over time, across several different countries. Past studies have shown that healthy adults are typically able to encode, on average, 7 items +/- 2 on simple short-term working memory tasks such as Digit Span, and 4 items +/- 1 on more complex working memory tests such as the Continuous Performance Test and Test of Memory Malingering (Miller, 1956; Cowan, 2001). In order to investigate the Flynn effect, a contemporary study examined change in average digit span forward and backward scores over a period of 80 years and found that the average number of items remained constant at 7 and 4 for simple and complex working memory tests, respectively, and that those numbers did not fluctuate significantly over time (Gignac, 2015). The digit span forward has been researched as an effective measure of short-term memory capacity and digit span backwards has been shown as an effective verbal working memory capacity measure (Oberauer et al., 2000).

Other studies examined focused on the topic of whether embedded effort measures perform like standalone effort measures and how valid the embedded effort measures are when compared to the standalone ones (Miele et al., 2011). Four standalone PVT's (e.g., Rey-15, TOMM, WMT, and VSVT) as well as 17 embedded validity indices, including the RDS, were investigated in a sample of 44 participants with mild

TBI (mean age = 45, men = 52%). After testing, the participants were grouped into either a suspected malingering group or a normal functioning group based on if they failed at least 2 of the standalone PVTs described above. Among the 17 embedded measures, the RDS was observed to have the best classification accuracy as well as sensitivity and specificity. The results also found that embedded measures when used alone are not sufficient to detect true malingering without the use of other standalone effort measures. It was recommended that at least two other standalone effort measures be included to rule in the possibility of true malingering. The Board of Directors of the American Academy of Clinical Neuropsychology created a "Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias and Malingering" which recommends using multiple effort measure of effort which tap into different cognitive domains which includes both embedded and standalone validity indices (Board of Neuropsychology, 2007). Therefore, the use of only standalone measures is also insufficient, and a combination of both embedded and standalone measures is recommended. The use of embedded measures may seem attractive to some clinicians (i.e., shorter administration time, less probability of coaching), but research does not support only using embedded measure or replacement of standalone effort measures for embedded ones.

Overall, the RDS has been shown to be an effective embedded measure of malingering, especially when used simultaneously with other standalone measures of malingering detection. The Reliable Digit Span is useful with assessing for cognitive impairment and traumatic brain injury but not recommended with participants with the presence of dementia due to increasing the risk of misinterpreting genuine cognitive

impairment as invalid performance.

Test of Memory Malingering (TOMM)

The Test of Memory Malingering is one of the most frequently used performance validity tests in the field of psychology (Tombaugh, 1996; Sharland et al., 2007). The Test of Memory Malingering is made up of 2 Learning Trials and an optional Delayed Retention Trial. The TOMM is available in both paper-and-pencil and computerized versions. The TOMM was validated on a group of 475 community-dwelling adults with ages ranging from 17 to 84 years old with no known history of neurological disease, psychiatric illness, head injury, and stroke (healthy controls) and 187 patients who underwent neuropsychological assessments and were diagnosed with no cognitive impairment (n = 13), aphasia (n = 21), cognitive impairment (n = 42), traumatic brain injury (n = 45), depression (n = 26), and dementia (n = 40). Most (96%) healthy controls demonstrated a perfect score of 50 on Learning Trial 2 and the Delayed Retention Trial, and rarely scored below 45, which led to the cutoff score of < 45. This criterion score identified 100% of the healthy controls as well as 95% of nondemented clinical patients as non-malingerers. The dementia group overall obtained an average score of 92% correct on the TOMM, but among moderate-to-severe dementia participants 10% scored below 40 on Trial 2, suggesting the TOMM may not be appropriate for dementia patients in more severe stages (Tombaugh, 1996). Nevertheless, the validation study served to demonstrate the validity of the TOMM to detect malingering in individuals with mild dementia who experience significant memory deficits, in addition to other clinical diagnostic groups.

The TOMM manual suggests < 45 for Trial 2 as an indicator of invalid responding, but several researchers have found this cutoff less sensitive to invalid responding than other PVT's. One study used the Word Memory Test (WMT) in conjunction with the TOMM and found the cutoff score of 45 to only detect 44% of individuals who failed the WMT (Greiffenstein et al., 2008). In addition, another study found the 45 cutoff to correctly identify 21% of litigation-seeking claimants who failed the WMT, 32% of individuals who failed the Nonverbal Medical Symptom Validity Test (NV-MSVT), and 35% of individuals who failed the Medical Symptom Validity Test (MSVT) (Green et al., 2008; Green et al., 2004). Lastly, an additional researcher evaluated the TOMM against criterion measures from multiple PVT's and found that it correctly identified 98% of patients with optimal efforts but detected less than half of those patients with suboptimal effort (Kulas et al., 2014). Overall, the cutoff score of 45 has proven to be an ineffective and more conservative cutoff score then some of the other scores examined in this review.

A group of researchers examined utilizing the cutoff score of < 45 to detect malingering in traumatic brain injuries (n = 84, 70.2% Mild TBI) versus psychiatric conditions (n = 68), such as depression, anxiety, and bipolar disorder (Erdodi et al., 2017). The TOMM Learning Trial 2 scores were compared against the WMT, the NV-MSVT, and the Erodi Index – Five (EI-5). Examining different cut offs from < 45 and up, the psychiatric condition group was half as likely as the TBI group to be classified as malingering using the escalating range of cut off scores. The standard cutoff of <45 demonstrated high specificity (.96-1.00), but low and variable sensitivity (.15 to .25 for psychiatric conditions and .21 to .50 for TBI's). Using a cutoff score of ≤ 49

demonstrated a sensitivity range of .38 to .67 and a specificity range of .89-.96 in psychiatric populations, and sensitivity of .42-.67 and specificity of .83-.91 in the TBI group. The TOMM Trial 2, using a cutoff score of ≤ 49, was able to detect approximately half of all patients who failed the other three PVTs. These results suggest that while the TOMM provides good specificity to minimize the number of participants being misclassified as malingerers, the sensitivity is low. Therefore, other measures may be needed to confirm or rule out the presence of malingering.

In another study of nonstandard cutoff scores in a TBI population, researchers examined Trial 1 and Trial 2 scores of the TOMM using cutoffs of ≤ 39 and < 49, respectively, in a sample of 44 patients referred to a private neuropsychological practice for mild TBI (Stenclik et al., 2013). Performance validity tests including the Rey 15-Item Test, the VSVT, Word Memory Test, and the RDS were included as criterion measures. After completion of testing, participants were categorized into performance-valid and performance-invalid groups in (based on performance of criterion measures previously mentioned). The Trial 2 cutoff score of < 49 demonstrated the highest sensitivity with relatively high specificity as well (sensitivity = .75, specificity = .92) as compared to both Trial 1 ≤ 39 (sensitivity = .60, specificity = .96) and the standard cut off score of < 45 for Trial 2 (sensitivity = .40, specificity = 1.00). Although high specificity was found utilizing different cutoffs, the highest sensitivity was observed using a cutoff of < 49 on Trial 2, which leads to further support of utilizing nonstandard cutoff scores in the assessment of malingering.

