An Electronic Validation of the Test of Memory Malingering (TOMM)

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Bonnie Maureen Smith

August 2023

Acknowledgements

Throughout my time in graduate school, I had the honor of working with exceptional faculty members who had a large impact on me, both personally and professionally. I wholeheartedly thank Dr. Merilee McCurdy, who has provided me with exceptional guidance and investment in my development as a clinician and professional. Thank you to Dr. Brian Wilhoit, who has given me numerous opportunities and unwavering support and encouragement. Thank you to Dr. Karee Dunn, who has given me teaching opportunities and has been a great cheerleader throughout my graduate program. Thank you to Dr. Sherry Mee Bell who has provided invaluable insight into the research process and has greatly assisted in my development as a researcher. An additional thank you goes to Dr. Steve McCallum for the years of guidance and investment into my development as a researcher. Further acknowledgement is due to my internship supervisor Dr. Holly Roberts for her support in data analysis and encouragement throughout this process. I would also like to thank my fellow School Psychology peers for their support and friendship over the years and I cannot wait to see what you accomplish next.

Lastly, and most importantly, I would like to thank my family for their unwavering support, encouragement, and love throughout my life. You never doubted in my ability to reach for the stars and gave me the strength to do so. Without you, this day would never have been possible, and I hope to make you proud as I continue on my journey. A final thanks goes to Dr. Jonathan Oravsky, who has offered support, motivation, and encouragement in countless ways throughout my journey.

Abstract

Within the context of a counterbalanced design, 46 students from a large state university in the southeast were administered the Test of Memory Malingering (TOMM) and the electronic version of the Test of Memory Malingering (TOMM-E). The TOMM was developed by Dr. Tom Tombaugh in 1996 and was originally designed to be administered in a paper and pencil format. The objective of this study was to establish equivalence between the computer-administered TOMM-E and the traditional paper-based TOMM. Each assessment consists of two trials and simple random assignment was used to determine which assessment format was administered to participants first. Time between assessment administrations varied but was approximately one week. Correlations between trials were high for both TOMM (.805) and TOMM-E (.755). Pairwise comparisons showed non-significant mean differences between all trials. ANOVA results indicated no significant mean differences between TOMM and TOMM-E trials ($F = 1.756$, $p = .159$, $\eta^2 = .038$). Levene's test revealed equal error variances for three trials: T1: $F(1, 44) = 2.728$, $p = .106$; T2: $F(1, 44) = .779$, $p = .382$), and E2 [($F(1, 44) = 2.002$, $p = .164$)], but unequal variances for E1 ($F(1, 44) = 10.849$, $p = .002$). Mauchly's test showed a significant effect of test format order on scores. Implications, limitations, and future research recommendations are discussed.

Table of Contents

List of Tables

CHAPTER I

Review of the Literature

Electronically based administration has become an increasingly popular way for clinicians and psychologists to administer assessments. Many practitioners have relied extensively on electronic assessments to reach clients during the pandemic. An electronic administration of an assessment offers many advantages, including cost-effectiveness, speed of administration, greater accessibility to diverse populations, reduced human error, and increased accuracy of results. However, equivalency between the traditional format of the TOMM and the electronic version of the TOMM (TOMM-E) must be established before the TOMM-E is used. Previously, one study has examined the equivalency between the TOMM and the TOMM-E but allowed examinees to self-administer the assessment, effectively introducing the additional potential for human error in administration. This study seeks to determine the equivalency of examiner-administered TOMM and TOMM-E in a sample of undergraduate university students.

Establishing Agreement

In today's world, electronically administered assessments hold many advantages over traditional paper-and-pencil methods. Electronic assessments reduce or eliminate human error, do not require testing booklets, allow for greater flexibility in scheduling, are accessible to a more diverse population (including those with disabilities), keep accurate time recording, allow for more secure storage of assessment data, may prevent the skipping of items by requiring a response before moving on, and can provide immediate feedback (Boo & Vispoel, 2012; Bracken et al., 2019; Gwaltney et al., 2008). Noyes and Garland (2008) argue that while complete equivalence between the two assessment forms is impossible, more sophisticated

computer software and more positive participant attitudes have resulted in continuing advances toward achieving total equivalence.

Most assessments that are in use today have been initially normed using a paper-and-pencil version. As a result, reliability and validity were established using a paper assessment format. It is crucial to establish equivalence between the forms to utilize results from an electronic version of an assessment to inform clinical decisions in the same way a paper-and-pencil administration would be used (Chuah et al., 2006).

Differences Between Administration Methods

While computerized assessments present items verbatim from the paper-and-pencil assessment, equivalence between the two forms must be explored and established before relying on computerized assessment to yield the same results. Gwaltney et al. (2008) suggest two primary reasons for inequivalence between these two forms: differences in how items or responses are presented and difficulties some examinees may have with computers. Some presentation differences may be minor, but some may involve significant alteration of item format, such as splitting a question between multiple screens due to space constraints. In addition, electronic items are often presented singularly, whereas items in a traditional paper assessment are often presented with multiple items at once. These presentation differences may alter response patterns as the examinee cannot return to previous items to compare or change their answers (Gwaltney et al., 2008). Studies examining the equivalency of paper to electronically administered assessments have been conducted since computers have become more widespread (Noyes & Garland, 2008). Early studies showed mixed results on the equivalency of electronic and paper administered assessments. For instance, Kak's (1981) study suggested a significant increase in reading speed when reading from a computer versus a paper

format of the same assessment. On the contrary, results from Heppner et al. (1985) found significant increases in reading performance scores on the paper administration of the Nelson-Denny Reading Test compared to a computerized version. A meta-analysis conducted by Ziefle in 1998 suggested that across a majority of studies, participant performance improved when presented with paper formats of assessments. Ziefle (1998) proposed that results may be due to increased eye strain when viewing electronic screens compared to paper.

Given the significant improvement in electronic screens over the last three decades and the increased familiarity of the general population with technology, we should expect to see the performance gap for electronically administered versions of assessments begin to shrink (Noyes & Garland, 2008). The use of electronic administration of assessments over traditional paper assessments has also been supported by user preference. Some studies suggest increased difficulty in manipulating cumbersome paper formats of assessments (Pinsoneault, 1996; Vispoel, 2000).

Determining the equivalence of paper and electronic formats can be conceptualized as test-retest or alternative form reliability. Gwaltney et al. (2008) suggest that this correlation should be at least 0.75 to establish excellent equivalence between forms. The advantage of utilizing the test-retest model to establish equivalency between forms lies in using the same subjects on both assessment forms. Theoretically, this should eliminate confounding variables due to heterogeneous individual characteristics (Hendrickson et al., 1993).

Effort

Effort can be defined as an individual's investment in performing at their maximum level (Bush et al., 2005). Effort is a crucial component of test performance and is connected to

cognitive abilities and behavioral engagement with the task at hand (Bigler, 2014). Adequate effort is such a crucial element in the assessment process that most cognitive, neuropsychological, and achievement assessments include specific phrasing to the examiner to direct the examinee to provide their best effort (Bigler, 2014).

Two separate terms have been proposed to further measure and describe effort on assessments: performance validity and symptom validity. These terms refer to two types of effort and potential deliberate exaggeration or falsification of symptoms. Performance validity refers to the accuracy of the examinee's performance on objective test measures (e.g., the TOMM). In contrast, symptom validity refers to the accuracy of self-reported symptoms compared with established symptomology of diagnoses (Larrabee, 2012). Although some variation exists between the two primary types of malingering (e.g., feigning cognitive impairments or exaggeration of psychological symptoms) and the profiles of individuals who might engage in one form of malingering over the other, the overarching principle remains the same. If an individual deliberately feigns symptoms for external gain, they will often fail in their performance on malingering assessments due to overrepresenting their challenges or symptoms (Walczyk et al., 2018).

Malingering

The concept of malingering in medical settings has been prevalent throughout recorded history, with the first record of malingering in English medical texts appearing in 1843. Shortly thereafter, a method involving ether was proposed to determine feigned illness from legitimate medical concerns (Lebourgeois, 2007). The term "malingering" was first used to describe soldiers during the 20th century who deliberately faked or exaggerated their symptoms to avoid or reduce their military service obligations (Walczyk et al., 2018). During World War II, British

soldiers produced literature for German soldiers with instructions on various methods to malinger illness in an attempt to gain military leave (Lebourgeois, 2007).

Currently, malingering within a medical or psychological context is defined as the deliberate falsifying or exaggerating of behavior or symptoms to achieve a desired outcome (Walczyk et al., 2018). This desired outcome may take many forms, including a prescription for medication, insurance benefits, personal liability payouts, or disability status. To differentiate malingering from other psychological disorders, the promise of secondary gain is necessary (Alozai & McPherson, 2022). Malingering is not currently included in the DSM-5 as a disorder. Although it shares some symptoms with somatic symptom disorder [DSM-5 300.82 (F45.1)], the primary difference is that malingering is the deliberate falsification of symptoms (American Psychiatric Association, 2013). In contrast, individuals with somatic symptom disorder experience the symptoms as their reality. Malingering is included in the DSM-5 as a V-Code, or "other conditions that may be a focus of clinical attention." The DSM-5 proposes that the presence of malingering should be evaluated when a patient presents with any combination of the four conditions: "1) The medicolegal context of the presentation, for example, a lawyer sending his client for evaluation or patient presents with an illness while facing trial, 2) Marked discrepancy between the individual's 'claimed stress or disability' and 'objective finding and observation,' 3) Lack of compliance with diagnostic evaluation, treatment regimen and follow up care, 4) Presence of antisocial personality disorder" (Ross, 2019). Malingering can be divided into three primary categories: exaggeration of symptoms, intentionally poor performance on neuropsychological or cognitive testing, and both exaggeration of symptoms and intentionally poor performance (Iverson & Binder, 2000). Still, other models propose that the concept of malingering should be further differentiated into four specific subtypes: invention (or the

claiming of symptoms that the person has not experienced), perseveration (or the continued presentation of symptoms when symptoms have subsided), exaggeration (or the deliberate over-presentation of symptoms), and transference (or when symptoms from one condition are knowingly misrepresented as stemming from another condition) (Huskey, 2002; Lipman, 1962).

