Fall 11-12-2020

Validation of the tri-choice naming and response bias measure

Chloe A. Huston
Cleveland State University

VALIDATION OF THE TRI-CHOICE NAMING AND RESPONSE BIAS MEASURE

CHLOE A. HUSTON

Bachelor of Science in Psychology

Baldwin Wallace University

May 2019

submitted in partial fulfillment of requirements for the degree

MASTER OF ARTS IN PSYCHOLOGY

at the

CLEVELAND STATE UNIVERSITY

May 2021

We hereby approve this thesis for

CHLOE A. HUSTON

Candidate for the Master of Arts in Clinical Psychology degree

for the Department of Psychology and the

CLEVELAND STATE UNIVERSITY'S

College of Graduate Studies by

____

Thesis Chairperson, Amir Poreh, Ph.D.

____

Department & Date

____

Thesis Committee Member, Elizabeth Goncy, Ph.D.

____

Department & Date

____

Thesis Committee Member, Colleen McMahon, Ph.D.

____

Department & Date

Student's Date of Defense: 11/12/2020

VALIDATION OF THE TRI-CHOICE NAMING AND RESPONSE BIAS MEASURE

CHLOE A. HUSTON

ABSTRACT

In the field of neuropsychology, feigning symptoms, also known as malingering or response bias, is an important issue as a large number of assessment referrals are for individuals that may receive benefits if the assessment results suggest cognitive impairment. Therefore, it is crucial to have valid and reliable measures that detect the feigning of symptoms (Slick et al., 1999). Currently, the tests that are routinely used to detect response bias are vulnerable to coaching. The goal of this study was to validate a new response bias test aimed at being less susceptible to coaching than existing measures: The Tri-Choice Naming and Response Bias Measure (N-Tri). To this end, 400 participants were assigned to either the coached malingerers' group, uncoached malingerers' group, or control group. Participants completed an online survey consisting of the N-Tri, Reliable Digit Span (RDS), and Portland Digit Recognition Test (PDRT). ROC curves demonstrated that all three tests were able to detect coached malingerers from controls, but the N-Tri had higher sensitivity than the RDS and PDRT. Thus, the N-Tri was able to better detect coached malingerers than compared to the RDS and PDRT. Several additional hypotheses were examined.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Introduction to Response Bias

Malingering, or response bias, is the exaggeration of cognitive dysfunction with the goal of acquiring material gain, such as money or services, or escaping legal accountability or formal duty, such as prison (Slick et al., 1999), and is a critical issue in neuropsychology. A large number of neuropsychological client referrals are for individuals who may receive favorable personal outcomes if the assessment results document cognitive impairment. Thus, it is important to have valid and reliable measures of assessing if cognitive impairment is legitimate or exaggerated (Slick et al., 1999).

Exact prevalence rates for malingering do not currently exist (Ali et al., 2015), but the feigning of psychological symptoms following a personal injury is estimated to range from 1% (Hickling et al., 1999) to more than 50% (Resnick, 1997). Hickling et al. (1999) suggest that, in a sample from the United States, more than 50% of individual with cases emerging from motor vehicle accidents were thought to be exaggerating psychological symptoms. Using stand-alone or within-measures malingering tests, Mittenburg et al. (2002) estimate malingering in personal injury cases to be approximately 30%, while malingering in cases involving mild head trauma is estimated to be around 41%.

The issue of malingering is not a new problem. Burkett and Whitley (1998) suggest that the majority of men that are claiming to be Vietnam veterans and have Post-Traumatic Stress Disorder (PTSD) were never in combat or never even served in the military. It is estimated that close to 75% of individuals receiving rewards and compensation for Vietnam PTSD were feigning symptoms (Burkett & Whitley, 1998). This estimate was supported by Frueh et al. (2005), who, through reviewing data on 100 patients at a VA Medical Center, determined that 94% had been diagnosed with PTSD, but only 41% had documentation of combat exposure. Estimates such as these demonstrate the importance of malingering measures that are well-validated, reliable, and customarily used (Janaski et al., 2011).

Response Bias Assessment

Detecting response bias can be done by tests that are within-measures and those that are stand-alone. On both types of response bias tests, malingerers often believe that they must perform poorly and display poor performance on items that individuals with real cognitive impairments typically perform well on (Walczyk et al., 2018). Thus, those who are exaggerating symptoms often end up with failure rates that are statistically improbable (Walczyk et al., 2018). Currently, there are several well-established within-measures response bias tests. These tests are embedded within commonly used neuropsychological assessments, and include the Digit Span subtest of the Weschler Adult Intelligence Scale (WAIS-R; Weschler, 1981; WAIS-III; Weschler, 1997b; WAIS-IV; Weschler, 2008) and the F scale of the MMPI-2 (Butcher et al., 1989). Reliable digit span (RDS) is a variation of the Digit Span subtest that was created precisely for the detection of malingering. The Digit Span subtest requires strings of digits that increase in length to be repeated in forward and backward order until two consecutive strings of the same length are answered incorrectly. The longest string of digits forwards and backwards, where both trials were successfully completed, are added together to calculate the RDS (Janaski et al., 2011). The F scale of the MMPI-2 was designed to identify infrequent or unusual responses (Arbisi & Ben-Porath, 1995). This scale includes items that were endorsed by 10% or less of the normative sample (Dahlstrom et al., 1972).

Additionally, there are several stand-alone tests for the detection of exaggerated neurocognitive impairment, such as the Test of Memory Malingering (TOMM; Tombaugh, 1997), Word Memory Test (WMT; Green et al., 1996), Validity Indicator Profile (VIP; Frederick, 1997), Victoria Symptom Validity Test (VSVT; Slick et al., 1997), and Portland Digit Recognition Test (PDRT; Binder & Willis, 1991). These tests are examples of a forced-choice paradigm, where individuals are shown a stimulus and then, after a delay, must select the stimulus (target) they were shown from multiple options (distractors) (Janaski et al., 2011). For example, the PDRT uses strings of digits in its forced-choice paradigm, where participants are read a string of five digits, must count backward for a specified amount of time, and then they must select the string of digits they heard previously out of two options (Binder & Willis, 1991).

Limitations of Response Bias Measures

While forced-choice malingering tests are widely used, an important limitation is their susceptibility to coaching. Coaching is when examinees are provided with information regarding the malingering tests they will be completing (Jasinski et al., 2011) and how to perform convincingly on them (Dunn et al., 2003). Rose et al. (1998) found that around 47% of those in the uncoached malingerers' group of their study were identified as exaggerating symptoms, while around only 29% of those in the coached malingerers' group were identified as malingering. Thus, coached individuals were better at avoiding detection (Rose et al., 1998). As was previously mentioned, within neuropsychology, client referrals are often for individuals who may receive benefits or compensation if the assessment results suggest cognitive impairment. Therefore, response bias measures not sufficiently detecting coached malingerers is an important problem.

There are multiple sources that individuals may receive coaching from. For example, clients may receive information regarding symptoms they should exaggerate from their attorneys or mental health professionals (Platt & Husband, 1986). Lees-Haley (1997), Gutheil (2003), and Wetter and Corrigan (1995) have all documented the impact of attorney coaching. If clients are being given information regarding the tests they will be completing, then conclusions based off of those tests may be speculative and misleading (Lees-Haley, 1997). Guteheil (2003) brought up the point that it is difficult to differentiate between corrupt, unethical coaching and attorney guidance, which plays a role in how malingering may never be eliminated. Another prominent factor in the problem of coaching is attorneys' attitudes. Wetter and Corrigan (1995) found that approximately 50% of attorneys and 33% of law students believe that clients should be given information regarding malingering assessments prior to evaluation.

Additionally, due to the amount of information available on the internet, individuals are set up well to coach themselves (Bauer & McCaffrey, 2006; Ruiz et al., 2002). Allen and Green (2001) report a decline in failure rates on the Computerized Assessment of Response Bias (CARB; Allen et al., 1997) over a six-year stretch that corresponded to increased internet access, which suggests that individuals are likely using the internet to coach themselves. Ruiz et al. (2002) investigated how much information existed on the web that could be used to help malingerers successfully portray cognitive impairment. They sorted websites into different categories based on how much assistance each website offered. Most websites fell into the "minimal threat" category based on the information on the website only offering minimal assistance. Fewer websites were categorized as "indirect threats." These websites described signs of symptom feigning and names of popular response bias assessments. A handful of websites were categorized as "direct threat" because they included actual test stimuli and test taking strategies to avoid detection on malingering tests (Ruiz et al., 2002).

Similarly, Bauer and McCaffrey (2006) explored Google and placed the top 50 results for the TOMM, WMT, and VSVT into categories based on the amount and type of information given about each test. They found that the TOMM had the most websites that explained scoring and/or provided specific cutoff scores. All tests had at least one website that revealed the format of the test or included a link to a study that examined the test. In addition, motivated individuals could also go right to the publisher of many psychological assessments: Psychological Assessment Resources (PAR, Inc.). Berry and Schipper (2007) searched the PAR, Inc. website and found an immense amount of information on common malingering measures, including the Memory Validity Profile (MVP; Sherman & Brooks, 2015), Miller Forensic Assessment of Symptoms Test (M-Fast; Miller, 2001), and VSVT, specifically the name, acronym, purpose, and most importantly, how response bias is detected for each test.