Research in the TOMM has heavily focused on Trial 2 as a predictor of invalid performance as individual differences in attention that may affect Trial 1 scores have less

of an influence on Trial 2 performance, after examinees have had multiple exposures to the target stimuli (Tombaugh, 1996). In an analysis of archival data collected from a sample of 77 mild TBI litigants seen in private practices, 45% of litigants scored 45 or greater on Trial 1, whereas 68% scored above 45 on Trial 2 (Gavett et al., 2005). Similarly, in the original validation data from the TOMM manual, 61% of participants scored above 45 on Trial 1, whereas 90% scored above 45 on Trial 2. Moreover, 100% of individuals who scored above 45 on Trial 1 subsequently scored higher than 45 on Trial 2 and Retention Trials. The results from this study demonstrate that good performance on Trial 1 leads to good performance on subsequent trials of the TOMM. However, several studies have examined using Trial 1 of the TOMM as a potentially more useful index than Trial 2 for detecting invalid effort (Denning et al 2013; Denning 2012; Kulas et al 2014; Schroeder et al 2011). One study examined this claim utilizing a total of 1,198 evaluees undergoing outpatient assessment were administered the TOMM, along with 2 other performance validity tests - the Word Memory Test, and the Computerized Assessment of Response Bias (CARB); 70% of the sample also had data from the California Verbal Learning Test - Second Edition Forced Choice Recognition Trial (Mossman et al., 2018). The study found that a significant number of participants with valid scores on the TOMM Trials 1 and 2 failed several of the other performance validity tests, and that perfect scores on the TOMM cannot rule out the possibility of malingering. However, utilizing a higher cutoff score for Trial 2 yielded better results at identifying individuals who are feigning symptoms as compared to using Trial 1 scores. This study demonstrated that Trial 1 may not be as/more effective than Trial 2 in detecting malingering on its own.

However, these interpretations must be made with caution and cannot be made solely based on this one performance validity measure. Comparing with other performance validity tests helps to increase the confidence of ruling in suboptimal effort and malingering.

The b Test

The b Test is a performance validity measure that consists of a booklet of 15 pages each containing 72 different stimuli including lowercase b's as the target, and lowercase d's, p's, q's, and morphed b's as distractors (Boone et al., 2002). Stimuli are presented in increasingly smaller formats and fonts throughout the booklet, and the examinee is scored based on the time they take to complete the test as well as the number of correct b's they circle. The overall completion time and total errors are combined to calculate an Effort Index (E-score) which is used to identify individuals exhibiting suboptimal effort. The manual suggests an E-score ≥ 90 as indicative of malingering in participants with TBI's. The b test has been shown to be effective at identifying highly suspicious individuals as malingering based on the observation that examinees with cognitive impairment and cerebral dysfunction are generally able to pass the b test. Additionally, the b test has also been effectively used to distinguish genuine psychiatric conditions such as depression and schizophrenia from definite malingering.

The b test was originally validated on a group of 34 suspected malingerers and 161 other subjects from various clinical populations including older depressed patients (*n* = 38), elderly controls (*n* = 17), head injury patients (*n* = 14), patients with left or right hemisphere cerebrovascular accidents (CVA) (*n* = 8), patients with learning disabilities (*n*

= 38), and patients with schizophrenia (*n* = 28) (Boone et al., 2000). The malingering group was comprised of individuals who exhibited suspect effort on two of the following six effort measures: Rey 15-Item Test, Rey Dot Counting Test, Rey Word Recognition, Warrington Recognition Memory Test-Words, Digit Span, and the Rey Auditory Verbal Learning Test. Overall results demonstrated suspected malingerers performed significantly worse on the b test as compared to the other clinical comparison groups including head injuries and patients with CVA's.

The b test, VSVT, and TOMM have been shown to be effective in a Spanish population with post concussive syndrome (PCS), including a PCS group involved in litigation, a PCS group not involved in litigation, and a healthy control group of simulated malingerers (fourth year psychology students with knowledge of brain damage and neuropsychology). The VSVT, TOMM, and the b test all demonstrated good sensitivity, identifying 97.1%, 88.6% and 80%, respectively, of the simulated malingering group as malingerers, and identifying 100%, 100% and 91.7%, respectively, of the non-litigation PCS group as non-malingerers (Vilar-Lopez et al., 2007). Additionally, the b test has been studied in the context of mild traumatic brain injuries. In a study by the same above author, four groups were examined; a group of non-compensation-seeking participants, a group of compensation-seeking participants (not suspected of malingering), a group of compensation-seeking participants suspected of malingering, and a group of simulated malingerers. After completion of the b test, the study demonstrated statistically significant differences between the malingering and the non-malingering groups. The b test correctly identified 81.3% of the non-malingering compensation-seeking group and 93% of non-compensation seeking participants as non-

malingerers. However, the *b* test classified only 37.5% and 27.6% of suspected malingers and simulated malingerers, respectively, as malingerers. This study demonstrated some initial efficacy of the b test in examining TBI participants and provided results that call for future research into increasing sensitivity for the *b* test in a TBI population (Vilar-Lopez et al., 2008).

The b test appears to perform well in older adult populations with mild cognitive impairment (MCI). Patients with MCI appear to perform similarly to healthy controls, suggesting that patients with genuine MCI can still pass the b test without being misclassified as a malingerer (Shandera et al., 2010). One study compared b test scores across 3 groups: one group with mild neurocognitive disorder (n = 21), healthy older adults asked to feign symptoms of mild neurocognitive disorder with standardized instruction (n = 21), and an age-matched healthy control group (n = 21). Results showed the malingering group performed significantly worse than both the control group and the diagnosed mild neurocognitive disorder group on error scores but scored comparably to the mild neurocognitive disorder group on response times. (Pace et al., 2019). The mild neurocognitive disorder group scored significantly worse on all scores than healthy controls, but both groups showed the same pattern of more omission errors than commission errors on the *b* test. In contrast, the malingering group demonstrated more commission errors than omissions, particularly more d errors (choosing a "*d*" over a "*b*") and omission errors on the b test. Overall, this study demonstrated the b test's ability to accurately distinguish malingerers from individuals diagnosed with mild neurocognitive disorder.

The Dot Counting Test

The Dot Counting Test (DCT) was originally designed and developed in 1941 as a test to detect cognitive malingering (Boone, Lu & Herzberg, 2002). The original format consists of 12 three by five-inch cards each containing a different number and arrangement of dots. The first set of six cards have randomly assigned and arranged dots while the second half of cards contains grouped dots. A participant is instructed to count the dots on the cards as quickly as possible while minimizing the number of errors. The test score (E-score) is based on both the response times and number of counting errors. This test is given in a wide variety of clinical contexts such as forensic populations, neuropsychological evaluations, community samples, and more.

The original study of the DCT examined the sensitivity and specificity of the DCT in several different subject groups (Boone et al., 2002b). Eighty-six clinical patients with noncredible symptoms were referred for neuropsychological assessment; these subjects were currently in litigation or seeking to obtain medical disability benefits for problems associated with medical or psychiatric disorders. In addition, these patients demonstrated non-credible effort on at least 2 of the following 6 performance validity tests: Rey 15-Item Test, Rey Word Recognition Test, The b Test, Warrington Recognition Memory Test-Words, Digit Span, and the Rey Auditory Verbal Learning Test. These participants also met behavioral conditions such as a pattern of neuropsychological scores inconsistent with profiles of genuine persons with medical or psychiatric conditions and indicating severity of symptoms significantly worse than genuine profiles of psychiatric or medical conditions. Additionally, participants also had implausible self-report symptoms and contradictions between self-report measures and

documented medical records and history. In addition to the 86 clinical patients, 14 additional forensic inpatient subjects were selected from Atascadero State Hospital in California. These participants were included in the group if they exhibited evidence of secondary gain, evidence of malingering on previous psychiatric testing, requests for medication, and abnormal reactions to medications.