Malingering can present a unique challenge to psychologists who often assume their client is being straightforward in their symptom presentation and need for assistance. Furthermore, some clinicians may be wary of confronting clients with questions related to malingering due to the possibility of client behavior escalation and hostility (Lebourgeois, 2007). Curtis & Hart (2015) determined that many counselors and psychologists hold negative views about clients who lie or misrepresent their symptoms. Clinicians reported decreased liking of their clients, decreased trust within the therapeutic relationship, and decreased desire to continue working with the client. As a result, these reactions may affect a clinician's case conceptualization, rapport, and treatment outcomes (Dickens & Curtis, 2019). Despite the potential challenges of working with a client who engages in malingering, it is a common phenomenon, particularly in cases where clients have an option for substantial gain. One meta-analysis discovered probable malingering in 30% of disability evaluations, 29% of personal injury evaluations, 19% of criminal evaluations, and 8% of medical evaluations (Mittenburg et al., 2002).

Assessment of Malingering

Like the clinical assessment of most presenting psychological concerns, an intake interview is the first step in assessing the potential for malingering. Lebourgeois (2007) suggests that a clinician should be mindful of posing open-ended questions to a client suspected of malingering to avoid inadvertently giving the client additional information about the symptoms

of a disorder. After the client has answered the open-ended questions concerning their symptoms, the clinician can more easily assess the presence of typical or atypical patterns of symptoms. Other client behavior patterns may also lead a clinician to suspect malingering. Cercey, Schretlen, & Brandt (1997) note that malingering may be present when an individual can function normally within the confines of work or school but demonstrates significant impairment on psychological assessments. Furthermore, clients presenting with malingering may display frustration or act uncooperative and evasive during clinic interviews, portray symptoms as beginning suddenly and with great intensity, and present inconsistent reports from family and friends (Iverson, 1995; Ruff, Wylie, & Tennant, 1993).

In cases where the clinician may suspect malingering, an assessment is often conducted to measure either the extent or likelihood of reported symptoms concerning a genuine mental health or medical condition. The American Academy of Clinical Neuropsychology has recommended routinely incorporating malingering assessments into most psychological evaluations, with a few noted population exceptions. Notably, examinees with dementia may perform poorly on malingering assessments due to legitimate memory impairments rather than attempted malingering (Heilbronner et al., 2009). In the broader field of psychology, some clinicians incorporate at least one aspect of malingering assessment into all cases. In contrast, others only utilize these measures if malingering is suspected. Drawbacks and benefits to both methods are present. Assessing for malingering when no clinician concern is present may frustrate the client, lengthen the time of assessment, and increase assessment cost while assessing for malingering only in suspected cases may allow for some sophisticated malingers to avoid detection (Gudmundsson et al., 2021). These assessments can either be embedded within another

measure or a standalone instrument. Both types are discussed in greater detail in the following sections.

There is no "gold standard" of malingering assessment, but rather several different measures that vary in many factors, including time of administration, format (self-report versus clinician-administered), and theoretical approach (Farkas et al., 2006). Assessments designed to measure malingering accurately must evaluate both sensitivity and specificity. Sensitivity refers to an assessment's ability to correctly identify individuals to whom the condition applies, while specificity refers to an assessment's ability to correctly identify individuals to whom the condition does not apply (Swift et al., 2020). In terms of assessing for performance malingering, sensitivity refers to the degree to which an assessment can measure the presence of cognitive challenges (whether genuine or malingered). In contrast, specificity refers to the ability of an assessment to identify the deliberate feigning of poor effort or performance. Sensitivity in assessing for symptom malingering refers to the degree to which an assessment can identify the presence of psychological symptoms, while specificity refers to the ability of an assessment to identify the likelihood that reported symptoms are experienced as the presentation of a psychological disorder. In general, sensitivity and specificity rates above .80 are considered strong for diagnostic tests in medicine and psychology (Mackin et al., 2010).

Performance Validity Tests

Performance validity tests, previously known as effort tests, are often used by psychologists to determine a purposeful lack of effort or feigned memory impairment (McWhirter et al., 2020). Most performance validity assessments involve a forced choice in which the examinee must select a previously displayed stimulus. Based solely on probability, an examinee should answer at least half of the items correctly. If the examinee correctly identifies

less than half of the stimuli, it may be assumed they are deliberately selecting the incorrect stimuli to feign their challenges with memory or other cognitive processes (McWhirter et al., 2020). Other test formats rely on the "floor effect," a predetermined cut score that is unlikely for an individual applying full cognitive effort to achieve below (McWhirter et al., 2020). In addition to the TOMM, several other empirically validated assessments designed to measure cognitive performance and effort are widely used today.

Test of Memory Malingering (TOMM). The TOMM is one of neuropsychologists' and clinical psychologists' most commonly used performance validity assessments (Grant & Werner, 2020). The TOMM is normed for use with ages 16– 84 and consists of two trials and an optional retention trial. The TOMM is composed of two parts: the first is a learning task, and the second is a recognition task. The learning task requires the examinee to memorize 50 simple, line-drawn pictures of everyday items presented for three seconds each. After the presentation of all 50 items, the examinee is presented with a recognition task using a forced choice model and instructed to select the item previously shown (Gudmundsson et al., 2021). A score of fewer than 45 correct items on either trial suggests the possibility of malingering and indicates the need to administer an additional retention trial. This trial is administered 15 minutes after the second trial and consists of only a forced choice paradigm, with no additional stimuli presented to the examinee (Gudmundsson et al., 2021). To avoid interference, no memory tasks should be administered between the second presentation and the retention trials (Tombaugh, 1996). Normative samples suggest that most examinees to whom the TOMM is administered make very few errors, including those with traumatic brain injuries (TBI) or epilepsy (Grant & Werner, 2020; Tombaugh, 1996). Individuals with lower cognitive abilities have also demonstrated

success on the TOMM, with the standard TOMM cutoff scores applicable to examinees with an estimated IQ of 80 or higher (Grant & Werner, 2020).

The TOMM is traditionally administered in a clinical setting with the examiner sitting across from the examinee. The examiner introduces the TOMM by stating that this test determines how well the examinee can learn and remember pictures of everyday objects. While the administration is straightforward and relatively uncomplicated, human error can result in the stimulus pictures being presented longer than prescribed. Current technology may mitigate this human error by implementing slideshow technology (e.g., Powerpoint or Google Slides) to automatically advance to the following picture at predetermined intervals. After presenting the images, the examiner can regain control of the slideshow to advance each picture during the assessment portion as needed.

Use of the TOMM with Individuals of Lower Cognitive Abilities. One of the most frequently cited challenges with the use of the TOMM is its specificity in individuals with lower cognitive abilities. Previously conducted studies examining this issue have suggested mixed results. Graue et al. (2007) conducted a study evaluating the use of the TOMM with individuals with mild intellectual disability (ID). Specificity rates were 0.69 on the first two trials of the TOMM and 0.89 on the retention trial, using the standard cutoff of 45 correct items. Similarly, Hurley & Deal (2006) reported specificity rates of 0.55 in their study with individuals possessing an IQ in the 50-70 range. The mean score on the first two trials with this population was 43.6, below the standard cutoff of 45 correct items. However, Grant & Werner (2020) note that three other performance validity tests were administered within the same time frame, elevating the possibility of flagging effort or fatigue, leading to artificially lower scores. Conversely, Simon

(2007) found high specificity rates (0.95) and high mean scores (49.4 on the first two trials and 48.7 on the retention trial) among individuals with mild ID in a prison setting.

Use of the TOMM with Individuals with Dementia. Another group that is given increased attention when examining the specificity of the TOMM is older individuals and those with dementia. Using the traditional cutoff score of 45 correct items, Teichner & Wagner (2004) determined a specificity rate of 0.76 with older individuals. However, when ruling out dementia through comprehensive neuropsychological measures, the specificity rate for older individuals increased to 0.96. Similarly, a study by Walter et al. (2014) to assess the use of the TOMM in a group of dementia patients found that approximately 20% of participants with moderate-severe levels of dementia failed the TOMM using the traditional cutoff scores. These results suggest that the TOMM may lead to false positives when examining individuals with dementia for malingering.

Use of the TOMM with Chronic Pain Conditions. Individuals with chronic pain are a population that deserves additional consideration in the assessment of malingering. Etherton et al. (2005) examined the impact of laboratory-induced pain (via the continuous application of ice) on TOMM performance. The study involved 39 participants without chronic pain who completed the TOMM in two conditions: with and without laboratory-induced pain. The study's results demonstrated that the performance on the TOMM was not affected by the laboratory-induced pain. This suggests that the TOMM remains a useful tool for assessing effort and malingering in clinical settings, even when patients are experiencing pain. Similarly, a study conducted by Meyers & Diep (2000) demonstrated that the TOMM was the most sensitive measure of malingering for patients with chronic pain conditions that were administered a battery of validity

assessments, including the Victoria Symptom Validity Test (VSVT) and the Dot Counting Test (DCT).

TOMM Psychometric Properties. Several studies have investigated the test-retest reliability of the TOMM. For example, Bianchini et al. (2005) found that the TOMM had high test-retest reliability, with a correlation coefficient of .86 between two test administrations. Similarly, Larrabee et al. (2007) found high test-retest reliability, with a correlation coefficient of .87. Several studies have investigated the inter-rater reliability of the TOMM. Larrabee et al. (2007) found that the TOMM had high inter-rater reliability, with an intraclass correlation coefficient (ICC) of .99. A study conducted by Tombaugh and colleagues (1998) also reported high inter-rater reliability, with the ICC ranging from .93 to .99.