Picture Recognition

Despite the limitations of forced choice malingering tests, namely susceptibility to coaching, there is a large amount of literature that suggests that recognition is the best procedure to detect symptom exaggeration (Tombaugh, 1997). This can best be achieved by use of pictures. There are several studies that suggest humans have a phenomenally high capacity for remembering visual information (Shepard, 1967; Nickerson, 1968; Brady et al., 2008; Standing et al., 1970). Standing et al. (1970) showed 1,100 images, with each being displayed for five seconds, to two healthy participants. After a 30-minute delay, they were shown 100 pairs of pictures, each consisting of a previously shown image and a new image. The participants correctly identified 95% and 99% of the pictures, respectively (Standing et al., 1970). Brady et al. (2008) showed healthy participants 2,500 images for three seconds each. They then were shown pairs of images, each pair consisted of the previously shown image paired with either a new image (novel condition), physically similar new image (exemplar condition), or the same image that was altered in some way, such as related to pose (state condition). It was found that participants successfully recalled 92%, 88%, and 87% of images in each condition, respectively (Brady et al., 2008).

The high capacity for picture recognition has not only been found in healthy individuals, but for various populations as well, such as older adults (Park et al., 1988; Park et al., 1986; Winograd et al., 1982), Alzheimer's disease (AD) patients (Kopelman, 1985; Freed et al., 1989), Korsakoff's syndrome patients (Kopelman, 1985), and Huntington's disease patients (Marone et al., 1986). Park et al. (1986) showed 27 older adults and 30 college students 60 total high, medium, and low detail pictures. The pictures were then paired with a never before seen picture and the participants were asked to select the image they previously saw. The results demonstrated that there was no significant difference in picture recognition between the two age groups. Freed et al. (1989) did a similar forced-choice picture recognition study but instead used AD patients and healthy controls. Picture recognition memory was tested after 10 minutes, 24 hours, and 72 hours. After 72 hours, AD patients did not perform significantly worse than the controls, suggesting the capacity for picture recognition is not significantly impacted in AD. Kopelman (1985) also conducted a picture recognition study with Korsakoff's syndrome patients and healthy controls. This study found that Korsakoff's syndrome patients did not perform significantly worse than controls in terms of picture recognition and that the rate of forgetting was not faster for individuals with Korsakoff's syndrome. Additionally, Marone et al. (1986) demonstrated that Huntington's disease patients also do not significantly differ from controls when it comes to picture recognition and do not have a faster rate of forgetting visual information than healthy controls.

Ideally, the pictures used in any malingering test would make the test appear difficult, but in reality, the test should be quite easy. That is, the judged or perceived difficulty should surpass the test's actually difficulty (Tombaugh, 1997). The reason for this is that individuals who are actually impaired should be able to perform well. However, malingerers should believe that the test would be difficult for people who are impaired, and therefore, perform poorly. Thus, in order to make the test appear difficult, the distractors used should not noticeably differ from the target stimulus, that is, the targets and distractors used should be homogeneous (Tombaugh, 1997).

Declarative Memory vs. Priming

Picture recognition is often thought of as being supported by declarative memory (Bird, 2017; Manns et al., 2003). Declarative memory is a type of long-term memory that consists of information that can be consciously brought to mind. Episodic and semantic memory are types of declarative memory. Episodic memories are those of personal experiences and events, which normally include information about the event as well as information regarding the time and place of the event (Dickerson & Eichenbaum, 2010). Semantic memory involves factual information that is not drawn from personal experiences (Camina & Guell, 2017).

Memory impairment is a symptom of brain injury that the general public is commonly aware of. Gouvier et al., (1988) found that 82% of the population reported that they knew that memory issues typically follow concussions. Additionally, individuals who exaggerate memory impairment often report multiple memory related issues (e.g. speed, forgetfulness) and perform poorly on memory assessments (Mittenberg et al., 1993). However, forced-choice picture recognition tests, such as the TOMM (Tombaugh, 1997), do not actually assess declarative memory. Instead, such tests have been found to be facilitated by priming (Warren & Morton, 1982).

Priming is a type of non-declarative memory, which involves unconscious memories, abilities, and skills (Camina & Guell, 2017). Priming is when being exposed to specific stimuli influences the responding to stimuli that are later presented (Camina & Guell, 2017). An example of priming is if an individual is asked to identify a picture after being shown only a small part of the image. If the individual was previously shown part of the picture, they will take less time to identify the image compared to if they had never been exposed to that image previously (Kolb & Wishaw, 2003). In other words, a repeated stimulus or image will be "remembered" and responded to quicker than stimuli not previously seen (Ellis et al., 1996). This type of paradigm is used in malingering tests that utilize picture recognition, such as the TOMM (Tombaugh, 1997) and 48-Picture Test (Signoret, 1979). Such tests use the presentation of an image to prime the examinee so they can quickly recognize the same image later (Warren & Morton, 1982).

Since picture recognition is facilitated by priming, it makes sense that priming has also been found to be preserved in populations such as older adults (Howard et al., 1981) and individuals with Alzheimer's disease (Chertkow et al., 1989), Huntington's disease (Shimamura, et al., 1987), Korsakoff's syndrome (Shimamura, et al., 1987), and closed head injuries (Mutter et al., 1990; Vakil & Oded, 2003; Vakil et al., 1994). Therefore, since picture recognition and priming have been found to be unaffected in many clinical populations, individuals with genuine impairments should still perform well on forced-choice picture recognition tests.

Sensitivity and Specificity

An important property of malingering tests, including picture recognition tests, is the test's sensitivity and specificity. How well an assessment does with the detection of malingering is determined using the sensitivity and specificity of the test. Sensitivity is how well a test does at correctly identifying individuals who have the condition that the test is designed to detect (Walczyk et al., 2018), in this case, those who are malingering. Specificity is how well a test does at correctly identifying individuals who do not have the condition (Walczyl et al., 2018), in this case, those who aren't malingering. Ideally, a test should have high sensitivity and specificity (Trevethan, 2017).

Summary and Conclusions

Brief final statement. With this evidence regarding the limitations of current forced-choice naming tests, namely their vulnerability to coaching, a recognition procedure using 31 images of common objects was the starting point for developing the Tri-Choice Naming and Response Bias Measure (N-Tri). The N-Tri, adapted from the Multilingual Naming Test (MINT; Gollan et al., 2012), is a novel assessment that was created to detect symptom exaggeration and aims to be less vulnerable to coaching than dual-choice tests.

Hypotheses. The purpose of this study is to validate the N-Tri for detection of malingering. To this end, volunteers were instructed to feign memory deficits with and without coaching or serve as the control group and were not given any special instructions. Several hypotheses were made. Given that the RDS and PDRT are well validated measures of malingering (Greve et al., 2007; Binder and Willis, 1991), the first hypothesis predicts that total score across all groups on the N-Tri will positively correlate highly with total score across all groups on the PDRT and RDS, and therefore, the N-Tri will demonstrate high convergent validity which provides support for the N-Tri having construct validity. The second hypothesis is that cutoff scores on the N-Tri will be able to detect coached and uncoached malingerers from controls. The third hypothesis predicts that compared to the Portland Digit Recognition Test and Reliable Digit Span total scores, scores on the N-Tri will better detect coached malingerers from controls. The fourth hypothesis is that education and age will not impact scores across all groups on any trials and total score of the N-Tri. This hypothesis comes from studies demonstrating that priming is preserved in older adults (e.g. Howard et al., 1981; Fleischman &

Gabrieli, 1998; Fleischman, 2007). The fifth hypothesis is that individuals in the uncoached malingerers' group will choose a higher number of top images on the N-Tri recognition phases compared to coached malingerers. The sixth and last hypothesis is that individuals in the uncoached malingerers' group will choose a higher number of overall incorrect answers, not just top images, on the N-Tri recognition phases compared to the coached malingerers' group. For hypotheses five and six, it is thought that the uncoached malingerers will choose more top images and overall incorrect answers since the coached malingerers were coached, and therefore given a test-taking strategy as described in more detail later. Due to receiving this information, the coached malingerers should not be giving answers that are very incorrect (top image) and shouldn't be giving as many incorrect answers overall as compared to the uncoached malingerers.

CHAPTER II

METHODS

Materials

The Portland Digit Recognition Test (PDRT). The PDRT is a forced-choice test and was designed for the detection of symptom exaggeration. The PDRT consists of 72 trials of digit recognition. Individuals are shown a string of 5-digits, are required to count backward out loud for 5 seconds for the first 18 trials, 15 seconds for the second 18 trials, and 30 seconds for the third and fourth 18 trials, and then are shown two strings of 5-digits and must select the string that they previously were shown (Binder & Willis, 1991).

For the purpose of this study, an online adaptation of the PDRT was developed and used. This version consisted of two blocks of 18 trials each, for a total of 36 trials. During the first 18 trials, participants listened to an audio file of the strings of 5-digits being read at a pace of 1 digit per second. They were then brought to a screen that showed a countdown clock and were instructed to count backwards out loud starting at 20 for 5 seconds. Participants were required to remain on this screen until the 5 seconds were complete. Lastly, participants were presented with two strings of 5-digits, arranged one on top of the other, and had to select the one they previously heard (recognition). The second 18 trials were the same as the first, with the stimuli being in the same order (e.g. the first 5-digit string in the first block (trial 1) was the same as the first 5-digit string in the second block (trial 19)), the main difference was the second block required the participants to count backwards starting at 50 for 10 seconds. Additionally, like the in-person PDRT, the position of the correct string during recognition changed between the two blocks (e.g. if the correct string was presented on top during the first block, it was then presented on the bottom during the second block).

The suggested PDRT cutoffs have been shown to have a sensitivity of between 20-50%, with a false positive error rate of 5% or less. However, it has been suggested that these cutoffs are conservative and using higher cutoffs could increase sensitivity to around 70% (Greve & Bianchini, 2006). Due to the PDRT that was used in this study being an adaptation, a new cutoff score was developed. This is described in more detail later.