These two study groups were compared with comparison groups composed of healthy controls and individuals with genuine learning disabilities, depression, head injury, cerebrovascular accidents, schizophrenia, and mild dementia. Notably, the moderate-severe dementia group was excluded for taking on average more than 30 seconds to complete the grouped dots counting task which would misidentify these participants as malingerers; this is consistent with other literature citing poor performance for moderate-severe dementia groups on other malingering tests (Arnett et al., 1995; Schretlen et al., 1991). Using a cutoff of E-scores ≥ 17 resulted in 100% sensitivity in the forensic suspect group, 75% sensitivity in the civil litigation and disability suspect effort group, and a specificity of $\geq 90\%$ for the clinical groups combined. These data demonstrate the ability of the DCT to detect noncredible cognitive symptoms in litigation/disability and forensic participants. Sensitivity and specificity rates of the DCT appear to be high even in clinical groups with genuine, measurable cognitive impairments.

More recent studies have examined different DCT scoring approaches in mixed clinical neuropsychiatric samples. A cross-sectional study sample of 132 patients with mixed neuropsychiatric conditions, including depression, chronic pain, bipolar disorder, anxiety, attention deficit/hyperactivity disorder, and substance use disorders, completed

the DCT along with 5 other criterion performance validity tests: Advanced Clinical Solutions Word Choice Test, Medical Symptom Validity Test, TOMM, Stroop Color-Word Test Word Reading, and the WAIS-IV Digit Span (Rhoads et al., 2021). Participants with ≤ 1 criterion PVT failure were categorized as valid (N = 102) and participants with ≥ 2 criterion PVT failures were classified as invalid (N = 30). Within the performance-valid group, 37 participants had normal scores on neuropsychological tests (cognitively intact) and 65 portrayed objective evidence of mild cognitive impairment (MCI). The traditional DCT E-score was examined, the DCT unrounded E-score, and the E-score for alternate DCT versions with four cards (DCT 4) and six cards (DCT 6) (Bailey et al., 2019). The DCT 6 is calculated by adding the mean completion time of ungrouped cards 1, 3, and 5 with grouped cards 8, 11, and 12 plus total errors on these six cards; the DCT 4 is calculated by adding the mean completion time of ungrouped cards 3 and 4 with grouped cards 8 and 11 plus the total errors on these four cards. Results showed an optimal cutoff score was ≥ 15.25 for the DCT 4 (sensitivity = .63, specificity = .89), and ≥ 14.83 for the DCT 6 (sensitivity = .67, specificity = .90) across all three groups. The DCT 4 had optimal psychometric properties of all four scores among the MCI group (sensitivity = .60, specificity = .91) as opposed to the DCT 6 which had a sensitivity of .43. Optimal cutoffs recommended in the literature are ≥ 15 (Soble et al., 2018) and ≥ 17 (Boone et al., 2002b) for the DCT E-score and the DCT unrounded cutoff of ≥ 13.80 (McCaul et al., 2018); this study found higher optimal cutoffs for the E-score (≥ 18) and the unrounded score (≥ 17.325) possibly due to the inclusion of more participants with cognitive impairments as compared two existing literature samples. Overall, the four scoring methods did not show significant differences for the cognitively

intact group, but the DCT 4 demonstrated better psychometric properties in cognitively impaired individuals. These findings support the claim that the DCT 4 scoring method is likely one of the most effective methods for detecting suspect effort among a wide range of clinical populations and presentations.

Rey 15-Item Memory Test (RMT)

The Rey 15-Item Memory Test (RMT) is a psychological measure used often to detect malingering in clinical populations. It is composed of 15 characters printed on one card which is presented to the examinee for 10 seconds, after which they are asked to immediately write down as many of the items as they can remember. In actuality, the test is presented in such a way so that it is meant to seem more difficult than it is; however, due to the simplicity of the symbols, and the redundancy among these items, even participants with significant impairments can perform well on this test (Rey, 1958). Normative observations of patients with memory disorders were gathered from 100 temporal lobe epilepsy patients demonstrating poor performance on 1 of 4 standardized memory tests compared with 16 neurological outpatients involved in litigation and 40 neurological outpatients not involved in litigation (Lee et al., 1992). Forty-two percent of the epilepsy patients achieved a perfect score of 15 and 96% scored 8 or above. In the non-litigation outpatient group, 27.5% of participants achieved a score of 15 and 95% obtained scores of 8 or above. In contrast, only 6.3% of participants in the litigation outpatient group had a perfect score of 15 and 62.5% achieved a score of 8 or above. The cutoff score of 7 on the RMT was at or below the 5th percentile for the epilepsy and non-litigation groups, but not the litigation-involved outpatient group which performed

significantly worse than the other two groups. This study demonstrates the ability of the RMT to detect differences in performance between neurological patients with and without external sources of motivation.

Studies in the past have found significant utility from the RMT even with severe psychiatric disorders. A group of researchers found a cutoff score of < 9 effective at identifying individuals exaggerating memory impairment in a population with severe psychiatric disorders (Goldberg et al., 1986). Additionally, a few researchers examined normal controls compared with participants with traumatic brain injuries and upheld the cutoff score of < 9 for identifying potential memory difficulties and impairment (specificity of .94) (Bernard et al., 1990). The cutoff score of 9 has been effective at differentiating healthy controls from neurological disorders (specificity of .96)(Lee et al., 1992), healthy controls asked to feign memory impairment from non-litigating psychiatric inpatients with depression (specificity of 1.0)(Guilmette et al., 1994), diagnosed malingerers from forensic impatience acquitted by reason of insanity (specificity of .96, sensitivity .05)(Simon, 1994), and older adults with major depressive disorder (specificity of .95)(Lee, et al., 2000). Overall, these studies demonstrated good specificity in these populations however sensitivity is another issue.

The cutoff scores for the RMT were examined in further detail with five outpatient participants with severe brain injuries referred for neuropsychological assessment at a clinic associated with a trauma center; all patients scored a 15 out of 15 on the RMT. (Taylor et al., 2003). All five of the participants, aged 20 to 58, performed poorly on standardized neuropsychological tests assessing memory, learning, and motor speed and dexterity while achieving a perfect score on the RMT. There are differing

opinions in the literature regarding cut off scores, as discussed previously 7 and 9 are some of the most prominent in literature. This research suggests a higher cutoff score may serve as a valid conservative score. Overall, the RMT has been a useful and valid screener for detecting the possibility of malingering; however, research suggests using higher cutoff scores for increasing sensitivity as much as possible; caution is needed as well as higher cutoffs may result in lower specificity rates.

A meta-analysis was conducted utilizing 13 studies which examined the RMT (Reznek, et al., 2005). Using the recommended cutoff score of 9, 837 out of the 983 genuine patients were correctly identified as non-malingerers (specificity of .85), but only 70 out of 192 known malingerers were classified as malingering (sensitivity .36). Lowering the cutoff score to 8 resulted in a slightly higher specificity, but decreased sensitivity (sensitivity .09). The high specificity rates of the RMT result in lower false positives or misidentifying people as malingerers when they are not malingering. The low sensitivity observed in this meta-analysis or the ability of the test to accurately identify patients which the test is designed to identify leads to mixed results. A lower cutoff score results in missing participants who may in fact be malingering, and this leads to low rates of false positives. However, a higher cutoff score results in better sensitivity while also running the risk of incorrectly identifying non-malingers as real malingerers. The relative importance of high specificity versus high sensitivity of a measure will depend on the relative severity of consequences associated with a false positive (incorrectly labeling someone with genuine impairments as a malingerer) compared to a false negative (failing to detect a malingerer). In a legal system in which participants may be diagnosed incorrectly as malingerers, specificity may be the important variable to examine here.

Overall, the Rey 15-Item Memory Test demonstrated high specificity across different populations which may be more important than having high sensitivity.