The TOMM has demonstrated strong validity across multiple measures. The TOMM has strong content validity because it includes both a learning task and a recognition task designed to assess different aspects of memory functioning (Tombaugh, 1996). Criterion validity, or the extent to which a test correlates with an established measure of the construct it is intended to measure, of the TOMM has been established in multiple studies. Slick et al. (1996) found that the TOMM had a high correlation with other measures of feigned cognitive impairment, such as the Word Memory Test (WMT) and the Victoria Symptom Validity Test (VSVT). Likewise, a study by Gervais et al. (2004) found that the TOMM highly correlated with the WMT (in assessing for malingering) and the Test of Memory and Learning (in assessing for general and specific memory functioning). In addition, the TOMM has been found to correlate negatively with measures of symptom validity and positively with measures of cognitive ability (Green et al., 1999).

The TOMM has been found to have high sensitivity and specificity in detecting malingering or exaggeration of memory impairment. In 1996, Tombaugh administered the TOMM to 63 individuals with a history of litigation or compensation claims related to cognitive impairment. The test had a sensitivity of .92 and a specificity of .90 in detecting malingering or exaggeration of cognitive symptoms. Similarly, Green et al. (1999) conducted a study with individuals seeking disability benefits due to cognitive impairment. In this study, the TOMM was reported to have a sensitivity of 1.00 and a specificity of .91. Sensitivity and specificity with the TOMM were also demonstrated to be strong among groups with various health conditions. Slick et al. (1996) administered the TOMM to patients with mild TBI and a control group; results were reported as a sensitivity of .83 and a specificity of .95. In a study by Galioto et al. (2020), the TOMM was administered to 135 patients with multiple sclerosis and 42 controls. Results indicated that the TOMM had a higher sensitivity and specificity (of .83 and 1.00, respectively) than the Victoria Symptom Validity Test for patients with multiple sclerosis.

TOMM Cultural Sensitivity. In the initial development and validation of the TOMM, Tombaugh (1996) noted the importance of cultural sensitivity and adaptation of the test to different populations. He recommended that the TOMM be used in conjunction with other assessment tools to evaluate cognitive functioning in diverse populations. Several studies have examined cultural sensitivity since the creation of the TOMM. In a study by Bridges & Ahern (2016), the TOMM was administered to a sample of individuals from diverse cultural backgrounds, including African American, Hispanic, and White examinees. The authors found that the TOMM had a sensitivity of .86 and a specificity of .91 for African American individuals, a sensitivity of .80 and a specificity of .80 for Hispanic individuals, and a sensitivity of .90 and a specificity of .78 for White individuals. Similarly, a study conducted with 60 participants in

Romania demonstrated similar performance to the normative group of the TOMM (Crișan & Erdodi, & 2022). This data suggests that the TOMM can be used confidently among Romanian populations. Nijdam-Jones et al. (2019) administered the TOMM to illiterate and literate individuals in Columbia. Results indicated that although the mean of both groups was above the cutoff score for malingering (illiterate group mean: 45.5; literate group mean: 49.3), the illiterate group scored significantly lower than the literate group, indicating potential concerns when using the TOMM with educationally diverse groups.

Rey Fifteen-Item Test (FIT). The FIT is one of the most popular tests of memory malingering, despite some objections to its clinical use (Reznek, 2005). The test has been normed for use with adults 18 and older, with some studies examining the use in children as young as eight. Test administration is very brief (completed in less than five minutes), test materials are limited (one stimulus card containing fifteen items organized in columns and rows), and scoring is straightforward and simple (requiring only the total of correct items) (Boone et al., 2002). After the presentation of the items and a brief delay, examinees are asked to recreate as many of the test items as possible from memory. Although the task appears challenging, items are grouped in themes to aid in memory. Like the TOMM, this assessment relies on the floor effect strategy (Frederick, 2000). Malingerers will often perform poorly on this task to feign cognitive challenges. Despite this assessment's popularity, questions have been raised since the test's inception. After conducting a meta-analysis of available studies, Reznek (2005) concluded that, apart from individuals with intellectual delays, the FIT proved to be a measure with low sensitivity and high specificity. Additionally, the FIT has demonstrated high test-retest reliability, with correlation coefficients ranging from .82 to .94 (Schoenberg et al., 2003; Tombaugh & McIntyre, 1992).

Word Memory Test (WMT). The WMT can either be administered electronically or through a paper-and-pencil method. The electronic version of the WMT generates a detailed printout of performance, including comparisons of individual performance to various normative groups (Bhowmick et al., 2019). The WMT assesses verbal memory, utilizing multiple components, including immediate and delayed recall and a paired recall component (Tracy, 2014). The WMT has a large normative sample, including individuals with traumatic brain injuries and neurological disease (Hartman, 2002). Additionally, extensive research has been conducted involving the WMT with various populations, including individuals with chronic pain, fibromyalgia, and children (Hartman, 2002). McWhirter et al. (2020) suggest that examinees with probable, mild, and moderate levels of dementia are likely to demonstrate poor performance on the WMT, and caution is recommended when working with this population. Despite this, most studies indicate that the WMT has high sensitivity and specificity with most examinee populations (Bhowmick et al., 2019). Internal consistency for the WMT has also been found to be high, with Cronbach's alpha coefficients ranging from .87 to .96. Test-retest reliability has also been found to be high, with intraclass correlation coefficients ranging from .73 to .91. Additionally, the WMT has been found to have high sensitivity and specificity in detecting malingered memory impairment, with sensitivity ranging from .81 to .96 and specificity ranging from .90 to .98 (Bhowmick et al., 2019).

Portland Digit Recognition Test (PDRT). Normed in 1993, the PDRT has largely fallen out of favor due to the significant time administration requirements. The assessment consists of 72 trials, broken into four sets of 18 (Tracy, 2014). The gaps between stimulus presentation and the forced choice item vary from five to 30 seconds, depending on the trial. Trials increase in difficulty, requiring the examinee to remember the string of numbers for successively longer

periods (Tracy, 2014). Allen (2011) notes that successive increases in time with each trial serve to increase the appearance of difficulty for examinees. Studies suggest that the PDRT has a high specificity and lower sensitivity, especially with compensation-seeking examinees (Bianchini et al., 2001). Internal consistency for the PDRT is somewhat strong, with Cronbach's alpha coefficients ranging from .85 to .97. Test-retest reliability has also been found to be high, with ICC ranging from .87 to .95 (Rose et al., 1995).

Symptom Validity Tests

Symptom validity is measured based on self-reported levels of symptoms. The assessment results are compared to the statistical probability of the symptoms being reported in a typical clinical presentation of a specific diagnosis. Most frequently, examinees tend to overreport their symptoms to malinger for external gain. Less commonly, examinees may underreport their symptoms, particularly on neuropsychological measures designed to determine living placements (e.g., a patient with dementia who desires to continue living independently) (Bush et al., 2005).

Minnesota Multiphasic Personality Inventory-2 (MMPI-2). The MMPI-2 includes several symptom validity scales (F-scales) embedded within the assessment that measure the likelihood of overstatement of psychosomatic symptoms (Greiffenstein, 2010). Items on these scales assess physical functioning, concerns with health, neurological problems, and cognitive problems related to attention, effort, and motivation (Wygant et al., 2007). Greiffenstein et al. (1995) suggest that while the MMPI-2 symptom validity scales can effectively measure psychological symptom reporting, they often fail as an accurate measure of cognitive effort and motivation. Larrabee (2003) reported that examinees seeking personal injury compensation who

failed a forced-choice assessment designed to identify malingering reported higher symptom levels on the MMPI-2 symptom validity scales than patients with traumatic brain injuries, depression, multiple sclerosis, spinal cord injuries, and chronic pain conditions. This study suggests the MMPI-2 symptom validity scales to be capable of accurately differentiating between exaggerated psychosomatic symptoms and accurate physical symptom reporting, characteristic of a chronic health condition. Specifically, malingers had elevated scores on scales measuring: Hypochondriasis, Hysteria, Paranoia, Psychasthenia (referring to an outdated psychological term describing patients with phobias, obsessions, compulsions, or excessive anxiety), Schizophrenia, and Social Introversion (Heaton et al., 1978). In contrast, patients with legitimate neurological conditions typically demonstrate elevated scores in Depression, Psychasthenia, and Schizophrenia (Cullum & Bigler, 1987). Despite this, Butcher et al. (2008) report that symptom validity scales on the MMPI-2 report many false positive profiles for malingering, particularly among psychiatric patients. These patients are more likely to be experiencing genuine psychological symptoms but may be reported as exaggerating them. As a result, some clinicians recommend that the symptom validity scales on this assessment not be used with a psychiatric or clinical population but still recommend for use among examinees seeking personal injury claims (Butcher et al., 2008).

Behavior Assessment System for Children, Third Edition (BASC-3). The BASC-3 is a commonly utilized rating scale designed to measure a wide range of behavioral symptoms, emotional symptoms, and adaptive skills among children aged 2 – 21 years 11 months. The BASC-3 includes three formats: a self-report, a teacher scale, and a parent scale. All three versions include five validity scales to assist clinicians with interpreting the scores (Kirk et al., 2014). The F-Index is of particular interest to the current study, which comprises items rarely

endorsed (less than 3% of the normative sample) by children, parents, or teachers. This index is sometimes referred to as a "faking bad" index, as elevated scores in this area should lead a clinician to consider the possibility of an overrepresentation of negative symptomology (Reynolds & Kamphaus, 2004). Kirk et al. (2014) concluded that children as young as eight could exaggerate their negative psychological symptoms, sometimes without the apparent promise of secondary gain. While the underlying reasons for an elevated F-Index may be as innocuous as parental frustration with child symptomology, further evaluation is warranted when the F-Index is elevated.

Victoria Symptom Validity Test (VSVT). The VSVT is a 48-item test divided into three blocks of 16 items; each normed for ages 18 – 72 years. Each block has eight difficult items and eight easy items (Slick et al., 1996). Slick et al. (1996) proposed adopting a three-tiered cut score system for interpreting scores. Using controls, experimental malingerers, and compensation-seeking patients, a classification of "malingered," "questionable," and "valid" were implemented to evaluate participant performance on the VSVT. This classification is applied to the number of easy items answered correctly, the number of difficult items answered correctly, and the total number of correct items. Utilizing this three-tiered system significantly increased sensitivity and showed little impact on specificity (Slick et al., 1996).