The Tri-Choice Naming and Response Bias Measure (N-Tri). The N-Tri, adapted from the MINT (Gollan et al., 2012), is a test developed for the detection of symptom exaggeration with the aim of being less vulnerable to coaching than dual-choice tests. The N-Tri is not only a test for response bias detection but serves as a naming test as well. Having both a naming test and malingering test in one measure will save clinicians time as well as ensure that the examinee can name the images included in the test. The N-Tri is similar to the forced choice paradigm that many common malingering tests utilize, but instead of the participant having to select the target image from two options, they now must select it from three options. The target image is paired with two distractors and are arranged in a triangular fashion, with the top image never being the target. The position of the target image out of the bottom two images was randomly chosen through a random number generator. Black and white line drawings, as opposed to real photographs, were used as the targets and distractors in order to ensure greater homogeneity across images.

The N-Tri consists of three trials. The first trial includes a naming and test phase, where participants are shown 31 line-drawings (targets) and are given approximately 20 seconds to name the objects they are shown. If they do not name the picture within 20 seconds or incorrectly name it, they are first given a semantic cue. If they still do not have the correct answer, they are given a phonemic cue. For example, if the image is of a butterfly, the semantic cue would be “it’s an insect”, while the phonemic cue would be to give the first consonant sound of the word, in this case, “bu”. The line-drawings are then paired with two new line drawings (distractors) and arranged in a triangular fashion. The participants must select the target image from the three image choices. The second trial consists of a study and test phase, where participants are shown the same targets from the first trial for 3 seconds each with a 1 second interval between each image, but this time they are not required to name the images. They then must again select the target image from three options. The last trial is a delayed test trial without a study phase, where participants must select the target images that they were shown in the previous two trials from distractors after 15-minute delay.

The online version of the N-Tri that was used in this study was administered as described above with one exception – there were no cues given during the naming phase of trial 1. Participants were shown the 31 images and required to type the name of the object into a text box, but they did not receive a cue if they failed to or incorrectly named the object. For the test phases of the online adaptation, participants were shown three images, two distractors and one target, each with a number from 1-3 below it. On the same screen, participants were given a multiple-choice question with answers ranging from 1-3 and were instructed to select the answer that corresponds to the image that they were previously shown. For the study phase of the N-Tri online version, the 31 images were flashed on the screen one at a time for 3 seconds with a 1 second interval consisting of a blank white screen between each image.

For the online version of the N-Tri, one point was awarded for each correct answer on all phases of all three trials. For the in-person version of the N-Tri, scoring is the same as in the online version with exception of the naming phase. During the naming phase of trial 1, a correct answer without a cue receives 3 points, a correct answer with a semantic cue receives 2 points, a correct answer with a phonemic cue receives 1 point, and not being able to correctly name the images receives 0 points.

The N-Tri produces two different scores: a naming score and a response bias score. The naming score is the score that was earned on the naming phase of trial 1 and is considered separate from the response bias score. The response bias score is the sum of the scores earned on the recognition phase of trial 1, trial 2, and trial 3. All references to N-Tri total score or trial 1 total score do not include the naming phase of trial 1.

Reliable Digit Span (RDS). The Digit Span (Yerkes, 1921) is a neuropsychological assessment that calls for strings of digits that increase in length to be repeated in forward and backward order until two consecutive strings of the same length are answered incorrectly. To calculate the RDS, the longest string of digits forwards and backwards, where both trials were successfully completed, are added together (Janaski et al., 2011). The RDS is a well validated measure of malingering (Greve et al., 2007). A score of 5 or below has been demonstrated to correctly identify 61% of individuals who are exaggerating symptoms with a false positive rate of 8% (Greve et al., 2007). Those with actual deficits resulting from brain injury perform well on the forward trial of the digit span. Thus, if an individual performs worse on the forward digit span compared to the backwards, response bias or malingering is suspected (Yerkes, 1921).

For this study, an online version of the RDS was used. Participants listened to an audio file of strings of digits being read at a pace of 1 digit per second and then were instructed to type the string they just heard into a text box. Participants completed all trials of the RDS regardless of incorrect answers, but the RDS was scored as described above. Therefore, if a participant incorrectly completed two strings of the same length and then successfully completed two strings of the same length, the successfully completed strings did not count since two consecutive strings of the same length were previously answered incorrectly. Two successfully completed strings were only included in the calculation of the RDS if they occurred before the participant failed out by incorrectly completing two strings of the same length.

Manipulation Check. Members of both the coached and uncoached malingerers' group (described below) were given a manipulation check. Participants were asked to recall as many of the details and instructions included in the scenarios they read at the beginning of the session (described below) as they could. They were also asked to rate on a five-point Likert scale how much effort they put in to following the instructions throughout testing. Participants that did not accurately recall the instructions pertaining to their group or that rated their effort as less than three were excluded from data analysis.

Attention Checks. Included in the surveys were 13 questions aimed at ensuring the participants were paying attention. These questions included text response questions (e.g. please type “123” into the box) and multiple-choice questions (please choose option “A”). It has been suggested that removing participants for incorrectly responding to one attention check question is too rigid (Berinsky et al., 2013). Therefore, Curran (2016) recommends removing participants that fail 50% or more of the attention check questions. Each survey included 13 attention check questions. Participants were excluded from data analysis if they incorrectly answered seven or more questions.

Eligibility Criteria. Participants were screened to ensure that all participants were neurotypical, meaning that they did not have any neurological, developmental, learning, or psychiatric (excluding anxiety and depression) impairments. If a participant reported not being neurotypical, they were excluded from data analysis.

Online Surveys. The order of tests and questions in the online surveys was constructed as follows: (1) informed consent, (2) eligibility criteria, (3) the scenario for the coached and uncoached malingering groups or instructions for participants to give their best effort for the control group, (4) the first two trials of the N-Tri, (5) both PDRT blocks of 18 trials, (6) the third trial of the N-Tri, (7) forward digit span, (8) backward digit span, (9) the demographics survey, and (10) the debriefing form. At the end of the survey, participants had the option of entering into a raffle for the opportunity to win $15. After data collection was complete, one winner was selected and mailed $15. Throughout the surveys, there were reminders for both malingering groups to act as if they have a head injury and are trying to get as much compensation as they can. The coached malingerers’ group was also reminded of test taking strategies.

Groups

Groups consisted of the uncoached malingerers' group, coached malingerers' group, and the control group. The only difference between the coached and uncoached malingering groups is that the scenario (described below) for coached malingerers included additional information consisting of a test-taking strategy and description of symptoms of a head injury. The control group did not read a scenario and did not complete a manipulation check. The experimental conditions are summarized in Table 1.

Uncoached Malingerers (UM). This group consisted of 118 adult volunteers. This group read the following scenario from Rose et al. (1998). The scenario for this group does not suggest a strategy for test-taking and does not describe the consequences of head injury.

"Three weeks ago you were in a car accident that was not your fault. A car ran a stop sign while you were in the intersection and hit you. You hit your head against the dashboard and were knocked out for about 20 minutes. Afterwards, you felt "dazed" for a while. You had to stay in the hospital for two days with a severe concussion. Because another person caused this accident, you are now in a lawsuit to decide how much money you will get from the person responsible. The more severe your problems, the more money you will get in the lawsuit.

As part of the lawsuit, a psychologist is about to examine you using several tests of your memory and problem solving ability. The purpose of the testing is to decide if there were any long-term effects from your head injury. Your goal is to convince your examiner, by your performance on these tests, that you have suffered brain damage from the accident." (Rose et al., 1998, p. 351).

Coached Malingerers (CM). This group consisted of 136 adult volunteers. This group read the same scenario from Rose et al. (1998) as the UM group that is listed above, but their scenario also included the following additional paragraph. This added information suggests a strategy for test-taking and gives information regarding the consequences and symptoms of head injury.

"[In order to convince the examiner that you have brain damage] you are to try to produce the most severe problems that you can *without* making it too obvious to the examiner. That means your "brain damage" must be believable. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems, you will not win your lawsuit and you will not get anything for your injuries. In addition, you may be fined or jailed for lying in court. People who have a head injury often have problems paying attention, cannot remember things as well, and do not learn things as easily as they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember, you are to try to produce the most severe problems that you can, mimicking the performance of persons who are truly injured." (Rose et al., 1998, p. 351).

Normal Controls (NC). This group consisted of 146 adult volunteers. This group did not receive any special instructions or scenarios and was told to complete the tests to the best of their ability.

Table 1. Experimental Conditions

Experimental Condition	Scenario Given	Given Symptoms of Head Injury	Given Test-taking Strategy	Instructions Given
Uncoached Malingerers	Yes	No	No	Feign Head Injury
Coached Malingerers	Yes	Yes	Yes	Feign Head Injury
Controls	No	No	No	Perform to Best Ability

Participants

One thousand, one hundred and twenty-six total participants were recruited from ResearchMatch. 726 participants were excluded from data analysis due to one of the following reasons: not meeting the eligibility criteria, not completing the entire survey, failing more than half of the attention check questions, not recalling the instructions from their group accurately, rating their effort during the survey as less than three on a five point scale, informing the researcher that they are trained in the tests administered, or informing the researcher that they did not follow instructions. After elimination of individuals that met the aforementioned criteria, participants were 400 neurotypical, English speaking volunteers who were 18 years and older (M = 42.21, SD = 15.17) with an average education of 17.70 years (*SD* = *3.06)*. Participants were randomly assigned to one of the three groups that were described above. Sample demographic information is summarized in Table 2.