Word Memory Test (WMT)

The Word Memory Test (WMT) measures verbal learning and memory and is designed to evaluate an individual's level of effort during neuropsychological testing. The test is composed of a list of 20 semantically-related word pairs (e.g., "dog-cat", "pig-bacon", "fish-fin", etc.) either presented verbally or via computer screen. After being shown the word pairs, the participant is presented with 40 new word pairs and asked to identify which word in the pair was from the original list of 20 (e.g. "pig" from "pig-squirrel"). The WMT includes an Immediate Recognition trial followed by a 30-minute Delayed Recognition Trial. Following administration of these procedures, four additional subtests are conducted including the Multiple Choice (match one word with its corresponding pair word among eight choices), Paired Associates (recall the second paired word without cues), Delayed Free Recall (recall as many words from the original list in any order), and Long Delayed Free Recall (free recall after 20 minutes has passed since Delayed Free Recall). Three scores are calculated from the Immediate recognition, Delayed recognition, and the consistency between Immediate and Delayed recognition; these three scores determine an individual's level of effort. The Word Memory Test has been effective in a variety of clinical populations and presentations.

In a sample of 40 healthy controls compared with 57 participants with moderate to severe TBI, the mean Immediate and Delayed Recognition scores from the WMT were above 95% or 38 out of 40 correct. The healthy controls composed of undergraduate

college students were 97.8% correct which is slightly higher than the brain injury participants (Green et al., 1999). Likewise, an additional study examined 15 patients with moderate to severe TBI's who were involved in head injury litigation (Iverson et al., 1999). These participants were given the WMT, and the Computerized Assessment of Response Bias (CARB), a computerized symptom validity test. None of the subjects demonstrated obvious evidence of exaggeration as measured by the CARB. The average score of these participants on the Immediate Recognition was 95.4% correct, 95.4% correct for Delayed Recognition, and 92% correct for the subtests. Psychology student simulators (n = 15) were also informed of the purpose of the WMT and were told to make their best attempt at simulating memory impairment without being detected by the test as faking. Simulation was assumed to be present if the person scored below 82% correct on Delayed recognition. All but one of these simulators scored below the cut off of 82%; there was also low consistency between immediate and delayed recognition trials. The results from this study show the WMT is effective in detecting simulated memory impairment, and insensitive to genuine cognitive impairments experienced by patients with moderate to severe TBI.

An additional study examined the TOMM and the WMT together as they are commonly utilized performance validity tests, and both have proven effective at detecting invalid performance. Some researchers even suggest that the WMT is more sensitive when compared to the TOMM (Bhowmick et al., 2019). A total of 268 examinees from clinical private practice, many of whom were disability claimants involved in litigation, were administered the WMT and the TOMM as part of a neuropsychological evaluation with tests including the Wechsler Memory Scale, WAIS-R, WRAT-3, CVLT-II, and the

Rey Complex Figure. The study found that 74% of participants passed both the WMT and TOMM, while the remaining 26% failed the WMT but passed the TOMM; there were no participants who passed the WMT but failed the TOMM. Individuals who passed the TOMM but failed the WMT exhibited lower overall test scores on the other neuropsychological measures described above. The results from this study show that the WMT may be more sensitive in predicting non-credible cognitive impairments than the TOMM. These findings help inform which measures of performance validity may be more effective in the detection of malingering.

CHAPTER FOUR

DISCUSSION AND CLINICAL RECOMMENDATIONS

Taking into account all the measures, it is useful to examine the overall utility of both psychiatric and cognitive malingering measures discussed in this review. Regarding psychiatric malingering measures, the M-FAST is an effective screener tool, which is cost-effective, quick, and efficient at detecting malingering of psychiatric disorders in forensic and clinical populations alike. Most studies described above validated it within multiple populations, and further demonstrated it was resistant to coaching. Likewise, the MMPI-2 is effective at detecting a wide range of psychiatric disorders including PTSD, schizophrenia, and various other mood disorders. The MMPI-2 takes longer to score and administer but provides overall good psychometric properties in terms of its validity. The MMPI-2 can provide a more complete profile on a participant as opposed to the M-FAST. When the clinician has time, access to the measure, and money, the MMPI-2 may be a more effective tool to assess for malingering as well as detect psychiatric illness. The SIRS is used by forensic psychologists in a wide variety of settings, but mostly relating to a forensic context. It is slightly shorter than the MMPI-2 resulting in wide clinical applications. However, several of the scales have the tendency to misclassify participants with comorbid intellectual disabilities as malingerers. Additionally, the newer SIR-2 was not effective at detecting feigned ADHD. More attention needs to be paid to PTSD and ADHD for the SIRS and SIRS-2 as several studies resulted in misclassifying genuine true responders as malingerers.

As far as cognitive malingering goes, the VSVT has various cutoffs used for different diagnostic groups in the detection of feigned cognitive impairment. It is

considered a “gold standard” for the detection of feigned cognitive impairment, and the studies described above help support this claim. The RDS is an easy-to-use embedded validity measure, which is difficult to be coached on. It is also used for detection of malingered cognitive impairment. When used in conjunction with other standalone or embedded measures, it was shown to be an effective predictor of malingering and negative response bias. The TOMM is effective in individuals with mild cognitive or memory impairments but loses sensitivity at lower cutoff scores. In addition, it is not as effective at assessing for malingering in those with more moderate to severe cognitive impairments. The b test is an additional test, which is relatively quick to administer and score, and performed well with individuals with severe cognitive dysfunction including dementia. The Dot Counting Test can effectively detect noncredible cognitive symptoms in litigation, disability claimants, and forensic populations. The Rey 15-Item Test demonstrated higher specificity than sensitivity at lowered cutoff scores but remained effective with detecting feigned memory impairment. This is an additional screener measure, which may be used in conjunction with other cognitive malingering tests to fully assess the presence of malingering. Lastly, the WMT is a relatively easy measure, which is difficult to fail; as such, it can be used in cases in which extreme exaggeration of cognitive impairment may be suspected. The WMT is also more effective and sensitive at detecting lower overall test scores in full neuropsychological batteries than the TOMM.

Regarding limitations, most of these studies either did not report race and if reported, most of the samples used were majority Caucasian. This may lead to a misclassification of participants of minority races and ethnicities as malingerers when they are genuine responders. Future research is needed with more diverse participants to

establish if these studied validity scores and effort scores are still effective. Additionally, this review does not include an exhaustive list of malingering measures, and there are more measures one could look at to examine malingering. Some of the measures discussed here are some of the most common measures used which is why we examined the validities and utilities. Future research can be aimed at examining different malingering measures in specific populations and how certain measures compare to the manual recommendations for a particular test. Additionally, there is a lack of literature on difference in the rates of cognitive vs psychiatric malingering in different populations and future research may want to examine the prevalence rates of psychiatric and cognitive malingering in different settings (i.e. personal injury, criminal cases, disability, etc.).

Lastly, the measures discussed in this review are helpful in the detection of malingering and suboptimal effort. To the untrained clinician, these measures may even be considered essential. It is the clinician's job to determine whether malingering is present or not based on not just the measures, but based on all the data available (e.g. behavioral observations, historical data, referral sources, etc.). It is important to note that measures alone are not sufficient to diagnose malingering (Mason et al., 2014). In addition to standalone and embedded measures, a clinical interview is necessary to examine how an individual presents to the visit. Collateral information from external sources is an excellent method to track someone's clinical history, this can include medical records, information obtained from family members or friends, and information from previous clinical providers. This ensures a clinician will have all the available data to work on with a participant. Should all the data be unavailable, the clinician can make a judgment over the validity of the measures based on the scores and how the individual

presents. Malingering is an issue which will continue to see research as participants attempt to obtain external benefits from testing, and with knowledge of these testing measures, clinicians can be well equipped to interpret results on these tests.