Structured Inventory of Malingered Symptomatology (SIMS). The SIMS is a self-report scale for use with clients from 18 years – 99 years containing 75 true/false questions relating to atypical and bizarre symptomology (Smith & Burger, 1997). In addition to an overall score, the assessment has been designed to detect malingering on five scales: low intelligence, affective disorders, neurological disorders, psychotic disorders, and amnestic disorders (Smith & Burger, 1997). The manual has suggested a cut score of 14 endorsed items as an indication of

possible malingering, but many studies have supported a cut score of 16 endorsed items (Gudmundsson et al., 2021). Smith & Burger (1997) suggest the overall SIMS score is the most accurate picture of malingering, while the five scales measuring various potential conditions have been supported for use as qualitative data.

Statement of the Problem and Rationale for the Study

The Test of Memory Malingering is one of the most frequently used measures to assess levels of malingering (Tombaugh, 1996). However, this test was normed to be delivered in an in-person clinical setting. Given the current pandemic, as a means for increasing efficiency, decreasing the potential for human error, and decreasing barriers to access, electronic administration of the TOMM would be beneficial in assessing levels of malingering remotely. However, before administering this malingering assessment electronically, it is crucial to determine the extent to which results from the TOMM and the TOMM-E can be considered equivalent. The American Psychological Association (APA) has recommended that equivalence between paper and electronic test versions be evaluated empirically and not assumed. Previous studies have examined the equivalency of the TOMM and the TOMM-E based on an examinee's self-administration of the TOMM (Vanderslice-Barr et al., 2011). Not only is examinee administered assessment not standardized, but it also introduces additional human error in the administration of the TOMM. Accordingly, this study seeks to determine the equivalence between examiner-administered TOMM and TOMM-E. The purpose of this study was to compare performance on the traditional paper version of the TOMM with the TOMM-E and to evaluate the equivalence of the two formats of this assessment.

Research Questions

The following research questions were addressed by this study:

1. Are there differences between the traditional administration of the TOMM means and the TOMM-E means?
    a. It is hypothesized that there will not be significant statistical differences between the traditional administration of the TOMM and the TOMM-E.
2. Are the distributions of scores for the TOMM and the TOMM-E equivalent?
    a. Again, it is hypothesized that there will not significant statistical differences between the TOMM and the TOMM-E, as this is a study of equivalence across the two presentation formats of the assessment.
3. Does the order effect of test format presentation (traditional compared to electronic) result in statistically significant differences in scores?
    a. As simple random assignment was used to determine which test presentation would be administered first, it is hypothesized that the order effect will not be statistically significant in obtained scores.

CHAPTER II

Methods

Participants and Setting

Participants were undergraduate students enrolled in a teacher education program at a large land-grant university in the southeastern United States. Demographic information was collected and included gender, age, and status of current ADHD diagnosis. This data is presented in Table 1. The sample was 84.7% female (n = 39) and 15.2% male (n = 7). Ages of participants ranged from 18 to 29 with 63.8% being 18-21 (n = 30), 23.4% being 22-25 (n = 11), and 10.6% being 26-29 (n=5). 19.6% of participants reported a current diagnosis of ADHD (n=9) while 80.4% (n=37) of participants did not report a current diagnosis of ADHD.

Prior to the implementation of this study, IRB permission was obtained. Preservice teachers were given a brief orientation to the research and were offered extra credit in their course to participate in the study. Alternatively, the preservice teachers were also given the option to complete a writing assignment for an equivalent amount of course credit. Preservice teachers were provided a link to complete informed consent and a demographic form. Additionally, they were provided contact information for the researchers to discuss any questions or concerns. After completing the informed consent and the demographic form, the lead researcher shared a Google Document with various options to schedule two appointments for assessment administration. Each assessment appointment was scheduled approximately one week apart. A random number generator was utilized for simple random assignment of participants to complete either the TOMM or the TOMM-E at their first appointment. The other format was administered at the second assessment appointment. During the assessment, scores

were recorded on physical protocols of the TOMM. Protocols from these assessment administrations are stored within a secured file cabinet.

Instruments

Test of Memory Malingering (TOMM)

The TOMM consists of two learning trials: a learning phase (during which pictures are presented for 3 seconds with a 1-second interval between pictures) and a test phase (Tombaugh, 1996). The same pictures are used across both trials but are presented in a different order on each trial. During the test phase, each target picture is paired with a novel drawing (distractor) and the examinee must select the picture to which they were previously exposed. The position of the target picture is counterbalanced for top and bottom positions on the page (Tombaugh, 1996). In this study, the TOMM was administered using the Zoom meeting platform and a document camera. The examiner provided a link to a Zoom meeting. Once the examinee was admitted into the Zoom meeting, the examiner's screen was shared with the examinee, which provided a view of the document camera projecting the TOMM. In both trials, the examiner presented each item to the examinee for three seconds. After presentation of the stimuli, the forced choice portion of the trial was presented to the examinee. Examinee voiced their selection between the two items, and the examiner marked their selection.

Test of Memory Malingering – Electronic (TOMM-E)

The TOMM-E was created via scans of the original stimuli on the TOMM. Full examiner instructions for administering the TOMM-E can be found in Appendix C. As with the administration of the TOMM, examiners shared a Zoom link with examinees. Examinees joined the Zoom meeting, and the examiner shared their screen with the TOMM-E PowerPoint.

Transitions were embedded within the PowerPoint to automatically advance stimuli after presenting for the three-second interval, with a one-second interval break between items. After presentation of the stimuli at each trial, examinees were presented with the forced choice portion of the assessment. Examinees verbalized their selections, and the examiner recorded their responses.

Procedures

Treatment Integrity

Treatment integrity refers to the extent to which an intervention or assessment is implemented as intended. In the case of the TOMM, treatment integrity is essential to ensure that the test is administered consistently and accurately. The TOMM manual provides detailed instructions for administering the test, including specific steps for presenting the stimuli and responding to examinee answers. Specifically, the manual states that the examiner should provide a brief response to the examinee when they select the correct choice (such as “that is correct” or “that is right”) without providing any additional feedback or reinforcement. Similarly, the manual advises examiners to avoid providing extensive feedback or reinforcement for incorrect responses, simply stating, “that is incorrect” or “that is wrong.” The TOMM manual notes that providing extensive feedback or reinforcement for responses could alter the test results by modifying the examinee’s effort. The manual also provides guidelines for monitoring the participant’s effort and response patterns during the test and assessing the consistency of their responses. In particular, the manual suggests that examiners monitor for signs of inconsistent effort, such as a lack of effort or engagement, inconsistent response patterns, or abnormal response latencies. The manual also provides guidelines for identifying potential sources of invalid or unreliable test results, such as technical errors, cognitive impairment, or psychological

factors that may impact test performance. In addition, the manual recommends that examiners monitor for the use of strategies that may impact test results. For example, the manual suggests that examiners note if participants appear to be guessing, intentionally slowing down their response time, or trying to appear more impaired. Finally, the TOMM manual states that it is essential to document any deviations from the standard procedures of the TOMM to ensure transparency and accountability.

Data Entry and Sampling

Upon completion of each administration of the TOMM, each participant's scores were entered into the database, which was housed in a password-protected Google Drive spreadsheet. After 10 data points were entered into the database, a sampling method was used to ensure the data was correctly entered. Three randomly selected data points were compared to the completed physical protocol to identify errors or discrepancies. Throughout this process, no errors or discrepancies were discovered; thus, the database data is considered valid and reliable.

Data Analyses

Initially, tables showing descriptive statistics will be created, including means, standard deviations, skewness, and kurtosis for all variables. To answer Research Question 1(Are there differences between the traditional administration of the TOMM means and the TOMM-E means?), a paired sample t-test will be conducted initially to determine the correlation between the two trials of each test format. Then, an ANOVA (Analysis of Variance) will be conducted to determine if there are statistically significant adjusted mean differences between the experimental (TOMM-E) and the control group (TOMM) means. An ANOVA was selected over conducting multiple independent t-tests for this dataset to increase the statistical power and, thus,

the likelihood of detecting a true difference between groups. In addition, an ANOVA can also be used with covariates, such as age, gender, and ADHD diagnosis, to control for individual differences. After the ANOVA has been run, Tukey's HSD (Honestly Significant Difference) post-hoc will be used to determine which group means are significantly different. Tukey's HSD test is a helpful tool for post-hoc analysis because it provides a way to control the overall Type I error rate while allowing for pairwise comparisons between group means.

Levene's Test of Equality of Error Variances will be utilized to answer Research Question 2, which explores the equivalence of score distributions between the TOMM and the TOMM-E. This test allows for the determination of statistically equivalent variability of scores. By examining the equality of error variances, it can be determined if there are significant differences in the spread of scores between the TOMM and the TOMM-E. In addition to performing Levene's test, the influence of outliers on the validity of the test results will be examined.

In order to address Research Question 3, which investigates the potential order effect of test format presentation (traditional compared to electronic) on scores, Mauchly's Test of Sphericity will be employed. This test is designed to examine whether the assumption of sphericity, which assumes that the variances of the differences between all pairs of conditions are equal, has been violated. If violations of sphericity are detected, it indicates that the order effect of test formats may have resulted in statistically significant differences in scores. Results will provide insights into the presence or absence of sphericity violations. If a violation of sphericity is detected, the Greenhouse-Geisser correction will be applied. This correction adjusts the degrees of freedom in subsequent statistical analyses, taking into account the observed violation.

By adjusting the degrees of freedom, the Greenhouse-Geisser correction helps to mitigate the potential impact of unequal variances and ensures that the statistical tests are conducted accurately.

CHAPTER III

Results

In this section, data cleaning procedures and descriptive statistics are presented first. Then data analyses addressing Research Questions one through three follow.