General Procedure

Volunteers 18 years of age and older recruited from ResearchMatch were randomly assigned to either the coached malingerers' group, uncoached malingerers' group, or the control group. All participants were given the online survey for their group as described above. Contents of the surveys varied based on group, but all participants in all groups completed the N-Tri, RDS, and PDRT. Once data collection was complete, analyses for all hypotheses were completed.

CHAPTER III

RESULTS

Descriptive Statistics

All analyses were computed using SPSS version 26 and Microsoft Excel version 16.0. In order to provide helpful auxiliary information, descriptive statistics were analyzed. Sample demographics are reported in Table 2. Participants were 81% female, 79% white, and 92% non-Hispanic/Latino. The three groups (coached, uncoached, controls) did not significantly differ in ethnicity, first language, years of education, race, age, or sex ($p > .05$). However, the groups did significantly differ in N-Tri total score ($F(2,397) = 497.896$, $p < .001$, $\eta^2 = .72$), N-Tri naming score ($F(2,397) = 252.090$, $p < .001$, $\eta^2 = .56$), PDRT total score ($F(2,397) = 328.464$, $p < .001$, $\eta^2 = .62$), and RDS score ($F(2,397) = 433.808$, $p < .001$, $\eta^2 = .69$).

Table 2. Sample Demographics by Group

	Coached	*Uncoached*	*Control*
N (# female)	136 (104)	118 (97)	146 (123)
Percent English as first language	93.4%	96.6%	94.5%
Percent white	79.4%	79.7%	78.1%
Percent non-Hispanic/Latino	94.1%	88.1%	93.2%
Age	41.24 ± 14.22	42.92 ± 15.31	42.53 ± 15.95
Years of Education	17.73 ± 2.98	17.66 ± 3.13	17.69 ± 3.11

A Tukey post hoc test revealed that N-Tri total scores were significantly lower for the coached (M = 48.15, SD = 17.12, p < .001) and uncoached (M = 29.06, SD = 23.61, p < .001) groups compared to controls (M = 91.23, SD = 5.34), and for the uncoached group (M = 29.06, SD = 23.61, p < .001) compared to the coached group (M = 48.15, SD = 17.12). Tukey post hoc also demonstrated that N-Tri naming scores were significantly lower for the coached (M = 13.60, SD = 7.73, p < .001) and uncoached groups (M = 10.75, SD = 10.05, p < .001) compared to controls (M = 28.91, SD = 2.21), and for the uncoached group (M = 10.75, SD = 10.05, p < .001) compared to the coached group (M = 13.60, SD = 7.73). PDRT total scores were also found to be significantly lower for the coached (M = 20.56, SD = 6.25, p < .001) and uncoached groups (M = 16.19, SD = 8.74, p <.001) compared to controls (M = 34.52, SD = 2.42), and for the uncoached group (M = 16.19, SD = 8.74, p <.001) compared to the coached group (M = 20.56, SD = 6.25). Lastly, the RDS scores were also revealed to be significantly lower for the coached (M = 2.68, SD = 2.41, p < .001) and uncoached (M = 1.43, SD = 2.16, p < .001) groups compared to controls (M = 10.18, SD = 3.20), and for the uncoached group (M = 1.43, SD = 2.16, p < .001) compared to the coached group (M = 2.68, SD = 2.41).

These results demonstrate that across scores for the N-Tri naming and total scores for the N-Tri, PDRT, and RDS that the uncoached and coached malingering groups consistently scored significantly lower than the control group, and that the uncoached malingering group consistently scored significantly lower than the coached group. These results suggest that the differences in information presented to each group at the beginning of testing (e.g. coached group receiving a test taking strategy) was influential in regard to participant responding and that the instructions given to each group were understood and followed.

Hypothesis One - N-Tri Correlations with PDRT and RDS

It was hypothesized that total score across all groups on the N-Tri will positively correlate highly with total score across all groups on the PDRT and RDS, and therefore, the N-Tri will demonstrate high convergent validity which provides evidence for construct validity. Pearson correlations were used to test this hypothesis and revealed a strong positive correlation between N-Tri total score across all groups and PDRT total score across all groups, $r(398) = .90$, $p < .001$ (one-tailed), and between N-Tri total score across all groups and RDS total score across all groups, $r(398) = .82$, $p < .001$ (one-tailed).

Stepwise multiple regressions were also conducted to provide further support for the N-Tri being correlated with the PDRT and RDS through the multiple correlation coefficient. Additionally, stepwise multiple regressions consider the multiple trials of the N-Tri that couldn't be accounted for in the Pearson correlations. The first stepwise multiple regression was conducted to determine the best combination of N-Tri trials to predict PDRT total score. At step 1 of the analysis, N-Tri trial 2 was entered into the regression equation and was significantly related to PDRT total score, $F(1, 398) = 1692.02$, $p < .001$. The multiple correlation coefficient was .90, indicating that 81% of the variance of PDRT total score could be accounted for by N-Tri trial 2 scores. At step 2 of the analysis, N-Tri trial 3 was added to the equation. N-Tri trial 3 scores explained an additional 0.8% of the variance of PDRT total score, R^2 change = .008, F change (1, 397) = 17.82, $p < .001$. Model 2 was significantly related to PDRT total score, $F(2, 397)$ = 890.678, $p < .001$. The multiple correlation coefficient was .90, indicating that 82% of the variance of PDRT score could be accounted for by N-Tri trials 2 and 3.

A stepwise multiple regression was also conducted to determine the best combination of N-Tri trials to predict RDS score. At step 1 of the analysis, N-Tri trial 3 was entered into the regression equation and was significantly related to RDS score, $F(1, 398)$ = 745.13, $p < .001$. The multiple correlation coefficient was .81, indicating that 65.2% of the variance of RDS total score could be accounted for by N-Tri trial 3 scores. At step 2 of the analysis, N-Tri trial 2 score was added to the equation. N-Tri trial 2 scores explained an additional 1.6% of the variance of RDS score, R^2 change = .016, F change (1, 397) = 19.72, $p < .001$. Model 2 was significantly related to RDS total score, $F(2, 397)$ = 399.94, $p < .001$. The multiple correlation coefficient was .82, indicating that 66.8% of the variance of RDS total score could be accounted for by N-Tri trial 3 and trial 2 scores. At step 3 of the analysis, N-Tri trial 1 score was added to the equation. N-Tri trial 1 scores explained an additional 0.4% of the variance of RDS score, R^2 change = .004, F change (1, 396) = 5.17, $p = .02$. Model 3 was significantly related to RDS total score, $F(3, 396)$ = 271.16, $p < .001$. The multiple correlation coefficient was .82, indicating that 67.3% of the variance of RDS total score could be accounted for by N-Tri trial 3, trial 2, and trial 1 scores.

This hypothesis was supported based on the Pearson correlations which demonstrated that the N-Tri highly positively correlates with the PDRT and RDS, and by the stepwise multiple regressions which demonstrated high multiple correlations coefficients. Based on these results, it can be suggested that the N-Tri has convergent validity, which provides support for construct validity and that the N-Tri is measuring response bias.

Hypothesis Two - Evaluation of Sensitivity and Specificity

It was hypothesized that cutoff scores on the N-Tri will be able to detect coached and uncoached malingerers from controls. ROC curves and Area Under the Curve (AUC) were used for this hypothesis. The AUC is a measure of diagnostic utility or the ability to differentiate between two groups (Millis & Volinsky, 2001). Fawcett (2006) suggests that an AUC of 0.8 and above is good, 0.65-0.7 is fair, and 0.5 and below is poor. The closer the AUC is to 1.0, the better the diagnostic utility (Millis & Volinsky, 2001).

For the first analysis, the uncoached and coached malingering groups were combined into one malingering group. ROC curves were used to determine optimal cutoff values. In order to maximize both sensitivity and specificity, the cutoff value for each trial and the total score of the N-Tri was determined by applying Youden's index, the sum of specificity and sensitivity minus one (Youden, 1950). Scores with the largest Youden's index were selected as the best cutoff values. For Trial 1, a score of 23.50 was determined to be the optimal cutoff value, with a sensitivity of 92.9% and a specificity of 98.6%. Trial 2 was found to have an optimal cutoff value of 27.50, with a sensitivity of 94.9% and a specificity of 97.9%, and trial 3 was found to have an optimal cutoff value of 28.50, with a sensitivity of 96.5% and a specificity of 97.3%. For total score, it was determined that 82.50 is the optimal cutoff value, with a sensitivity of 96.9% and a specificity of 96.6%. The ROC Curve for N-Tri total score for simulated malingerers and controls is displayed in Figure 1.

ROC curves were then used to determine the N-Tri's sensitivity and specificity of detecting coached and uncoached malingerers separately from controls. Using the suggested cutoffs for each trial and total score it was found that in regard to detecting coached malingerers from controls, N-Tri trial 1 detected 91.9% of coached malingerers and correctly identified 98.6% of controls as not being coached malingerers. N-Tri trial 2 correctly identified 94.9% of coached malingerers and 97.9% of controls as not being coached malingerers. N-Tri trial 3 correctly identified 96.3% of coached malingerers and 97.3% of controls as not being coached malingerers. N-Tri total score correctly identified 97.1% of coached malingerers and 96.6% of controls as not being coached malingerers. The ROC Curve for N-Tri total score for coached malingerers and controls is displayed in Figure 2.

With regard to detecting uncoached malingerers from controls, N-Tri trial 1 detected 94.1% of uncoached malingerers and identified 98.6% of controls as not being uncoached malingerers. N-Tri trial 2 detected 94.9% of uncoached malingerers and identified 97.9% of controls as not being uncoached malingerers. N-Tri trial 3 detected 96.6% of uncoached malingerers and identified 97.3% of controls as not being uncoached malingerers. N-Tri total score detected 96.6% of uncoached malingerers and identified 96.6% of controls as not being uncoached malingerers. The ROC Curve for N-Tri total score for uncoached malingerers and controls is displayed in Figure 3.