REFERENCES

Ahmadi, K., Lashani, Z., Afzali, M. H., Tavalaie, S. A., & Mirzaee, J. (2013). Malingering and PTSD: Detecting malingering and war related PTSD by Miller Forensic Assessment of Symptoms Test (M-FAST). *BMC Psychiatry*, *13*(1), 1–5. https://doi.org/10.1186/1471-244x-13-154

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Appelbaum, K. L. (1994). Assessment of Criminal-Justice-Related Competencies in Defendants with Mental Retardation. *The Journal of Psychiatry & Law*, *22*(3), 311–327. https://doi.org/10.1177/009318539402200301

Arbisi, P. A. (2005). MMPI-2: A Practitioner's Guide. In J. N. Butcher (E.), *Use of the MMPI-2 in personal injury and disability evaluations* (Illustrated ed., pp. 407–441). American Psychological Association (APA).

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale, F(p). *Psychological Assessment*, *7*(4), 424–431. https://doi.org/10.1037/1040-3590.7.4.424

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A Survey of Psychological Test Use Patterns Among Forensic Psychologists. *Journal of Personality Assessment*, *87*(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07

Arnett, P. A., Hammeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-item test in neurological patients and malingerers. The Clinical Neuropsychologist,9, 17–26.

Aronoff, G. M., Mandel, S., Genovese, E., Maitz, E. A., Dorto, A. J., Klimek, E. H., & Staats, T. E. (2007). Evaluating Malingering in Contested Injury or Illness. *Pain Practice*, *7*(2), 178–204. https://doi.org/10.1111/j.1533-2500.2007.00126.x

Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting Incomplete Effort with Digit Span from the Wechsler Adult Intelligence Scale—Third Edition. *The Clinical Neuropsychologist*, *20*(3), 513–523. https://doi.org/10.1080/13854040590967117

Bacchiochi, J. R., & Bagby, R. M. (2006). Development and Validation of the Malingering Discriminant Function Index for the MMPI–2. *Journal of Personality Assessment*, *87*(1), 51–61. https://doi.org/10.1207/s15327752jpa8701_04

Bagby, R. M., Nicholson, R. A., Bacchiochi, J. R., Ryder, A. G., & Bury, A. S. (2002). The Predictive Capacity of the MMPI-2 and PAI Validity Scales and Indexes to Detect Coached and Uncoached Feigning. *Journal of Personality Assessment*, *78*(1), 69–86. https://doi.org/10.1207/s15327752jpa7801_05

Bailey, K. C., Webber, T. A., Phillips, J. I., Kraemer, L. D. R., Marceaux, J. C., & Soble, J. R. (2019). When Time is of the Essence: Preliminary Findings for a Quick Administration of the Dot Counting Test. *Archives of Clinical Neuropsychology*, *36*(3), 403–413. https://doi.org/10.1093/arclin/acz058

Becke, M., Fuermaier, A. B. M., Buehren, J., Weisbrod, M., Aschenbrenner, S., Tucha, O., & Tucha, L. (2019). Utility of the Structured Interview of Reported Symptoms (SIRS-2) in detecting feigned adult attention-deficit/hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology*, *41*(8), 786–802. https://doi.org/10.1080/13803395.2019.1621268

Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology*, *46*(4), 432–436. https://pubmed.ncbi.nlm.nih.gov/2212046/

Bhowmick, C., Hirst, R., & Green, P. (2019). Comparison of the Word Memory Test and the Test of Memory Malingering in detecting invalid performance in neuropsychological testing. *Applied Neuropsychology: Adult*, 1–11. https://doi.org/10.1080/23279095.2019.1658585

Bianchini, K. J., Aguerrevere, L. E., Curtis, K. L., Roebuck-Spencer, T. M., Frey, F. C., Greve, K. W., & Calamia, M. (2018). Classification accuracy of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2)-Restructured form validity scales in detecting malingered pain-related disability. *Psychological Assessment*, *30*(7), 857–869. https://doi.org/10.1037/pas0000532

Board of Directors of American Academy of Clinical Neuropsychology. (2007). Practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, 21, 209–231.

Boone, K. (2000). Validation of a New Technique to Detect Malingering of Cognitive Symptoms The b Test. *Archives of Clinical Neuropsychology*, *15*(3), 227–241. https://doi.org/10.1016/s0887-6177(99)00014-1

Boone, K., Lu, P., and Herzberg, D. S. (2002). *The b Test manual*. Los Angeles: Western Psychological Service. A

Boone, K. B., Lu, P., & Herzberg, D. (2002). The Dot Counting Test. Los Angeles, CA: Western Psychological Services. B

Brodaty, H., Clarke, J., Ganguli, M., Grek, A., Jorm, A. F., Khachaturian, Z., & Scherr, P. (1998). Screening for cognitive impairment in general practice: toward a consensus. *Alzheimer Disease and Associated Disorders*, *12*(1), 1–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9539404

Brand, B. L., McNary, S. W., Loewenstein, R. J., Kolos, A. C., & Barr, S. R. (2006). Assessment of Genuine and Simulated Dissociative Identity Disorder on the Structured Interview of Reported Symptoms. *Journal of Trauma & Dissociation*, *7*(1), 63–85. https://doi.org/10.1300/j229v07n01_06

Brandt J, Rubinsky E, Lassen G. Uncovering malingered amnesia. Ann N Y Acad Sci. 1985;444:502-3. doi: 10.1111/j.1749-6632.1985.tb37625.x. PMID: 3860110.

Bury, A. S., & Bagby, R. M. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment*, *14*(4), 472–484. https://doi.org/10.1037/1040-3590.14.4.472

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2. An administrative and interpretive guide*. Minneapolis: University of Minnesota Press.

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

Carlat, D. J. (1998). The psychiatric review of symptoms: a screening tool for family physicians. *American Family Physician*, *58*(7), 1617–1624. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9824959

Chacón, F., Godino, A., Lösel, F., Santolaya, F., Sierra, J. C., Tortosa, F., … Wexler, D. (2012). In Seacrh of a Fast Screening Method for Detecting the Malingering of Cognitive Impairment. *The European Journal of Psychology Applied to Legal Context Eur. J. Psychol. Appl. Legal Context*, *4*(42), 99–196. Retrieved from www.usc.es/sepjf

Conroy, M. A., & Kwartner, P. P. (2006). Malingering. *Applied Psychology in Criminal Justice*, *2*(3), 29–51. http://dev.cjcenter.org/_files/apcj/2_3_Malingering.pdf

Correa AA, Rogers R, Hoersting R. Validation of the Spanish SIRS with monolingual Hispanic outpatients. *Journal of Personality Assessment*. 2010 Sep;92(5):458-64. doi: 10.1080/00223891.2010.497430. PMID: 20706932.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/s0140525x01003922

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, *23*(1), 133–152. https://doi.org/10.1080/13854040701819050

DeClue, G. D. (2011). Harry Potter and the Structured Interview of Reported Symptoms? *Open Access Journal of Forensic Psychology*, *3*, 1–17. https://a091bcda-ef2e-423a-8e26-23fa0ffe04ed.filesusr.com/ugd/166e3f_8e12c7ab6cb54b0b9343b2d18e318230.pdf

Denning, J. H. (2012). The Efficiency and Accuracy of The Test of Memory Malingering Trial 1, Errors on the First 10 Items of The Test of Memory Malingering, and Five Embedded Measures in Predicting Invalid Test Performance. *Archives of Clinical Neuropsychology*, *27*(4), 417–432. https://doi.org/10.1093/arclin/acs044

Denning, J. H. (2013). Combining the Test of Memory Malingering Trial 1 With Behavioral Responses Improves the Detection of Effort Test Failure. *Applied Neuropsychology: Adult*, *21*(4), 269–277. https://doi.org/10.1080/23279095.2013.811076

Detullio, D., Messer, S. C., Kennedy, T. D., & Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment*, *31*(11), 1319–1328. https://doi.org/10.1037/pas0000753

Docktor, B., O'Sullivan, P., Greene, R., Arredondo, R., & Mann, A. (1993 March). *Effects of setting, gender, and code-type on the MMPI-2 Addiction Potential Scale and Addiction Acknowledgement Scale*. Paper presented at the meeting of the Society for Personality Assessment, San Francisco.