Data Cleaning Procedures

Prior to conducting the analyses, several data cleaning processes were implemented to ensure the integrity and quality of the dataset. The initial dataset consisted of responses from 56 participants at the completion of the data collection phase. However, one participant was excluded from the analysis immediately as they did not fall within the predetermined age range specified for this study, which was set to include individuals between the ages of 18 and 29. Six participants who did not return for the second administration of the assessments were also excluded from the analysis. This exclusion was necessary to maintain consistency and continuity in the measurements obtained from the participants. Additionally, three participants who encountered technical issues, specifically glitches in their Wi-Fi connection during the assessment sessions, were removed from the analysis. This step aimed to minimize any potential disruptions or inconsistencies caused by unreliable internet connections, thereby preserving the quality and reliability of the collected data.

After implementing these data cleaning procedures, the analysis's final participant pool comprised 46 individuals. Removing ineligible participants and those who did not complete the entire assessment protocol contributed to a more focused and representative dataset, enhancing the robustness and validity of the subsequent statistical analyses.

Descriptive Statistics

Following the removal of incomplete administrations of the TOMM by respondents, the adjusted totals and means for the TOMM and TOMM-E assessments were obtained from the final pool of participants (N =46). Means, standard deviations, skewness, and kurtosis are presented below in Table 2. Normality of data was evaluated based on Kline's (2005) recommendation that data is considered to be normal if skewness is between -3 to +3 and kurtosis is between -10 to +10.

TOMM scores ranged from 45 to 50 (SD= 1.36; 1.30) with a mean of 49.17 and a mode of 50 across both administrations. TOMM-E scores ranged from 43 to 50 (SD= 1.58; 1.10) with a mean of 49.08 and a mode of 50 across both administrations. Scores on the TOMM are approximately normally distributed, with skewness ranging from -1.18 to -1.84 and kurtosis ranging from 3.40 to 6.26. These measures suggest that the data is negatively skewed. Negative skewness can suggest that the data may be more concentrated around a higher value, with fewer values towards the lower end. The mode of both administrations of the TOMM was 50, suggesting that data was clustered around the maximum score for the TOMM and thus resulting in negative skewness.

Similarly, scores on the TOMM-E were approximately normally distributed, with skewness ranging from -1.65 to -1.89 and kurtosis ranging from 6.18 to 7.27. Again, negative skewness in this dataset suggests the data is concentrated around a higher value. As the mode for the TOMM-E is also 50, this suggests that many participants achieved the maximum score on the TOMM-E. Results from data analyses pertaining to each research question are presented below.

Research Question 1: Evaluating Statistically Significant Differences Between the TOMM and the TOMM-E

Research Question 1 addressed the following question: Are there statistically significant differences between the means of scores from the traditional administration of the TOMM and the TOMM-E formats? A repeated measures ANOVA was run to determine if the scores on the traditional version of the TOMM (T1 and T2) differed from scores on the electronic version of the TOMM-E (E1 and E2).

Initially, a paired sample t-test was conducted. The test results indicated strong positive correlations between the scores obtained by each participant on the two trials of the TOMM and the TOMM-E (Table 3). Specifically, a high correlation coefficient of .805 was observed between the scores of the first and second trials on the traditional assessment version. Similarly, a correlation coefficient of .755 was found between the scores of the first and second trials on the electronic version of the assessment. These correlation coefficients suggest a strong relationship between the scores obtained in the two trials, indicating consistency in performance across the trials for each version of the assessment.

A repeated measures ANOVA was conducted to analyze the effect of test format on participant scores. The results revealed a non-significant main effect. The F-value of 1.756 with an associated p-value of .159 indicates no statistically significant difference in the means between the trials of the TOMM or the TOMM-E. The effect size, represented by the partial eta squared (η^2) of .038, suggests a small magnitude of effect. The lack of a significant main effect indicates that performance outcomes across the assessments' four total trials (T1, T2, E1, E2) were comparable. The analysis also accounted for the order in which the tests were administered as a between-subjects factor. This factor will be explored in Research Question 3. These results

suggest that the test format did not have a substantial impact on the scores obtained from participants. See Table 4.

Pairwise comparisons were conducted to examine the mean differences between each trial. The results (Table 5) revealed a non-significant mean difference of .004 (SE = .227, p = 1.000, 95% CI [-.625, .632]) for T1 and E1; a non-significant mean difference of -.325 (SE = .159, p = .280, 95% CI [-.763, .114]) for T1 and E2; a non-significant mean difference of .230 (SE = .232, p = 1.000, 95% CI [-.412, .871]) for T2 and E1; and a non-significant mean difference of -.099 (SE = .153, p = 1.000, 95% CI [-.521, .323]) for T2 and E2. These findings indicate no statistically significant difference between any trials (T1, T2, E1, E2).

Research Question 2: Equivalence in Distribution Scores for the TOMM and the TOMM-E

Research Question 2 addressed the following question: Are the distributions of scores for the TOMM and the TOMM-E equivalent? Levene's Test of Equality of Error Variances was examined to answer this question. If Levene's test result is statistically significant, this suggests that the data do not show homogeneity of variance. If Levene's test result is not significant, it can be assumed that the data show homogeneity of variance.

The results of Levene's test indicated that the assumption of equality of error variances was not violated for T1 ($F(1, 44) = 2.728$, $p = .106$). Similarly, there was no significant difference in variances observed for T2 ($F(1, 44) = .779$, $p = .382$) or E2 ($F(1, 44) = 2.002$, $p = .164$). These findings suggest that the variances of scores were similar and did not significantly differ for these trials. However, for E1, the assumption of equal variances was violated ($F(1, 44) = 10.849$, $p = .002$), indicating a significant difference in variances for this trial. These results

suggest that the variances of scores were not equivalent for the E1, suggesting potential heterogeneity in the score distribution compared to the other trials. See Table 6.

Research Question 3: Order Effect for Test Formats

Research Question 3 addressed the following question: Is the order effect of test formats (traditional compared to electronic) statistically significant?

Mauchly's Test of Sphericity was analyzed to determine if the order effect of test formats resulted in statistically significant differences in scores. Mauchly's test evaluates whether the assumption of sphericity is violated in repeated measures designs. Sphericity refers to the assumption that the variances of the differences between all possible pairs of conditions are equal. If the assumption of sphericity is violated, it suggests a systematic order effect is present in the data. Descriptive statistics are in Table 7.

The test yielded a significant result (Mauchly's $W = .320$, Approx. Chi-Square = 48.708, $df = 5$, $p < .001$), indicating a violation of the assumption of sphericity. Therefore, the sphericity assumption was not met for the data. See Table 8. Considering this violation, the Greenhouse-Geisser correction ($\varepsilon = .568$) was applied to adjust the degrees of freedom in subsequent statistical analyses to account for the violation of sphericity. To account for the violation of the sphericity assumption, the Greenhouse-Geisser correction was applied to the data (Table 9). The corrected analysis revealed a significant main effect of the independent variable on the dependent variable (Sum of Squares = 21.492, $df = 1.704$, mean square = 12.615, $F = 10.115$, $p < .001$, $\eta^2 = .187$). These findings indicate that the order of test administration had a statistically significant effect on the scores, even after adjusting for the violation of the sphericity assumption using the Greenhouse-Geisser correction.

CHAPTER IV

Discussion

Detecting malingering in psychological assessments is paramount as it ensures evaluations' reliability and validity. Malingering refers to the deliberate exaggeration or feigning of symptoms or impairments to obtain personal gains, such as financial compensation or prescription medication. Unrecognized malingering can lead to misdiagnosis, inappropriate treatment, and allocation of limited resources to individuals who do not genuinely require them. Furthermore, it undermines the integrity of the assessment process and may affect the rapport between clinician and client (Dickens & Curtis, 2019).

Multiple studies have highlighted the significance of detecting malingering in psychological assessments. For instance, Rogers et al. (1998) emphasized the need for clinicians to employ robust techniques and measures to identify malingering, considering its prevalence and potential impact on clinical decision-making. Similarly, a study by Bianchini, Greve, and Glynn (2005) emphasized the importance of utilizing validated malingering assessment tools to differentiate genuine psychiatric disorders from feigned symptoms.

Two types of malingering assessments can be conceptualized as symptom validity assessments and performance validity assessments. Symptom validity measures are typically self-report measures used to evaluate the credibility and validity of reported symptoms compared to typically reported symptoms of various diagnoses (Bush et al., 2005). Performance validity assessments are measures designed to assess an individual's effort and motivation in performing cognitive and psychological tasks (McWhirter et al., 2020). Both types of malingering

assessments aid clinicians in identifying potential exaggeration or feigning of symptoms, thereby ensuring the integrity and validity of the evaluation process.

The TOMM is a frequently used performance validity assessment of malingering. However, current administration procedures necessitate the use of the traditional flipbook. An electronic format of the TOMM would reduce human error in administration, allow for greater efficiency in administration, and decrease barriers to access for remote populations. Thus, the current study's purpose was to validate an electronic version of the TOMM (the TOMM-E). Previous studies used an examinee-administered electronic version of the TOMM, introducing the additional possibility of human error and disregarding standardized administration guidelines. This study compared a clinician-administered electronic version of the TOMM to the traditionally administered version of the TOMM across the same participants. Using the same participants across both assessment versions can be conceptualized as a test-retest model, reducing individual heterogeneous variables across administrations. Results provide support for the validation of the electronic version of the TOMM. These results, limitations, and implications for the use of the TOMM-E are described below.

Research Question 1

The present study aimed to investigate the potential differences between the traditional version of the TOMM and its electronic counterpart, the TOMM-E. Research Question 1 examined statistically significant differences between these two assessment versions. To address this question, a paired sample t-test was initially conducted to assess the correlation between scores obtained on the two trials within the TOMM and the TOMM-E. Analysis revealed a strong positive correlation of .805 and .755, respectively, indicating a significant association

between scores obtained on the two trials of each assessment. Furthermore, a repeated measures analysis of variance (ANOVA) was employed to explore the effects of the within-subjects factor, which included the two trials of each assessment format (T1, T2, E1, E2), on participant scores. The main objective was to investigate if there were any notable differences in mean scores across trials on the two versions of the TOMM. The results of the repeated measures ANOVA did not yield statistically significant main effects ($F(3) = 1.756$, $p = .159$, $\eta^2 = .038$), indicating that there were no substantial differences in mean scores between the various levels of the TOMM formats. These findings highlight the comparable nature of the TOMM and the TOMM-E, as indicated by the high correlation between their scores and the lack of significant differences in mean scores across trials.