These results support this hypothesis and demonstrate that the N-Tri was able to distinguish between coached malingerers and controls, uncoached malingerers and controls, and simulated malingerers (coached and uncoached combined) and controls with high levels of sensitivity and specificity for all trials and the total score.

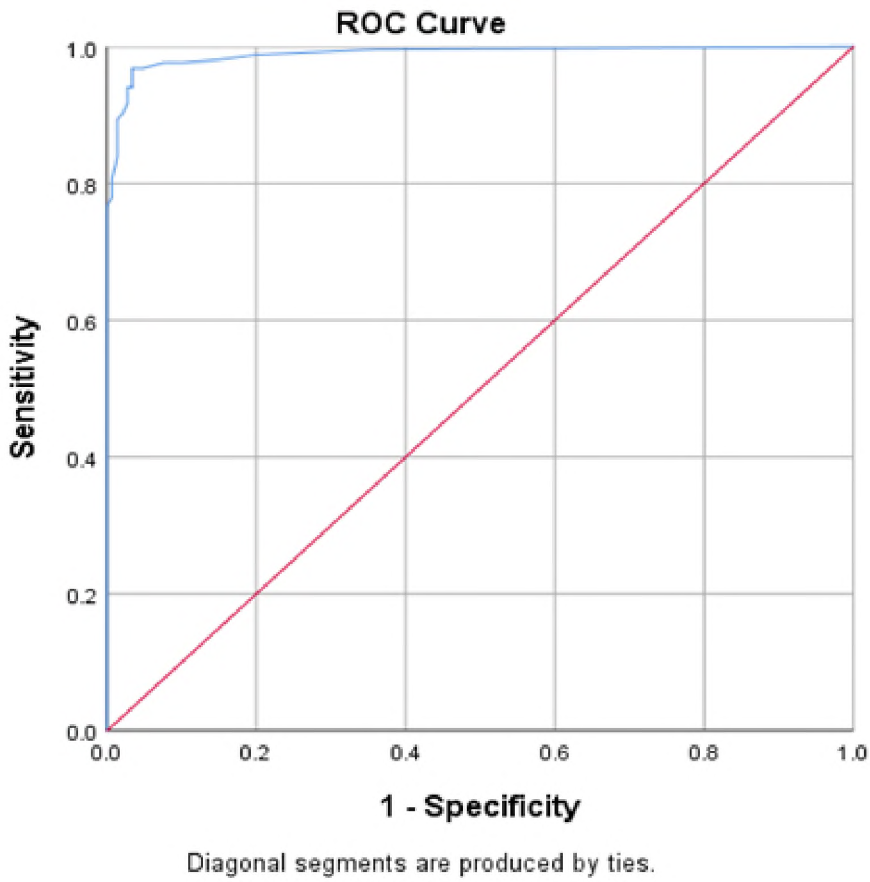

Figure 1. ROC Curve for N-Tri Total Score: Simulated Malingerers and Controls (AUC = .989)

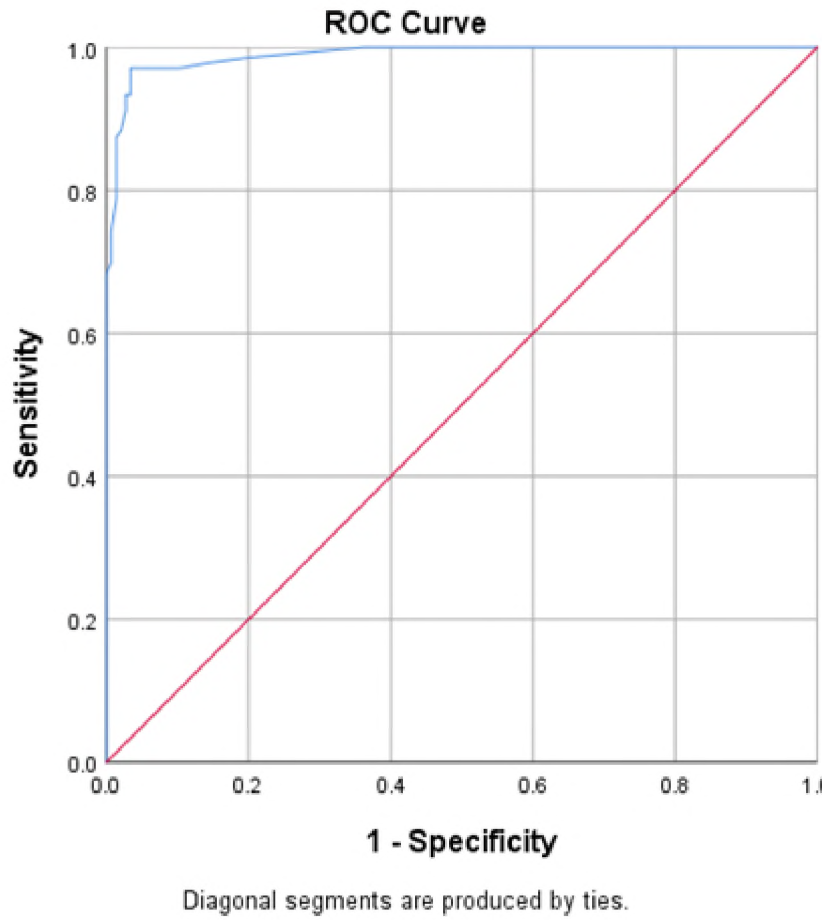

Figure 2. ROC Curve for N-Tri total score: Coached Malingerers and Controls (AUC = .989)

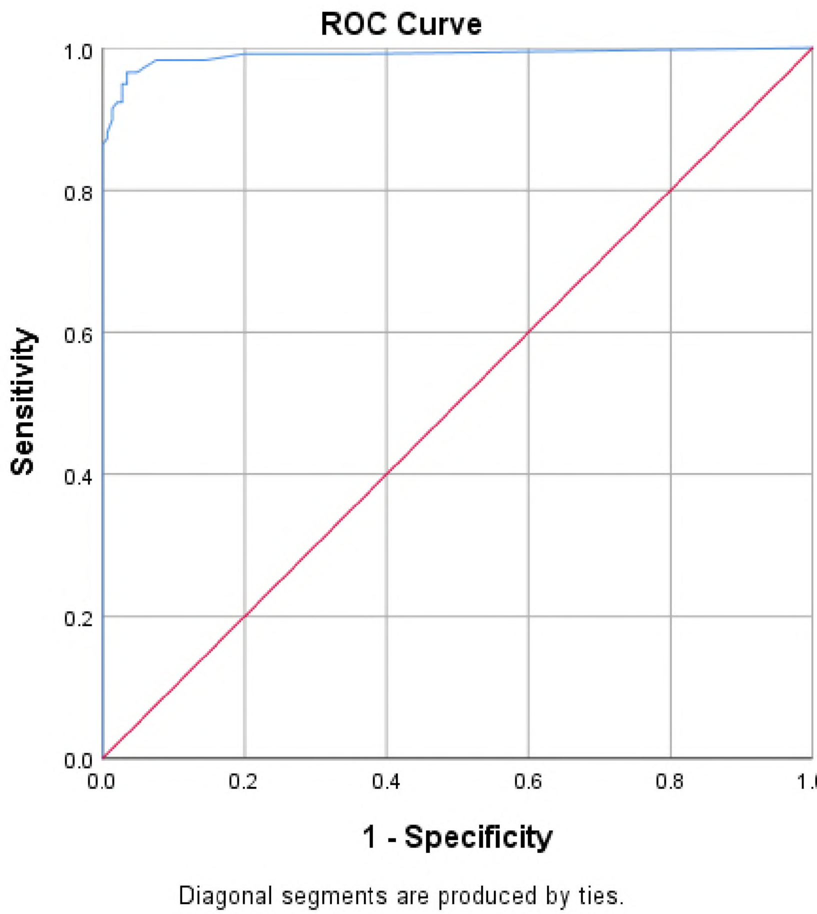

Figure 3. ROC Curve for N-Tri Total Score: Uncoached Malingerers and Controls (AUC = .990)

Hypothesis Three - Coached Malingering Detection

It was hypothesized that when compared to PDRT total score and RDS total score, N-Tri total score will better detect coached malingerers from controls. ROC curves were used to determine the sensitivity of each measure. A cutoff score of ≤ 6 or ≤ 7 is commonly used for the RDS (Schroeder et al., 2012). Using these cutoffs, a sensitivity of 89% and 94.1% were found for ≤ 6 and ≤ 7, respectively, with an AUC of .955. The suggested cutoff score for the PDRT adaptation used in this study was developed by using Youden's index (Youden, 1950) and was determined to be 29.50. Using the suggested cutoff of 29.50, PDRT sensitivity was found to be 92.6% with an AUC of .971.

Using the suggested cutoff value of 82.50 for the N-Tri, sensitivity was found to be 97.1% with an AUC of .989. Thus, the N-Tri correctly identified 97.1% of coached malingerers, while the RDS correctly identified 89% or 94.1%, depending on the cutoff, and the PDRT correctly identified 92.6% of coached malingerers. These results support this hypothesis and suggest that the N-Tri can better detect coached malingerers that the RDS and PDRT. The ROC curve for detecting coached malingerers from controls for the N-Tri is displayed in Figure 2. ROC curves for detecting coached malingerers from controls for the PDRT and RDS are shown in Figures 4 and 5.