Drayton, M. (2009). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). *Occupational Medicine*, *59*(2), 135–136. https://doi.org/10.1093/occmed/kqn182

Elhai, J. D., Gold, S. N., Sellers, A. H., & Dorfman, W. I. (2001). The Detection of Malingered Posttraumatic Stress Disorder with MMPI-2 Fake Bad Indices. *Assessment*, *8*(2), 221–236. https://doi.org/10.1177/107319110100800210

Erdodi, L. A., & Rai, J. K. (2017). A single error is one too many: Examining alternative cutoffs on Trial 2 of the TOMM. *Brain Injury*, *31*(10), 1362–1368. https://doi.org/10.1080/02699052.2017.1332386

Erdődi, L. A., Seke, K. R., Shahein, A., Tyson, B. T., Sagar, S., & Roth, R. M. (2017). Low scores on the Grooved Pegboard Test are associated with invalid responding and psychiatric symptoms. *Psychology & Neuroscience, 10*(3), 325–344. https://doi.org/10.1037/pne0000103

Frazier, T., Frazier, A., Busch, R., Kerwood, M., & Demaree, H. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology, 23*(5), 501–509. https://doi.org/10.1016/j.acn.2008.04.001

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, *158*(3), 374–380. https://doi.org/10.1016/j.psychres.2007.04.002

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. https://doi.org/10.1207/s15324826an1201_1

Gervais, R., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, *19*(4), 475–487. https://doi.org/10.1016/j.acn.2003.05.001

Gendreau, P., Irvine, M., & Knight, S. (1973). Evaluating response set styles on the MMPI with prisoners: Faking good adjustment and maladjustment. *Canadian Journal of Behavioural Science, 5*(2), 183-194. http://dx.doi.org/10.1037/h0082343

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence*, *48*, 85–95. https://doi.org/10.1016/j.intell.2014.11.001

Goldberg, J. O., & Miller, H. R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42*(5), 792–795. https://doi.org/10.1002/1097-4679(198609)42:5<792::AID-JCLP2270420519>3.0.CO;2-8

Goldman, M. (2001). Systematic treatment of criminal offenders having developmental delay. *NADD Bulletin*, *4*(3), 56–57.

Gough, H. G. (1947). Simulated patterns on the Minnesota Multiphasic Personality Inventory. *The Journal of Abnormal and Social Psychology, 42*(2), 215–225. https://doi.org/10.1037/h0063295

Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of Feigned Mental Retardation Using the New Generation of Malingering Detection Instruments: Preliminary Findings. *The Clinical Neuropsychologist*, *21*(6), 929–942. https://doi.org/10.1080/13854040600932137

Green, D., & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment*, *23*(1), 95–107. https://doi.org/10.1037/a0021149

Green, D., Rosenfeld, B., & Belfi, B. (2012). New and Improved? A Comparison of the Original and Revised Versions of the Structured Interview of Reported Symptoms. *Assessment*, *20*(2), 210–218. https://doi.org/10.1177/1073191112464389

Green P. (2008). *Green's non-verbal medical symptom validity test user's manual*. Edmonton, Canada: Green's Publishing.

Green P. (2004) *Green's medical symptom validity test user's manual*. Edmonton, Canada: Green's Publishing.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, *13*(10), 813–819. https://doi.org/10.1080/026990599121205

Greiffenstein, M. F. (2010). The MMPI-2 Symptom Validity Scale (FBS) Not Influenced by Medical Impairment: A Large Sleep Center Investigation. *Assessment*, *17*(2), 269–277. https://doi.org/10.1177/1073191109358823

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. https://doi.org/10.1037/1040-3590.6.3.218

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 failure concordance rates. *Archives of Clinical Neuropsychology*, *23*(7–8), 801–807. https://doi.org/10.1016/j.acn.2008.07.005

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*, *22*(6), 709–719. https://doi.org/10.1076/jcen.22.6.709.958

Guilmette, T. J., Hart, K. J., Giuliano, A. J., & Leininger, B. E. (1994). Detecting simulated memory impairment: Comparison of the rey fifteen-item test and the hiscock forced-choice procedure. *Clinical Neuropsychologist*, *8*(3), 283–294. https://doi.org/10.1080/13854049408404135

Guy, L. S., & Miller, H. A. (2004). Screening for Malingered Psychopathology in a Correctional Setting. *Criminal Justice and Behavior*, *31*(6), 695–716. https://doi.org/10.1177/0093854804268754

Hall, R. C. W., & Hall, R. C. W. (2007). Detection of Malingered PTSD: An Overview of Clinical, Psychometric, and Physiological Assessment: Where Do We Stand? *Journal of Forensic Sciences*, *52*(3), 717–725. https://doi.org/10.1111/j.1556-4029.2007.00434.

Hathaway, S. R., & McKinley, J. C. (1951). *Minnesota Multiphasic Personality Inventory; manual* (Revised). Psychological Corporation

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. https://doi.org/10.1177/1073191105281099

Hiscock, C. K., Branham, J. D., & Hiscock, M. (1994). Detection of feigned cognitive impairment: The two-alternative forced-choice method compared with selected conventional tests. *Journal of Psychopathology and Behavioral Assessment*, *16*(2), 95–110. https://doi.org/10.1007/bf02232721

Iverson, G., Green, P., & Gervais, R. (1999). Using the word memory test to detect biased responding in head injury litigation. *Journal of Cognitive Rehabilitation*, *17*(2), 4–8.

Iverson, G. L., & Tulsky, D. S. (2003). Detecting malingering on the WAIS-III: Unusual Digit Span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology*, *18*(1), 1–9. https://doi.org/10.1016/S0887-6177(01)00176-7

Jones, A. (2013). Victoria Symptom Validity Test: Cutoff Scores for Psychometrically Defined Malingering Groups in a Military Sample. *The Clinical Neuropsychologist*, *27*(8), 1373–1394. https://doi.org/10.1080/13854046.2013.851740

Keary, T. A., Frazier, T. W., Belzile, C. J., Chapin, J. S., Naugle, R. I., Najm, I. M., & Busch, R. M. (2013). Working Memory and Intelligence Are Associated with Victoria Symptom Validity Test Hard Item Performance in Patients With Intractable Epilepsy. *Journal of the International Neuropsychological Society*, *19*(3), 314–323. https://doi.org/10.1017/s1355617712001397

Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology*, *43*(1), 48–55. https://doi.org/10.1037/h0076266

Kulas, J. F., Axelrod, B. N., & Rinaldi, A. R. (2014). Cross-Validation of Supplemental Test of Memory Malingering Scores as Performance Validity Measures. *Psychological Injury and Law*, *7*(3), 236–244. https://doi.org/10.1007/s12207-014-9200-4

Lee, A., Brauer Boone, K., Lesser, I., Wohl, M., Wilkins, S., & Parks, C. (2000). Performance of Older Depressed Patients on Two Cognitive Malingering Tests: False Positive Rates for the Rey 15-item Memorization and Dot Counting Tests. *The Clinical Neuropsychologist*, *14*(3), 303–308. https://doi.org/10.1076/1385-4046(200008)14:3;1-p;ft303

Lee, G. P., Loring, D. W., & Martin, R. C. (1992). Rey's 15-item visual memory test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment*, *4*(1), 43–46. https://doi.org/10.1037/1040-3590.4.1.43

Lezak, M. D., (1995). *Neuropsychological assessment* (3rd ed.) New York: Oxford University Press.