Research Question 2

Research Question 2 sought to determine if the distribution of scores on the TOMM and the TOMM-E were equivalent. When values are described as "statistically equivalent," the observed difference between groups is smaller than a meaningful threshold and falls within the range specified by the equivalence bounds (Lakens et al., 2018). Results from Levene's Test of Equality of Error Variances indicated that the assumption of equal variances was not violated for T1 ($F(1, 44) = 2.728$, $p = .106$); T2 ($F(1, 44) = .779$, $p = .382$); and E2 ($F(1, 44) = 2.002$, $p = .164$). The assumption of equal variances was violated for E1 ($F(1, 44) = 10.849$, $p = .002$). O'Brien (1992) suggests that outliers can significantly impact the results of statistical tests, including tests of equal variances like Levene's test. One of the critical consequences of outliers is their ability to inflate estimates of variability. In the present study, examining minimum and maximum values for each trial provides valuable insights into the presence of outliers and their potential influence on the results.

In Table 2, the descriptive statistics for the TOMM and TOMM-E assessments present the range of values for each trial. Notably, the maximum values across all trials were consistent, with a maximum score of 50. This consistency suggests that the upper range of performance remained stable across the different trials and versions of the assessment. However, the analysis revealed a disparity in the minimum values. Specifically, the minimum value observed for E1 (43) was significantly lower than the minimum values for T1 (46), T2 (45), and E2 (45). This discrepancy indicates the presence of a potential outlier in the E1 trial, as evidenced by the notably lower score compared to the other trials. Identifying this outlier emphasizes the need for a cautious interpretation of Levene's Test of Equality of Error Variances results. Overall, examining minimum and maximum values in the TOMM and TOMM-E assessments provides valuable information regarding the presence of outliers and their potential impact on the observed data.

Research Question 3

Research Question 3 examined if the order of administration of test formats (traditional prior to electronic vs. electronic prior to traditional) had a significant result on the obtained outcomes. Test format assignment was counterbalanced through simple random assignment of participants, with 21 completing the traditional version first and 25 completing the electronic version first. The results of the present study demonstrated a violation of the assumption of sphericity, as evidenced by a significant result obtained from Mauchly's Test of Sphericity (Mauchly's W = .320, Approx. Chi-Square = 48.708, df = 5, $p < .001$). This violation indicated that the variances of the differences between all possible pairs of conditions were not equal. As a result, the data did not meet the sphericity assumption, highlighting the need for further analysis.

To address the violation of sphericity, the Greenhouse-Geisser correction was employed, with an epsilon value of .568. This correction adjusted the degrees of freedom in subsequent statistical analyses to account for the violation of sphericity. Applying the Greenhouse-Geisser correction (Table 9) allowed for a more accurate examination of the main effect of the test format on the dependent variable. The corrected analysis revealed a significant main effect of the independent variable on the dependent variable (Sum of Squares = 21.492, df = 1.704, mean square = 12.615, $F = 10.115$, $p < .001$, $\eta^2 = .187$). These findings indicate that the order of test administration had a statistically significant effect on the scores, even after adjusting for the violation of the sphericity assumption using the Greenhouse-Geisser correction. Therefore, it can be concluded that the sequence in which the tests were administered influenced the participants' performance on obtained scores.

One possible explanation for the order effect might be practice effects. Repeated administrations of an assessment can introduce practice effects, which are improvements in performance due to increased familiarity and experience with the task. Practice effects are a common concern in repeated measures studies, as they can confound the interpretation of change over time (Schretlen et al., 2003). These effects can manifest as increased speed, accuracy, or efficiency in task completion. Practice effects are driven by factors such as improved testing strategies, reduced anxiety, and reduced novelty-related distractions (Kramer et al., 1995; Roche et al., 2015). Another possible explanation for the order effect might be attributed to changes in motivation and effort. Participants may initially be highly motivated and engaged, leading to better performance on the first test format. However, as the testing session progresses, motivation may decline, affecting participants' effort and performance in later test conditions (Bridger et al., 2013).

Despite the presence of the order effect in this data, in applied settings, two formats of the assessment would not be administered to the same client. As such, although the order effect is statistically significant, it is unlikely to impact clinical uses for the TOMM or the TOMM-E.

Implications, Limitations, and Future Directions

In summary, results indicate that the electronic version (TOMM-E) is a valid tool for assessing memory malingering. In addition, the TOMM-E has been demonstrated to be psychometrically equivalent to the traditional version of the TOMM. The TOMM-E fills a gap in the field of practice by allowing clinicians to administer an electronic assessment of memory malingering while reducing the possibility of human error in administration. The only previous study to examine the validity of the TOMM-E was examinee administered, thus failing to adhere to standardized administration procedures and introducing an additional element of human error (Vanderslice-Barr et al., 2011).

A notable limitation of this study is the lack of generalizability of these findings, given the relatively small sample size and highly specific population from which participants were drawn (undergraduate students enrolled in a teacher preparation course). The sample consisted of only 46 participants after data cleaning was complete, 84.7% of whom were females. Future research would benefit from obtaining larger sample sizes of a more heterogenous nature to represent differing demographics better. In addition, less than 50 participants yielded usable results out of more than 225 potential participants. There are several possible reasons for this paucity of participation, including time conflicts when scheduling assessment appointments, satisfaction with current grade in the course (and thus not attempting to earn extra credit), and other obligations on potential participant's time, including work, school assignments, and family

obligations. Future directions of study could include validating the TOMM-E with other populations, including elderly individuals, those with cognitive delays, and those with chronic pain conditions.

Another notable limitation in this study is the ceiling effect on the TOMM and the TOMM-E. The ceiling effect refers to a phenomenon where a majority of individuals, particularly those with intact cognitive abilities, achieve near-perfect scores on the test (Larrabee, 2012). In this study, all but one participant scored at least 45 (out of 50) across all trials. The ceiling effect can limit the sensitivity of the TOMM in detecting malingering, as a vast majority of individuals obtain high scores on this test. To address the ceiling effect, future studies may examine a performance discrepancy between the initial two trials of the TOMM and the retention trial. It is expected that participants would perform similarly across the initial trials and the retention trial, and any discrepancies in performance could suggest the presence of malingering.

Overall, the TOMM-E demonstrates its utility as a valuable tool for assessing memory malingering. The validation of the electronic version offers several advantages over the traditional paper-based administration. Firstly, the TOMM-E minimizes the potential for human error in the administration process. With the automated presentation of stimuli, the electronic format ensures consistent assessment delivery, reducing variability associated with human factors. Furthermore, the TOMM-E offers increased efficiency in the administration process. The electronic format allows for quicker administration, as clinicians can complete the assessment without the need for manual page-turning or manipulation of the flipbook. This streamlined process saves time for both the evaluator and the participant.

In addition, the TOMM-E addresses the barriers to access that can hinder traditional in-person assessments. By utilizing remote administration, the TOMM-E can increase opportunities for individuals with limited mobility, geographical constraints, or other limitations preventing them from attending in-person assessments. This inclusion of remote and diverse populations ensures a more comprehensive representation of individuals in memory malingering assessments and promotes equity in access to psychological evaluations.

In conclusion, the TOMM-E has been validated as statistically equivalent to the TOMM across both trials. It provides a valuable solution for assessing memory malingering remotely, overcoming the limitations associated with traditional paper-based assessments. Its advantages in reducing human error, improving efficiency, and increasing access to diverse populations make it a valuable tool in the field of psychological assessment.

REFERENCES

Allen, D.N. (2011). Portland Digit Recognition Test. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3206

Alozai, U. U., & McPherson, P. K. (2022). *Malingering*. StatPearls Publishing.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

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, 28*(4), 486-496. https://doi.org/10.1080/23279095.2019.1658585

Bianchini, K. J., Mathias, C. W., Greve, K. W., Houston, R. J., & Crouch, J. A. (2001). Classification accuracy of the Portland Digit Recognition test in traumatic brain injury. *The Clinical Neuropsychologist*, *15*(4), 461–470. https://doi.org/10.1076/clin.15.4.461.1890

Bianchini, K.J., Greve, K.W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5*(4), 404-417. https://doi.org/10.1016/j.spinee.2004.11.016.

Bigler, E.D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury*, *28*(13-14), 1623-1638. https://doi.org/10.3109/02699052.2014.947627

Boo, J., & Vispoel, W. (2012). Computer versus Paper-and-Pencil Assessment of Educational Development: A Comparison of Psychometric Features and Examinee Preferences. *Psychological Reports*, *111*(2), 443-460. https://doi.org/10.2466/10.03.11.PR0.111.5.443-460

Boone, K., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-Item recognition trial: A technique to enhance sensitivity of the Rey 15-Item memorization test. *Journal of Clinical and Experimental Neuropsychology, 24,* 561-573. https://doi.org/10.1076/jcen.24.5.561.1004

Bracken, M.R., Mazur-Mosiewicz, A., & Glazek, K. (2019). Trail Making Test: Comparison of paper-and-pencil and electronic versions. *Applied Neuropsychology: Adult, (26)*6, 522-532. https://10.1080/23279095.2018.1460371

Bridger, E. K., Balcetis, E., & Dunning, D. (2013). Order in the face of chaos: Effects of order of presentation on judgments of competence from facial appearance. *Social Cognition, 31*(6), 733-746.

Bridges, A. J., & Ahern, D. C. (2016). Assessment of malingering in minority populations. In F. R. Ferraro (Ed.), *Minority and cross-cultural aspects of neuropsychological assessment: Enduring and emerging trends* (pp. 436–458). Taylor & Francis.