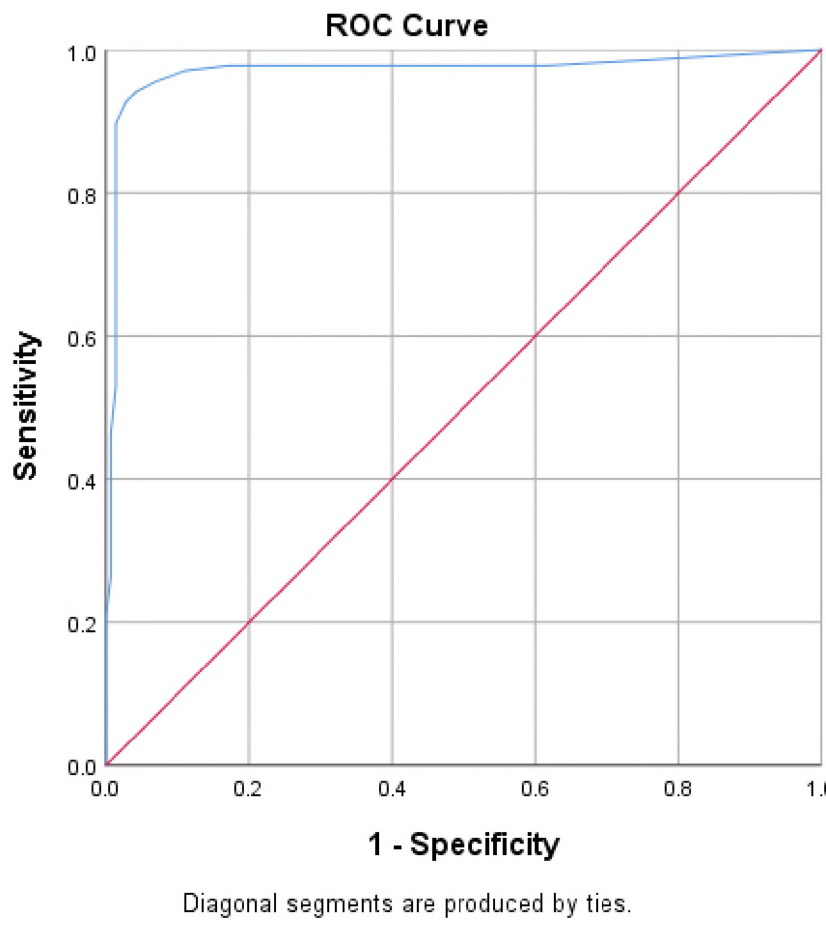

Figure 4. ROC Curve for PDRT Total: Coached Malingerers and Controls (AUC = .971)

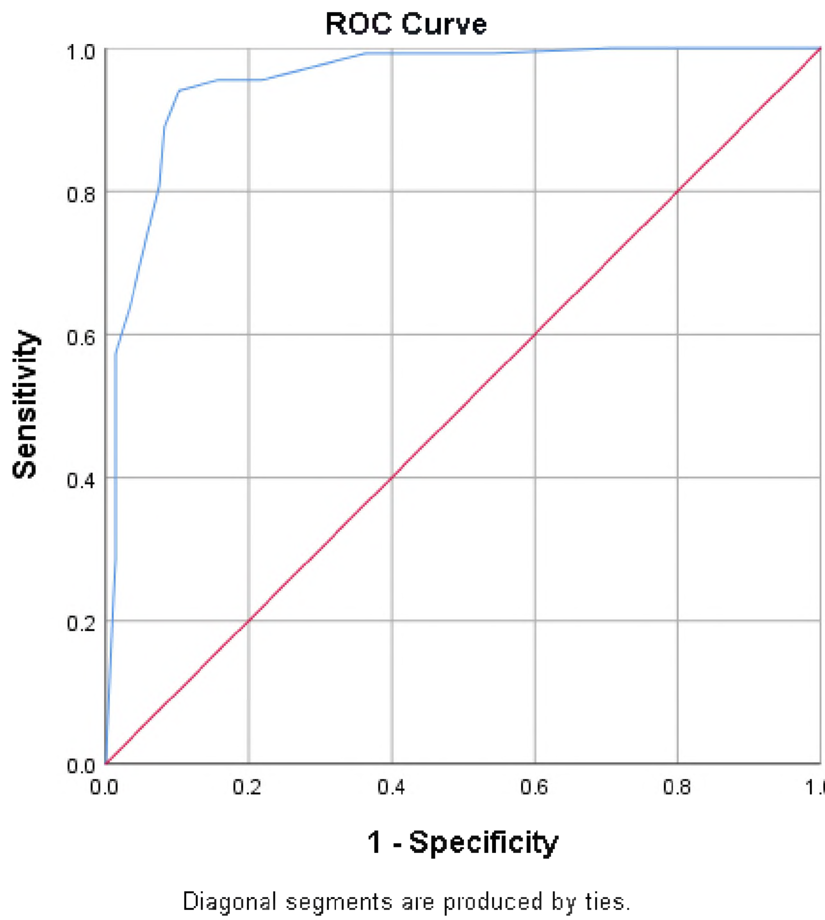

Figure 5. ROC Curve for RDS: Coached Malingerers and Controls (AUC = .955)

Hypothesis Four - Influence of Education and Age on N-Tri Scores

Multiple studies have found priming, which facilitates picture recognition (Warren & Morton, 1982), to be preserved in older adults (Howard et al., 1981; Fleischman & Gabrieli, 1998; Fleischman, 2007). If this is the case, then older adults should perform just as well on priming tasks, such as picture recognition, as younger adults. However, some studies have suggested that priming decreases with age (Chiarello and Hoyer, 1988; Abbenhuis et al., 1990; Davis et al., 1990; Hultsch et al., 1991; Ward et al., 2013). In order to ensure that individuals can perform well on the N-Tri regardless of age or education, it was hypothesized that education and age will not impact scores on any trials or total score of the N-Tri. To determine if this hypothesis is correct, Pearson correlations were used to correlate education with each trial and total score of the N-Tri, and to correlate age with each trial and total score of the N-Tri. Pearson correlation revealed no significant correlations between years of formal education and N-Tri trial 1, $r(394) = -.03$, $p = .60$ (two-tailed), years of formal education and N-Tri trial 2, $r(394) = -.00$, $p = .96$ (two-tailed), years of formal education an N-Tri trial 3 scores, $r(394) = -.01$, $p = .91$ (two-tailed), and years of formal education and N-Tri total scores, $r(394) = -.01$, $p = .82$ (two-tailed).

A Pearson correlation also demonstrated that there are no significant correlations between highest degree achieved and N-Tri trial 1 scores, $r(398) = -.05$, $p = .36$ (two-tailed), highest degree achieved and N-Tri trial 2 scores, $r(398) = -.04$, $p = .39$ (two-tailed), highest degree achieved and N-Trial trial 3 scores, $r(398) = -.04$, $p = .39$ (two-tailed), and highest degree achieved and N-Tri total score, $r(398) = -.05$, $p = .37$ (two-tailed).

In addition, it was revealed that there are no significant correlations between age in years and N-Tri trial 1 score, $r(398) = -.03$, $p = .62$ (two-tailed), age in years and N-Tri trial 2 score, $r(398) = -.04$, $p = .44$ (two-tailed), age in years and N-Tri trial 3 score, $r(398) = -.02$, $p = .77$ (two-tailed), and age in years and N-Tri total score, $r(398) = -.03$, $p = .56$ (two-tailed). It was also revealed that there are no significant correlations between age group and N-Tri trial 1 scores, $r(398) = -.03$, $p = .57$ (two-tailed), age group and N-Tri trial 2 scores, $r(398) = -.05$, $p = .34$ (two-tailed), age group and N-Tri trial 3 scores, $r(398) = -.03$, $p = .61$ (two-tailed), and age group and N-Tri total score, $r(398) = -.04$, $p = .49$ (two-tailed).

These results support this hypothesis by demonstrating that total score on the N-Tri and scores on individual trials of the N-Tri are not influenced by age or education. Therefore, regardless of age or education level, examinees can still perform well on the N-Tri. These results also provide support for picture recognition and priming not being significantly impacted by age, which align with and support the findings of Howard et al. (1981), Fleischman & Gabrieli, (1998), and Fleischman (2007), who also found priming to be intact in older adults.

Hypotheses Five and Six - Incorrect Responses

It was hypothesized that participants in the uncoached malingerers' group will choose more top images during the recognition trials of the N-Tri than coached malingerers and that participants in the uncoached malingerers' group will choose a higher number of overall incorrect images, not just top images, during the recognition trials of the N-Tri. These two hypotheses aim to ensure, beyond the manipulation check, that the coached and uncoached malingering groups understood and followed instructions. Additionally, these hypotheses look to support that including the additional information into the coached malingerers' scenario was influential to responding.

Two independent samples t-test were conducted to compare the mean number of top images selected between the coached malingerers' group and uncoached malingerers' group, and to compare the mean number of incorrect images selected between the coached malingerers' group and uncoached malingerers' group. The 118 participants in the uncoached malingering group (M = 37.36, SD = 18.07) compared to the 136 in the coached malingering group (M = 25.13, SD = 11.06) selected a significantly higher number of top images during the N-Tri recognition trials, $t(188.225) = -6.39$, $p < .001$, d

= .82. The 118 participants in the uncoached malingering group (M = 63.94, SD = 23.61) compared to the 136 in the coached malingering group (M = 45.47, SD = 17.01) chose a significantly higher number of incorrect answers during the N-Tri recognition trials, $t(209.303) = -7.06$, $p < .001$, $d = .90$. Therefore, these results support the hypotheses and suggest that coached and uncoached malingering groups understood and followed instructions and that including the additional information into the coached malingerers' scenario was influential to responding.