Liu, C., Liu, Z., Chiu, H. F., Carl, T. W., Zhang, H., Wang, P., Wu, G., Mwansisya, T. E., Cao, L., Hu, A., Wang, Y., & Xue, Z. (2013). Detection of malingering: psychometric evaluation of the Chinese version of the structured interview of reported symptoms-2. *BMC psychiatry*, *13*, 254. https://doi.org/10.1186/1471-244X-13-254

Loring, D. W., Larrabee, G. J., Lee, G. P., & Meador, K. J. (2007). Victoria Symptom Validity Test Performance in a Heterogenous Clinical Sample. *The Clinical Neuropsychologist*, *21*(3), 522–531. https://doi.org/10.1080/13854040600611384

Loring, D. W., Lee, G. P., & Meador, K. J. (2005). Victoria Symptom Validity Test Performance in Non-Litigating Epilepsy Surgery Candidates. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 610–617. https://doi.org/10.1080/13803390490918471

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

Mason, A. M., Cardell, R., & Armstrong, M. (2014). Malingering psychosis: guidelines for assessment and management. *Perspectives in Psychiatric Care*, *50*(1), 51–7. https://doi.org/10.1111/ppc.12025

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, 29 (3), 308–318.

Meyer, R. G., & Deitsch, S. M. (1995). The assessment of malingering in psychodiagnostic evaluations: Research-based concepts and methods for consultants. *Consulting Psychology Journal: Practice and Research*, *47*(4), 234–245. https://doi.org/10.1037/1061-4087.47.4.234

McCaul, C., Boone, K. B., Ermshar, A., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2018). Cross-validation of the Dot Counting Test in a large sample of credible and non-credible patients referred for neuropsychological testing. *The Clinical Neuropsychologist*, *32*(6), 1054–1067. https://doi.org/10.1080/13854046.2018.1425481

McDermott, B. E., Dualan, I. V., & Scott, C. L. (2013). Malingering in the correctional system: Does incentive affect prevalence? *International Journal of Law and Psychiatry*, *36*(3), 287–292. https://doi.org/10.1016/j.ijlp.2013.04.013

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. (1948). The Minnesota Multiphasic Personality Inventory: VI. The K Scale. *Journal of Consulting Psychology*, *12*(1), 20–31. https://doi.org/10.1037/h0061377

Megargee, E. (2006). *Using the MMPI-2 in Criminal Justice and Correctional Settings* (First edition). Univ Of Minnesota Press.

Megargee, E. L. (2005). MMPI-2: A Practitioner's Guide. In J. N. Butcher (Ed.), *Use of the MMPI-2 in correctional settings.* (Illustrated ed., pp. 327–360). American Psychological Association (APA).

Meyers, J. E., & Volbrecht, M. (1998). Validation of Reliable Digits for Detection of Malingering. *Assessment*, *5*(3), 303–307. https://doi.org/10.1177/107319119800500309

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2011). Are Embedded Validity Indices Equivalent to Free-Standing Symptom Validity Tests? *Archives of Clinical Neuropsychology*, *27*(1), 10–22. https://doi.org/10.1093/arclin/acr084

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, H. A. (2001). The Miller Forensic of Symptoms Test (M-FAST) professional manual. Odessa, FL: Psychological Assessment Resources.

Miller, H. A. (2005). The Miller-Forensic Assessment of Symptoms Test (M-Fast) Test Generalizability and Utility Across Race, Literacy, and Clinical Opinion. *Criminal Justice and Behavior*, *32*(6), 591–611. https://doi.org/10.1177/0093854805278805

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base Rates of Malingering and Symptom Exeggeration. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, *24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Montes, O., & Guyton, M. R. (2014). Performance of Hispanic inmates on the Spanish Miller Forensic Assessment of Symptoms Test (M-FAST). *Law and Human Behavior*, *38*(5), 428–438. https://doi.org/10.1037/lhb0000074

Mossman, D., Wygant, D. B., Gervais, R. O., & Hart, K. J. (2018). Trial 1 versus Trial 2 of the Test of Memory Malingering: Evaluating accuracy without a "gold standard". *Psychological Assessment*, *30*(1), 74–85. https://doi.org/10.1037/pas0000449

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships Between Eight Measures of Suspect Effort. *The Clinical Neuropsychologist*, *17*(2), 263–272. https://doi.org/10.1076/clin.17.2.263.16511

Norris, M. P., & May, M. C. (1998). Screening for malingering in a correctional setting. *Law and Human Behavior*, 22, 315–323.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). *Individual differences in working memory capacity and reasoning ability.* In A. R. A. Conway, C. Jarrold, M. J. Kane (Eds.) & A. Miyake & J. N. Towse (Ed.), *Variation in working memory* (p. 49–75). Oxford University Press.

Pace, G., Orrù, G., Monaro, M., Gnoato, F., Vitaliani, R., Boone, K. B., Gemignani, A., & Sartori, G. (2019). Malingering Detection of Cognitive Impairment With the b Test Is Boosted Using Machine Learning. *Frontiers in Psychology*, *10*, 1–8. https://doi.org/10.3389/fpsyg.2019.01650

Pelfrey, W. V. (2004). The Relationship Between Malingerers' Intelligence and MMPI-2 Knowledge and Their Ability to Avoid Detection. *International Journal of Offender Therapy and Comparative Criminology*, *48*(6), 649–663. https://doi.org/10.1177/0306624x04265085

Pollock, P., Quigley, B., Worley, K., & Bashford, C. (1997). Feigned mental disorder in prisoners referred to forensic mental health services. *Journal of Psychiatric and Mental Health Nursing*, 4, 9–15.

Resch, Z.J., Webber, T.A., Bernstein, M.T. *et al.* Victoria Symptom Validity Test: A Systematic Review and Cross-Validation Study. *Neuropsychol Rev* (2021). https://doi.org/10.1007/s11065-021-09477-5

Rey, A. (1958). *L'examen clinique en psychologie.* [The clinical examination in psychology]. Presses Universitaries De France.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, *19*(7), 539–543. https://doi.org/10.1080/02699050400005242

Rhoads, T., Resch, Z. J., Ovsiew, G. P., White, D. J., Abramson, D. A., & Soble, J. R. (2021). Every second counts: A comparison of four dot counting test scoring procedures for detecting invalid neuropsychological test performance. *Psychological Assessment*, *33*(2), 133–141. https://doi.org/10.1037/pas0000970

Rogers, R. (2012). Research response styles. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception*, (3rd ed., pg 301-322). The Guilford Press.

Rogers, R. (2018b). Structured interviews and dissimulation. In R. Rogers & S. D. Bender (Eds.), Clinical assessment of malingering and deception (4th ed., pp. 422–448). New York, NY: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). Structured Interview of Reported Symptoms (SIRS) and professional manual. Odessa, FL: Psychological Assessment Resources.