Bush, S.S., Ruff, R.M., Troster, A.I., Barth, J.T., Koffler, S.P., Pliskin, N.H., Reynolds, C.R., & Silver, C.H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20,* 419-426.

Butcher, J.N., Arbisi, P.A., Atlis, M.M., & McNulty, J.L. (2008). The construct validity of the Lees-Haley Fake Bad Scale: Does this scale measure somatic malingering and feigned

emotional distress? *Archives of Clinical Neuropsychology, 23*(7-8), 855-864. https://doi.org/10.1016/j.acn.2008.10.001.

Cercy, S.P., Schretlen, D.J., & Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomena. R. Rogers (Ed.), Clinical assessment of malingering and deception (2nd ed), Guilford Press, New York (1997), pp. 85-107

Chuah, S.C., Drasgow, F., & Roberts, B.W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality, 40*(4), 359-376. https://doi.org/10.1016/j.jrp.2005.01.006

Crişan, I., & Erdodi, L. (2022). Examining the cross-cultural validity of the Test of Memory Malingering and the Rey 15-item test. *Applied Neuropsychology: Adult.* Advance online publication. https://doi.org/10.1080/23279095.2022.2064753

Cullum, C.M. & Bigler, E.D. (1987). Lateralized cerebral dysfunction and the MMPI revisited. Paper presented at the 15th annual meeting of the International Neuropsychological Society, Washington, D.C.

Curtis, D. A., & Hart, C. L. (2015). Pinocchio's nose in therapy: Therapists' beliefs and attitudes toward client deception. *International Journal for the Advancement of Counselling*, *37*, 279–292. doi:10.1007/s10447-015-9243-6

Dickens, C. R., & Curtis, D. A. (2019). Lies within the law: Therapist' beliefs and attitudes about deception. *Journal of Forensic Psychology Research & Practice*, *19*(5), 359–375. https://doi-org.utk.idm.oclc.org/10.1080/24732850.2019.1666604

Etherton, J.L., Bianchini, K.J., Greve, K.W., & Ciota, M.A. (2005). Test of Memory Malingering Performance is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of Clinical Neuropsychology, 20* (3), 375-384. https://doi.org/10.1016/j.acn.2004.09.007

Frederick, R.I. (2000). A personal floor effect strategy to evaluate the validity of performance on memory tests. *Journal of Clinical and Experimental Neuropsychology*, *22*(6), 720-730. https://doi.org/10/1076/jcen.22.6.720.951

Galioto, R., Dhima, K., Berenholz, O., & Busch, R. (2020). Performance validity testing in multiple sclerosis. *Journal of the International Neuropsychological Society*, *26*(10), 1028-1035. https://doi:10.1017/S1355617720000466

Gervais, R. O., 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: The Official Journal of the National Academy of Neuropsychologists*, *19*(4), 475–487. https://doi.org/10.1016/j.acn.2003.05.001

Grant, A. F., & Werner, N. J. (2020). Retrospective analysis of the Test of Memory Malingering in a low intellectual quotient intractable epilepsy sample. *Archives of clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *35*(6), 726–734. https://doi.org/10.1093/arclin/acaa022

Graue, L.O., Berry, D.T., Clark, J.A., Sollman, M.J., Cardi, M., Hopkins J., & Dellynda, W. (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, 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., Gola, T., & Baker, W.J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist, 9*(3), 230-240. https://doi.org/10.1080/13854049508400485

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.

Gudmundsson, S., Mortensen, E. L., & Sestoft, D. (2021). The Use of Psychological Assessment and Validity Tests in Forensic Psychiatric Examinations. *Journal of Forensic Psychology Research & Practice, 21*(2), 118–130. https://doiorg.utk.idm.oclc.org/10.1080/24732850.2020.1836885

Gwaltney, C.J., Shields, A.L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health, 11,* 322-333.

Hartman, D.E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology, 17,* 709-715.

Heaton, R.K., Smith, H.H., Lehman, R.A., & Vogt, A.T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology,* 46, 892-900.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American Academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129.

Hendrickson, A. R., Massey, P. D., & Cronan, T. P. (1993). On the Test-Retest Reliability of Perceived Usefulness and Perceived Ease of Use Scales. *MIS Quarterly*, *17*(2), 227–230. https://doi.org/10.2307/249803

Hurley, K.E., & Deal, W.P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation, 44*(2), 112-119.

Huskey, M.L. (2002). Portland Digit Recognition Test, The Test of Memory Malingering and standard neuropsychological measures: Comparing efficacy, the effects of coaching, and face validity. *Graduate Student Theses, Dissertations, & Professional Papers.* 4993.

Iverson, G.L. (1995). Qualitative aspects of malingered memory deficits. *Brain Injury*, *9*(1), 35-40.

Iverson, G.L., & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation (15)*2, 829-858.

Kak, A. (1981). Relationship between readability of printed and CRT-displayed text. *Proceedings of the 25th annual meeting of the Human Factors Society.* Santa Monica, CA: Human Factors and Ergonomics Society, 137-140.

Kline, R.B. (2005). Principles and Practice of Structural Equation Modeling (2nd Edition ed.). New York: The Guilford Press.

Kirk, J.W., Hutaff-Lee, C.F., Connery, A.K., Baker, D.A., & Kirkwood, M.W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21*(5). 562-569.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training and transfer effects in task switching. *Memory & Cognition, 23*(4), 517-526.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269.

Larrabee, G. (2003). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology, 18*(6). 673-686. https://doi.org/10.1016/S0887-6177(02)00157-9

Larrabee, G. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625-630. https://doi:10.1017/S1355617712000240

Lebourgeois, H.W. (2007). Malingering: Key points in assessment. *Psychiatric Times, 24*(4).

Lipman, F.D. (1962). Malingering in personal injury cases. *Temple Law Quarterly, 35*, 141–62.

Mackin R.S., Ayalon L., Feliciano L., & Areán P.A. (2010). The sensitivity and specificity of cognitive screening instruments to detect cognitive impairment in older adults with severe

psychiatric illness. *Journal of Geriatric Psychiatry, 23*(2), 91-99. https://doi: 10.1177/0891988709358589.

McWhirter L, Ritchie CW, Stone J, & Carson, A. (2020). Performance validity test failure in clinical populations - a systematic review. *Journal of Neurology, Neurosurgery & Psychiatry, 91,* 945-952.

Meyers, J.E., & Diep, A.N. (2000). Assessment of Malingering in Chronic Pain Patients Using Neuropsychological Tests. *Applied Neuropsychology, 7*, 133 - 139.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of clinical and experimental neuropsychology, 24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2019). The effect of literacy and culture on cognitive effort test performance: An examination of the Test of Memory Malingering in Colombia. *Journal of clinical and experimental neuropsychology, 41*(10), 1015–1023. https://doi.org/10.1080/13803395.2019.1644294

Noyes, J.M., & Garland, K.J. (2008). Computer-vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*(9), 1352-1375.

Pinsoneault, T.B., 1996. Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. *Computers in Human Behavior, 12*, 291–300.

Reynolds, C.R & Kamphaus, R.W. (2004). *Behavioral Assessment System for Children-Second Edition manual.* Circle Pines, MN: AGS Publishing.

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

Roche, N. L., Fleming, J. M., & Shores, E. A. (2015). A comprehensive examination of the influence of demographic and psychosocial factors on neuropsychological test performance in adults. *Archives of Clinical Neuropsychology, 30*(6), 640-657.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and non-forensic malingerers: A prototypical analysis of explanatory models. *Law and human behavior*, *22*(4), 353–367. https://doi.org/10.1023/a:1025714808591

Rose, F.E., Hall, S., & Szalda-Petree, A.D. (1995). Portland digit recognition test-computerized: Measuring response latency improves the detection of malingering, *The Clinical Neuropsychologist, 9*(2), 124-134. https://doi.org/10.1080/13854049508401594

Ross C. A. (2019). Problems with factitious disorder, malingering, and somatic symptoms in DSM-5. *Psychosomatics*, *60*(4), 432–433. https://doi.org/10.1016/j.psym.2018.11.003

Ruff, R.M., Wylie, T.F., & Tennant, W. (1993). Malingering and malingering-like aspects of mild closed head injury. *Journal of Head Trauma Rehabilitation, 8*, 60-73.

Schoenberg, M. R., Dorr, D., & Morgan, C. D. (2003). The ability of the Millon Clinical Multiaxial Inventory-Third Edition to detect malingering. *Psychological Assessment, 15*(2), 198–204. https://doi.org/10.1037/1040-3590.15.2.198

Schretlen, D. J., Testa, S. M., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14*(3), 436-445.

Slick, D.J., Hopp, G, Strauss, E., & Spellacy, F.J. (1996). Victoria symptom validity test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology. 18*(6), 911-922. https://doi.org/10.1080/01688639608408313

Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology*. *63*(4), 339–344. https://doi.org/10.1002/jclp.20351

Smith, G.P., & Burger, G.K. (1997) Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law. 25*(2), 183-189.

Swift, A., Heale, R., & Twycross, A. (2020). What are sensitivity and specificity? *Evidence-Based Nursing,* 23, 2-4.

Teichner, G., & Wagner, M.T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology, 19,* 455-464.

Tombaugh, T.N., & McIntyre, N.J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*(9), 922 – 935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.

Tombaugh, T. N. (1996). *Test of memory malingering*. North Tonawanda, NY: Multi-Health Systems.

Tombaugh, T. N., Rees, L., & McIntyre, N. (1998). Normative data for the Trail Making Test. Personal communication cited in Spreen and Strauss. A compendium of neuropsychological tests: Administration, norms and commentary (2nd ed.). New York: Oxford University Press.

Tracy, D. (2014). Evaluating malingering in cognitive and memory examinations: A guide for clinicians. *Advances in Psychiatric Treatment, 20*(6), 405-412. https://doi:10.1192/apt.bp.114.012906

Vanderslice-Barr, J.L., Miele, A.S., Jardin, B & McCaffrey, R.J. (2011). Comparison of computerized versus booklet versions of the TOMM. *Applied Neuropsychology, 18*(1), 34-36. https://doi:10.1080/09084282.2010.523377

Vispoel, W.P., 2000. Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. *Measurement and Evaluation in Counseling and Development, 33,* 130–143.