CHAPTER IV

DISCUSSION

The goal of the present study was to validate the N-Tri for the detection of malingering. To this end, it was hypothesized that the N-Tri would positively correlate highly with the PDRT and RDS. This hypothesis was supported by Pearson correlations demonstrating that the N-Tri has strong positive correlations with the PDRT and RDS. This hypothesis was also supported by stepwise multiple regressions displaying high multiple regression coefficients. These results demonstrate high convergent validity, and thus, provide support for construct validity and the N-Tri being a measure of response bias since both the PDRT and RDS have been shown to accurately identify response bias (Vickery et al., 2001; Greve et al., 2007; Mathias et al., 2002).

Additional support for the validation of the N-Tri comes from hypothesis two, which demonstrated that the N-Tri can detect both coached and uncoached malingerers with high levels of sensitivity and specificity. Therefore, the N-Tri is capable of identifying a high percentage of examinees as malingering and is also capable of identifying a high number of examinees as not malingering. Furthermore, hypothesis three demonstrated that the N-Tri has higher sensitivity when detecting coached malingerers then the PDRT and RDS, which suggests that the N-Tri is able to detect a higher percentage of coached malingerers than both the PDRT and RDS. Thus, it is appropriate to say that based off of this study, the N-Tri is better at identifying coached malingerers than the PDRT and RDS.

Hypothesis four demonstrated that scores on all trials and total score of the N-Tri were not impacted by age or education. This implies that regardless of age or education, examinees can still perform well on the N-Tri. Additionally, these findings suggest that priming does not decrease in older adults since older adults performed just as well on the N-Tri as younger adults. Thus, these results align with and support those of Howard et al. (1981), Fleischman and Gabrieli (1998), and Fleischman (2007), who also found priming in older adults to not be significantly impaired.

Hypotheses five and six demonstrated that when compared to the coached malingerers' group, the uncoached malingerers' group chose a higher number of top images, as well as more overall incorrect answers on the N-Tri recognition trials. This was expected since the uncoached malingerers' group was not given a test-taking strategy, and therefore, was not told that "major exaggerations" and "remembering absolutely nothing" are not realistic. Therefore, when compared to the coached malingerers' group, they should have given more answers that were far from being the correct answer, such as by selecting the top image instead of selecting the incorrect bottom image. Thus, the uncoached malingerers' group should have selected more top images than that coached malingerers' group, which they did. Additionally, the uncoached malingerers' group should have also been getting more incorrect answers overall compared to the coached malingerers' group, which they also did. This suggests that the coached and uncoached malingering groups understood and followed instructions, and that including the additional information into the coached malingerers' scenario was influential to responding.

Taken together, this study looked to validate the N-Tri for the detection of response bias. The results of this study suggest that the N-Tri has convergent validity, which provides support for construct validity and the N-Tri being a measure of response bias. It was also suggested that the N-Tri can identify high percentages of malingerers overall, and when compared to the PDRT and RDS, is better at detecting coached malingerers specifically. Lastly, regarding age and education, this study suggested that anyone regardless of their level of education or age can perform well on the N-Tri.

Clinical Implications

Having an increased ability to detect malingerers, such as with the use of the N-Tri, would be beneficial for many fields, including medicine. Thirteen percent of clients who go to an emergency room for mental health related symptoms are feigning impairment (Yates et al., 1996). Not only does this waste emergency room workers' time, but it is also expensive for insurance companies. In Texas, it has been found that individuals who malinger cost insurance companies $150 billion every year, which increases insurance costs per family by $1800 (Texas Department of Insurance: Consumer). Furthermore, most psychiatric inpatient facilities have a limited number of rooms and beds. It is frequently the case that patients with genuine mental illnesses are unable to be admitted to inpatient care facilities and instead spend days in an emergency department because individuals who are malingering are blocking patients who are actually ill from receiving care (Garriga, 2007). If the individuals who are working in emergency departments could accurately identify patients who are feigning symptoms, both coached and uncoached, not only would their time not be wasted, but it could save insurance companies money and help individuals with true mental illnesses receive care faster.

Additionally, in fields such as psychology, there is a need for well-validated measures of malingering. Further, there is a need for malingering measures that accurately detect coached individuals specifically. If clinicians had more confidence in the ability of the response bias tests they use to detect coached malingerers, then they could be more confident in the diagnoses and treatment recommendations made off any given patient's assessment results (Bush et al., 2005). In addition, psychologists being able to accurately detect coached and uncoached malingerers could save the Social Security Administration (SSA) billions of dollars a year. It was found that in 2011, the malingering of mental disorders in Social Security Disability examinations cost the SSA $20.02 billion (Chafetz & Underhill, 2013).

The results of this study suggest that the N-Tri could be one of these well-validated measures that isn't as vulnerable to coaching as current tests. When used in conjunction with additional response bias tests, the N-Tri could help make the detection of exaggerated symptoms more accurate and identify more malingerers. Specifically, the N-Tri could help detect some of the coached malingerers that typically aren't detected using existing malingering tests since this study demonstrated that when compared to the PDRT and the RDS, the N-Tri had higher sensitivity when detecting coached malingerers. Incorporating the N-Tri into the fields of psychology and medicine could potentially help patients with mental health problems receive adequate care quicker and ensure that they are receiving diagnoses and treatment that are accurate and would benefit them. It could also help save insurance companies and the SSA large amounts of money.

In order to incorporate the N-Tri into clinical practice, it first must be extensively tested beyond the current study. It must be shown that the N-Tri possesses high rates of validity and reliability, and that the results of this study stand for more diverse populations as well as with in-person administration. Norms for the N-Tri must be established as well. If the time comes when the N-Tri is ready for clinical use, it is suggested to be used in conjunction with other malingering assessments in order to ensure that outcomes are consistent across multiple measures.

Limitations

This study, as with all studies, had several limitations. First, online surveys were used instead of administering the assessments in person. Thus, there was no way of knowing if the participants understood the instructions, responded conscientiously, experienced any technical difficulties, or if testing environments were free from distractions. Second, adaptations of the N-Tri, RDS, and PDRT were used. The results may not replicate with in-person administration of the non-adaptation versions. Third, the malingering groups were composed of participants who were instructed to feign cognitive impairment. It is possible that some participants simulated response bias more realistically than others and were more representative of true response bias in the larger population, so these results may not apply to or be generalizable to the larger population. Next, the sample used in this study was largely white, female, and non-Hispanic/Latino. Therefore, the results may not be representative of populations that aren't white, female, and non-Hispanic/Latino. Lastly, only neurotypical participants were used. It is proposed that the N-Tri be further validated by use of a sample that includes clinical populations in order to demonstrate that scores on the N-Tri are not influenced by the presence of actual impairments.

Future Research

As was suggested, future studies that examine performance on the N-Tri for clinical populations in order to ensure that scores are not influenced by genuine impairments should be conducted. Studies that aim to assess the N-Tri's reliability should also be considered. Additional simulation studies to further assess the N-Tri's validity especially with more diverse samples would be useful, as would studies that aim to create norms for the N-Tri. It would also be useful to have studies that aim to validate the naming portion of the N-Tri, such as by correlating it to other naming tests, in order to demonstrate that not only the response bias part of the N-Tri is valid, but the naming part is valid as well. The reliability of the N-Tri naming portion should also be examined. In conclusion, the N-Tri can be a useful measure for the detection of response bias in malingerers that did and did not receive coaching, but research that further examines its validity and reliability are needed.

References

Abbenhuis, M. A., Raaijmakers, W. G. M., Raaijmakers, J. G. W., & Van Woerden G. J. M. (1990). Episodic memory in dementia of the Alzheimer type and in normal ageing: similar impairment in automatic processing. *The Quarterly Journal of Experimental Psychology, 42*, 569–583.

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB 97 manual for the Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

Allen, L. M., & Green, P. (2001). Declining CARB failure rates over six years of testing: what's wrong with this picture? *Archives of Clinical Neuropsychology, 16*, 846.

Ali, S., Jabeen, S., & Alam, F. (2015). Multimodal approach to identifying malingering posttraumatic stress disorder: a review. *Innovative Clinical Neuroscience, 13*(1-2), 12-20.

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

Bauer, L., & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: is test security threatened? *Archives of Clinical Neuropsychology, 21*(1), 121-126.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2013). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science, 58*(3), 1-15

Berry, D. T. R., & Schipper, L. J. (2007). Detection of feigned psychiatric symptoms during forensic neuropsychological evaluations. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 226–263). New York, NY: Oxford University Press.

Bird, C. M. (2017). The role of the hippocampus in recognition memory. *Cortex, 93*, 155-165.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 175-181

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object detail. *PNAS, 105*(38), 14325-14329.

Burkett, B. G., & Whitley, G. (1998). *Stolen Valor: how the Vietnam generation was robbed of its heroes and its history.* Dallas: Verity Press.

Bush, S. S., Ruff, R. M., Tröster, 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 : the official journal of the National Academy of Neuropsychologists*, *20*(4), 419–426.

Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press

Camina, E., & Guell, F. (2017). The neuroanatomical, neurophysiological and psychological basis of memory: current models and their origins. *Frontiers in Pharmacology, 8*, 438.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, *28*(7), 633–639.

Chertkow, H., & Bub, D. (1989). Priming and semantic memory loss in Alzheimer's disease. *Brain and Language, 36*, 420-446.

Chiarello, C., & Hoyer, W. J. (1988). Adult age differences in implicit and explicit memory: time course and encoding effects. *Psychology and Aging, 3*, 358–366.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). A MMPI handbook. Vol. 1: Clinical interpretation (rev. ed.). Minneapolis: University of Minnesota Press.