Rogers, R., Kropp, P. R., Bagby, R. M., & Dickens, S. E. (1992). Faking specific disorders: A study of the structured interview of reported symptoms (SIRS). *Journal of Clinical Psychology*, *48*(5), 643–648. https://0-doi-org.catalog.llu.edu/10.1002/1097-4679(199209)48:5<643::AID-JCLP2270480511>3.0.CO;2-2

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). Structured Interview of Reported Symptoms (SIRS), 2nd Edition, Professional manual. Lutz, FL: Psychological Assessment Resources, Inc.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of Feigned Mental Disorders. *Assessment*, *10*(2), 160–177. https://doi.org/10.1177/1073191103010002007

Rogers, R., Velsor, S. F., & Williams, M. M. (2020). A Detailed Analyis of SIRS Versus SIRS-2 Critiques. *Psychological Injury and Law*, *13*(3), 275–283. https://doi.org/10.1007/s12207-020-09379-6

Rothke, S. E., Friedman, A. F., Dahlstrom, W. G., Greene, R. L., Arredondo, R., & Mann, A. W. (1994). MMPI-2 Normative Data for the F-K Index: Implications for Clinical Neuropsychological, and Forensic Practice. *Assessment*, *1*(1), 1–15. https://doi.org/10.1177/1073191194001001002

Ruocco, A. C. (2016). Compliance on neuropsychological performance validity testing in patients with borderline personality disorder. *Psychological Assessment*, *28*(3), 345–350. https://doi.org/10.1037/a0039481

Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-item memory test to detect malingered amnesia. *Psychological Assessment*: *A Journal of Consulting and Clinical Psychology*, 3, 667–672.

Schroeder, R. W., Baade, L. E., Peck, C. P., & Heinrichs, R. J. (2011). Use of Test of Memory Malingering Trial 1 as a measure of response bias. *The Clinical Neuropsychologist*, 26, 564.

Shandera, A. L., Berry, D. T. R., Clark, J. A., Shipper, L. J., Graue, L. O., and Harp, J. P. (2010). Detection of malingered mental retardation. Psychol. Assess. 22, 50–56. doi: 10.1037/a0016585

Sharland M. J., Gfeller J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, *22*, 213–223.

Silk-Eglit, G. M., Lynch, J. K., & McCaffrey, R. J. (2016). Validation of Victoria Symptom Validity Test Cutoff Scores among Mild Traumatic Brain Injury Litigants Using a Known-Groups Design. *Archives of Clinical Neuropsychology*, *31*(3), 231–245. https://doi.org/10.1093/arclin/acv108

Slick, D. J. (1997). *The Victoria Symptom Validity Test*. Amsterdam University Press.

Soble, J. R., Santos, O. A., Bain, K. M., Kirton, J. W., Bailey, K. C., Critchfield, E. A., O'Rourke, J. J. F., Highsmith, J. M., & González, D. A. (2017). The Dot Counting Test adds up: Validation and response pattern analysis in a mixed clinical veteran sample. *Journal of Clinical and Experimental Neuropsychology*, *40*(4), 317–325. https://doi.org/10.1080/13803395.2017.1342773

Steffan, J. S., Morgan, R. D., Lee, J., & Sellbom, M. (2010). A Comparative Analysis of MMPI-2 Malingering Detection Models Among Inmates. *Assessment*, *17*(2), 185–196. https://doi.org/10.1177/1073191109359382

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (3rd ed.). Oxford University Press.

Stenclik, J. H., Miele, A. S., Silk-Eglit, G., Lynch, J. K., & McCaffrey, R. J. (2013). Can the Sensitivity and Specificity of the TOMM Be Increased with Differential Cutoff Scores? *Applied Neuropsychology: Adult, 20*(4), 243–248. https://doi.org/10.1080/09084282.2012.704603

Storm, J., & Graham, J. R. (2000). Detection of coached general malingering on the MMPI—2. *Psychological Assessment, 12*(2), 158–165. https://doi.org/10.1037/1040-3590.12.2.158

Sullivan, K., Lange, R. T., & Dawes, S. (2006). Methods of Detecting Malingering and Estimated Symptom Exaggeration Base Rates in Australia. *Journal of Forensic Neuropsychology, 4*(4), 49–70. https://doi.org/10.1300/J151v04n04_04

Tarescavage, A. M., & Glassmire, D. M. (2016). Differences between Structured Interview of Reported Symptoms (SIRS) and SIRS-2 sensitivity estimates among forensic inpatients: a criterion groups comparison. Law and Human Behavior, 40, 488–502. https://doi. org/10.1037/lhb0000191.

Taylor, L. A., Kreutzer, J. S., & West, D. D. (2003). Evaluation of malingering cut-off scores for the Rey 15-Item Test: a brain injury case study series. *Brain Injury, 17*(4), 295–308. https://doi.org/10.1080/0269905021000038401

Tolin, D. F., Steenkamp, M. M., Marx, B. P., & Litz, B. T. (2010). Detecting symptom exaggeration in combat veterans using the MMPI–2 symptom validity scales: A mixed group validation. *Psychological Assessment, 22*(4), 729–736. https://doi.org/10.1037/a0020973

Tombaugh, T. N. (1996). *Test of Memory Malingering*. MHS.

Toomey, J. A., Kucharski, L. T., & Duncan, S. (2008). The Utility of the MMPI-2 Malingering Discriminant Function Index in the Detection of Malingering. *Assessment, 16*(1), 115–121. https://doi.org/10.1177/1073191108319713

Veazey, C. H., Wagner, A. L., Hays, J. R., & Miller, H. A. (2005). Validity of the Miller Forensic Assessment of Symptoms Test in Psychiatric Inpatients. *Psychological Reports, 96*(3), 771–774. https://doi.org/10.2466/pr0.96.3.771-774

Vilar-Lòpez, R., Gomez-Rio, M., Caracuel-Romero, A., Llamas-Elvira, J., and Perez-Garcia, M. (2008). Use of specific malingering measures in a Spanish sample. *Journal of Clinical Neuropsychology*. 30, 710–722. doi: 10.1080/13803390701684562

Vilar-Lòpez, R., Santiago-Ramajo, S., Gomez-Rio, M., Verdejo-Garcia, A., Llamas, J. M., and Perez-Garcia, M. (2007). Detection of malingering in a Spanish population using three specific malingering tests. *Archives of Clinical Neuropsychology*. 22, 379–338. doi: 10.1016/j.acn.2007.01.012

Vitacco, M. J., Jackson, R. L., Rogers, R., Neumann, C. S., Miller, H. A., & Gabel, J. (2008). Detection Strategies for Malingering With the Miller Forensic Assessment of Symptoms Test. *Assessment*, *15*(1), 97–103. https://doi.org/10.1177/1073191107308085

Vrij, A. (2008). Detecting lies and deceit: Pitfalls and opportunities, 2nd edition: Discovery Service for Loma Linda University Library. Chichester: John Wiley.

Walczyk, J. J., Sewell, N., & DiBenedetto, M. B. (2018). A Review of Approaches to Detecting Malingering in Forensic Contexts and Promising Cognitive Load-Inducing Lie Detection Techniques. *Frontiers in Psychiatry*, *9*, 700. https://doi.org/10.3389/fpsyt.2018.00700

Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A Validation of the Test of Memory Malingering in a Forensic Psychiatric Setting. *Journal of Clinical and Experimental Neuropsychology*, *25*(7), 979–990. https://doi.org/10.1076/jcen.25.7.979.16481

Weiss, R. A., Rosenfeld, B., & Farkas, M. R. (2011). The Utility of the Structured Interview of Reported Symptoms in a Sample of Individuals With Intellectual Disabilities. *Assessment*, *18*(3), 284–290. https://doi.org/10.1177/1073191111408230

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology: Research and Practice*, *26*(5), 474–477. https://doi.org/10.1037/0735-7028.26.5.474

Willis, P. F., Farrer, T. J., & Bigler, E. D. (2011). Are Effort Measures Sensitive to Cognitive Impairment?. Retrieved January 11, 2017, from http://0-eds.a.ebscohost.com.catalog.llu.edu/eds/detail/detail?sid=bbdbc680-1975-4b4c-911e 6fc799e3cd8f%40sessionmgr4008&vid=0&hid=4203&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=69632302&db=aph

Zenisek, R. A., 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*(5), 426–433. https://doi.org/10.1093/arclin/acw025