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*. https://doi.org/10.3389/fpsyt.2018.00700

Walter, J., Morris, J., Swier-Vosnos, A., & Pliskin, N. (2014). Effects of severity of dementia on a symptom validity measure. *The Clinical neuropsychologist*, *28*(7), 1197–1208. https://doi.org/10.1080/13854046.2014.960454

Wygant, D.B., Sellbom, M., Ben-Porath, Y.S., Stafford, K.P., Freeman, D.B., & Heilbronner, R.L. (2007). The relation between symptom validity testing and MMPI-2 scores as a

function of forensic evaluation context. *Archives of Clinical Neuropsychology, 22*(4), 489-499. https://doi.org/10.1016/j.acn.2007.01.027

Ziefle, M. (1998). Effects of display resolution on visual performance. *Human Factors, 40,* 554-568.

APPENDIX

Appendix A

Institutional Review Board (IRB) Informed Consent

Consent for Research Participation

Research Study Title: *Test of Memory Malingering – Electronic Validation Study*

Researcher(s): *Dr. Brian Wilhoit,* University of Tennessee, Knoxville
Dr. Steve McCallum, University of Tennessee, Knoxville
Bonnie Smith, University of Tennessee, Knoxville

Why am I being asked to be in this research study?

We are asking you to be in this research study because you are a college student between the ages of 18 and 29.

What is this research study about?

The purpose of the research study is to validate an electronic assessment for a memory task.

How long will I be in the research study?

If you agree to be in the study, your participation will last the length of two administrations of the memory assessment (on separate days) for a total of 45 - 60 minutes.

What will happen if I say "Yes, I want to be in this research study"?

If you agree to be in this study, we will ask you to:

- Schedule two time slots that are convenient for you to participate.
- Have access to an electronic device that is connected to the internet.
- Join a Zoom meeting with the research assistant.

Participate in two memory tasks, one of which will be administered via a document camera and the second of which will be administered on screen via a slide software program. The administration of this assessment via the slide software program is experimental.

What happens if I say "No, I do not want to be in this research study"?

Being in this study is up to you. You can say no now or leave the study later.

Either way, your decision won't affect your grades, your relationship with your instructors, or standing with the University of Tennessee.

Instead of participating in the study, options available to you include:

- Writing a 400-500 word paper on which of the learning theories (behavioral, constructivism, cognitive or social-cognitive) discussed in this class would be the most beneficial for use in your future classroom. Please contact your instructor for additional details on this assignment.

What happens if I say "Yes" but change my mind later?

Even if you decide to be in the study now, you can change your mind and stop at any time.

If you decide to stop before the study is completed:

- At any point during this study, you are free to withdraw your participation. Simply let the researcher know that you no longer wish to participate and the assessment will immediately stop.
- Any information that has been collected prior to your withdrawal will be destroyed and not used in any research activities.

After completing both of the memory tasks, your data will be assigned a random number and de-identified. At this point, you will no longer be able to withdraw your data from this study.

Are there any possible risks to me?

We don't know of any risks to you from being in the study.

Are there any benefits to being in this research study?

We do not expect you to benefit from being in this study. Your participation may help us to learn more about memory malingering among college students. We hope the knowledge gained from this study will benefit others in the future.

Who can see or use the information collected for this research study?

We will protect the confidentiality of your information by de-identifying your information after completion of both memory assessments and assigning your data a random number.

All study data, including electronic consent forms, will be stored on a secure database for a period of five years. The only people who will have access to this data are those who are directly involved in this study including the principal investigator and research assistants.

If information from this study is published or presented at scientific meetings, your name and other personal information will not be used.

We will make every effort to prevent anyone who is not on the research team from knowing that you gave us information or what information came from you. Although it is unlikely, there are times when others may need to see the information we collect about you. These include:

- People at the University of Tennessee, Knoxville who oversee research to make sure it is conducted properly.

- Government agencies (such as the Office for Human Research Protections in the U.S. Department of Health and Human Services), and others responsible for watching over the safety, effectiveness, and conduct of the research.
- If a law or court requires us to share the information, we would have to follow that law or final court ruling.

What will happen to my information after this study is over?

Future Research:

- Your information may be used for future research studies or shared with other researchers for use in future studies without obtaining additional informed consent from you. If this happens, all of your identifiable information will be removed before any future use or sharing with other researchers.

Will I be paid for being in this research study?

- For your participation in this research study, you will be offered 30 points of extra credit towards your final grade in EDPY 401.
- If you withdraw prior to completing both assessments, you will still be eligible for partial credit. In addition, you will still be eligible to complete a portion of the alternative assignment to earn a total of 30 points

What else do I need to know?

We may need to stop your participation in the study without your consent if:

- You do not follow the study instructions

If we learn about any new information that may change your mind about being in the study, we will tell you. If that happens, you may be asked to sign a new consent form.

Your research data may be used to create products or to deliver services, including some that may be sold or make money for others. If this happens, there are no plans to provide financial payment to you or your family.

Who can answer my questions about this research study?

If you have questions or concerns about this study, or have experienced a research related problem or injury, contact the researchers, Dr. Brian Wilhoit, **bwilhoit@utk.edu**, (865)974-6177.

For questions or concerns about your rights or to speak with someone other than the research team about the study, please contact:

Institutional Review Board

The University of Tennessee, Knoxville

1534 White Avenue

Blount Hall, Room 408

Knoxville, TN 37996-1529

Phone: 865-974-7697

Email: utkirb@utk.edu

STATEMENT OF CONSENT

I have read this form and the research study has been explained to me. I have been given the chance to ask questions and my questions have been answered. If I have more questions, I have been told who to contact. By signing this document, I am agreeing to be in this study. I will receive a copy of this document after I sign it.

Name of Adult Participant	Signature of Adult Participant	Date

Researcher Signature (to be completed at time of informed consent)

I have explained the study to the participant and answered all of his/her questions. I believe that he/she understands the information described in this consent form and freely consents to be in the study.

Name of Research Team Member	Signature of Research Team Member	Date

Appendix B

Table 1

Demographic Information

	N	%
Gender	46	--
Female	39	84.7
Male	7	15.2
Age	46	--
18-21	30	63.8
22-25	11	23.4
26-29	5	10.6
Participants	46	--
ADHD Diagnosis	9	19.6

Table 2

TOMM and TOMM-E Descriptive Statistics

	N	Min	Max	Mean	SD	Skewness	Kurtosis
TOMM: Trial 1	46	46.00	50.00	48.98	1.36	-1.18	3.40
TOMM: Trial 2	46	45.00	50.00	49.17	1.30	-1.84	6.26
TOMM-E: Trial 1	46	43.00	50.00	48.89	1.58	-1.65	6.18
TOMM-E: Trial 2	46	45.00	50.00	49.26	1.10	-1.89	7.27

Table 3

Paired Samples Correlations

	N	Correlation	Significance 2-tailed
Pair 1: Doc Cam Score 1 – Doc Cam Score 2	46	.805	P < 0.001
Pair 2: Electronic Score 1 – Electronic Score 2	46	.755	P < 0.001

Table 4

Test of Within Subjects Effects

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
	3.731	3	1.244	1.756	0.159	.038

Table 5

Pairwise Comparisons Across Trials

		Mean Difference	Std. Error	Sig.	Confidence Interval: Lower Bound*	Confidence Interval: Upper Bound*
Traditional: Trial 1 (T1)	Electronic: Trial 1 (E1)	.004	.227	1.000	-.625	.632
	Electronic: Trial 2 (E2)	-.325	.159	.280	-.763	.114
Traditional: Trial 2 (T2)	Electronic: Trial 1 (E1)	.230	.232	1.000	-.412	.871
	Electronic: Trial 2 (E2)	-.099	.153	1.000	-.521	.323

*95% Confidence Interval

Table 6

Levene's Test of Equality of Error Variances

Based on Mean	Levene Statistic	df1	df2	Sig.
Traditional: Trial 1 (T1)	2.728	1	44	.106
Traditional: Trial 2 (T2)	.779	1	44	.382
Electronic: Trial 1 (E1)	10.849	1	44	.002
Electronic: Trial 2 (E2)	2.002	1	44	.164

Table 7

Descriptive Statistics for Trials

	Administered First	Mean	Std. Deviation	*N*
Traditional Score: Trial 1 (T1)	Traditional	48.48	1.470	21
	Electronic	49.40	1.118	25
Traditional Score: Trial 2 (T2)	Traditional	49.05	1.396	21
	Electronic	49.28	1.242	25
Electronic Score: Trial 1 (E1)	Traditional	49.43	.926	21
	Electronic	48.44	1.873	25
Electronic Score: Trial 2 (E2)	Traditional	49.29	.902	21
	Electronic	49.24	1.268	25

Table 8

Mauchly's Test of Sphericity

Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser
.320	48.708	5	<.001	.568

Table 9

Greenhouse-Geisser Correction

Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta. Squared
21.492	1.704	12.615	10.115	<.001	.187

VITA

Bonnie Smith was born in Gainesville, Florida to Beth and Clay Smith. She graduated from Flagler College in 2014 with Bachelor of Arts degrees in Psychology and Elementary Exceptional Student Education. From 2015 until 2018, she worked as a first-grade teacher in an inclusion classroom in Jacksonville, Florida. During this time, she earned her Master of Science degree in Psychology with a Concentration in Child and Developmental Psychology from the University of Southern New Hampshire. In August 2018, Bonnie accepted a position in the School Psychology doctoral program at the University of Tennessee, Knoxville. She earned a Master of Science degree in Applied Behavioral Analysis in August 2021. In July 2022, Bonnie accepted a predoctoral internship position at the Munroe-Meyer Institute in Omaha, Nebraska working in Behavioral Pediatrics and Integrated Primary Care. She earned her doctorate degree in August 2023.