Davis, H. P., Cohen, A., Gandy, M., Colombo, P., Van Dusseldorp, G., Simolke, N., & Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience, 104*, 288–297.

Dickerson, B. C., & Eichenbaum, H. (2010). The episodic memory system: neurocircuitry and disorders. *Neuropharmacology, 35*, 86-104.

Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: effects of coaching and information. *Archives of Clinical Psychology, 18*, 121-134.

Ellis, A. W., Flude, B. M., Young, A., & Burton, A. M. (1996). Two loci of repetition priming in the recognition of familiar faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 295–308.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters, 27*, 861-874.

Fleischman, D. A. (2007). Repetition priming in ageing and Alzheimer's disease: an integrative review and future directions. *Cortex, 43*, 889–897.

Fleischman, D. A., and Gabrieli, J. D. E. (1998). Repetition priming in normal ageing and Alzheimer's disease: a review of findings and theories. *Psychology and Aging, 13*, 88–119.

Frederick, R. I. (1997). *Validity indicator profile manual.* Minnetonka, MN: NCS Assessments.

Freed, D. M., Corkin, S., Growdon, J. H., & Nissen, M. J. (1989). Selective attention in Alzheimer's disease: characterizing cognitive subgroups of patients. *Neuropsychologia, 27*(3), 325-339.

Frueh, B. C., Elhai, J. D., Grubaugh, A. L., Monnier, J., Kashdan, T., Sauvageot, J. A,. Hamner, M. B., Burkett, B. G., & Arana, G. W. (2005). Documented combat exposure of US veterans seeking treatment for combat-related post-traumatic stress disorder. *British Journal of Psychiatry, 186,* 467-472.

Garriga, M. (2007, March 1). *Malingering in the clinical setting.* Psychiatric Times.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15,* 594-615.

Gouvier, W. D., Prestholdt, P. H., & Warner, M. S. (1988). A survey of common misconceptions about head injury recovery. *Archives of Clinical Neuropsychology, 3*, 331-343.

Green, P., Allen, L., & Astner, K. (1996). Manual for Computerised Word Memory Test. Durham, NC: CogniSyst.

Greve, K. W., & Bianchini, K. J. (2006). Classification accuracy of the Portland Digit Recognition Test in traumatic brain injury: results of a known-group analysis. *The Clinical Neuropsychologist, 20*(4), 816-830.

Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., … Ciota, M. A. (2007). Malingering in toxic exposure: classification accuracy of Reliable Digit Span and WAIS-III Digit Span scaled scores. *Assessment, 14*(1), 12-21.

Gutheil, T. G. (2003). Reflections on coaching by attorneys. *Journal of the American Academy of Psychiatry and the Law, 31*, 6-9.

Hickling, E. J., Taylor, A. E., Blanchard, E. B., & Devineni, T. (1999). *Simulation of motor vehicle accident related PTSD: Effects of coaching with DSM-IV criteria.* In E. J. Hickling, & E. B. Blanchard (Eds.), The international handbook of road traffic accidents and psychological trauma: Current understanding, treatment, and law (pp. 305 – 320). New York: Elsevier.

Horton, A. M. (2000). Book and Test Review. *Archives of Clinical Neuropsychology, 15*(7), 649-651.

Howard, D. V., McAndrews, M. P., & Lasaga, M. I. (1981). Semantic priming of lexical decisions in young and old adults. *Journal of Gerontology, 36*(6), 707-714.

Hultsch, D. F., Masson, M. E. J., & Small, B. J. (1991). Adult age differences in direct and indirect tests of memory. *Journal of Gerontology, 46*, 22–30.

Jasinski, L. J., Berry, D. T. R., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: a meta-analytic review. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 300-314.

Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of Human Neuropsychology*. New York, NY: Worth Publishers.

Kopelman, M. D. (1985). Rates of forgetting in Alzheimer's- type dementia and Korsakoff's syndrome. *Neuropsychologia, 23*(5), 623-638.

Lees-Haley, P. R. (1997). Attorneys influence expert evidence in forensic psychological and neurological cases. *Assessment, 4*(4), 321-324.

Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron, 37*, 171-180.

Martone, M., Butters, N., & Trauner, D. (1986). Some analyses of forgetting of pictorial material in amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 8*(3), 161-178.

Millis, S.R., & Volinsky, C.T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology, 23* (6), 809-828.

Mittenberg, W., Azrin, R, Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale – Revised. *Psychological Assessment, 5*(1), 34-40.

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.

Mutter, S. A., Howard, D. V., Howard, J. H., & Wiggs, C. L. (1990). Performance on direct and indirect tests of memory after mild closed head injury. *Cognitive Neuropsychology, 7*(4), 329-346.

Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science, 11*(2), 58.

Park, D. C., Puglisi, J. T., & Smith, A. D. (1986). Memory for pictures: does an age-related decline exist? *Psychology and Aging, 1*(1), 11-17.

Park, D. C., Royal, D., Dudley, W., & Morrell, R. (1988). Forgetting of pictures over a long retention interval in young and older adults. *Psychology and Aging, 3*(1), 94-95.

Platt, J. J., & Husband, S. D. (1986). Post-traumatic stress disorder in forensic practice. *American Journal of Forensic Psychology, 4*(1), 29-56.

Resnick, P. J. (1997). *Malingering of posttraumatic disorders.* In R. Rogers (Ed.), Clinical assessment of malingering and deception (2nd ed.) (pp. 130–152). New York, NY: Guilford Press.

Rose, F. E., Hall, S., Szalda-Petree, A. D., & Bach, P. J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13*(4), 349-363.

Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: misuse of the internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology: Research and Practice, 33*(3), 294-299.

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment, 19*(1), 21-30.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior, 6*(1), 156-163.

Shimamura, A. P., Salmon, D. P., Squire, L. R., & Butters, N. (1987). Memory dysfunction and word priming in dementia and amnesia. *Behavioral Neuroscience, 101*(3), 347-351.

Signoret, J. L. (1979). *Batterie d'Efficience Mnésique BEM 144* [Memory Performance Battery BEM 144]. Unpublished psychological test battery. La Salpetrière, Paris.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). The Victoria symptom validity test. *Odessa, FL: Psychological Assessment Resources.*

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545-561.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73-74.

Texas Department of Insurance: Consumer. (2006, August). *Insurance fraud.*

Tombaugh, T. (1997). The test of memory malingering (TOMM): normative data from cognitively impaired individuals. *Psychological Assessment, 9*(3), 260-268.

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, liabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*, 1-7.

Vakil, E., & Oded, Y. (2003). Comparison between three memory tests: cued recall, priming and saving closed-head injured patients and controls. *Journal of Clinical and Experimental Neuropsychology, 25*(2), 274-282.

Vakil, E., Biederman, Y., Liran, G., Groswasser, Z., & Aberbuch, S. (1994). Head injured patients and control groups: implicit vs. explicit measures of frequency judgment. *Journal of Clinical and Experimental Neuropsychology, 16*, 539-546.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16*, 45 –73.

Walczyk, J. J., Sewell, N., & DiBenedetto, M. B. (2018). A review of approaches to detecting malingering in forensic contexts and promising cognitive load-inducing lie detection techniques. *Frontiers in Psychiatry, 9*, 700.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology, 73*(1), 117-129.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised manual*. New York, NY: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale III: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *WAIS–IV: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: a survey of attorneys' and law students' attitudes. Professional Psychology: *Research and Practice, 26*(5), 474-477.

Winograd, E., Smith, A. D., & Simon, E. W. (1982). Aging and the picture superiority effect in recall. *Journal of Gerontology, 37*(1), 70-75.

Yates, B. D., Nordquist, C. R., & Schultz-Ross, R. A. (1996). Feigned psychiatric symptoms in the emergency room. *Psychiatric services (Washington, D.C.)*, *47*(9), 998–1000.

Yerkes, R. M. (1921). *Psychological examining in the United States Army.* Washington, DC: Government Printing Office.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.

Appendix A: Eligibility Criteria

Eligibility Criteria

Ineligible if answers to "yes" on any of the following:

Criterion	**Yes**	**No**
Diagnosed with brain injury/concussion		
Neurological Disease		
Diagnosed severe psychiatric disorder (Excluding depression and anxiety)		
Central nervous system disease (e.g. epilepsy)		
Stroke		
Developmental disorder (e.g., autism spectrum disorder)		
Learning disability		
Brain Tumor		
Brain Surgery		

Appendix B: Manipulation Check for Malingering Groups

Manipulation Check

Please recall below any instructions from the scenario you read at the beginning of the session and that you were asked to follow throughout testing. Be sure to include what your goal was during the testing and any related details. Please write as many as you can remember.

On a scale from 1 to 5, during the tests you completed today, how much did you try to follow the instructions from the scenario you read at the beginning of the session?

(Please circle your response)

1	2	3	4	5
Didn't try at all				Tried very hard

Appendix C: N-Tri Response Form (Trial 1)

N-Tri Response Form – Trial One

Trial 1, Part 1: Naming

Part 1:Naming

Image	**Correct without cue**	**With semantic cue** Correct	incorrect	**With phonemic cue** correct	incorrect	**Points**
Butterfly **(An insect)**	______	______	______	______	______	0 1 2 3
Glove **(Item of clothing)**	______	______	______	______	______	0 1 2 3
Lightbulb **(Produces light)**	______	______	______	______	______	0 1 2 3
Watch **(Use it to tell time)**	______	______	______	______	______	0 1 2 3
Candle **(Produces light)**	______	______	______	______	______	0 1 2 3
Clown **(Seen at the circus)**	______	______	______	______	______	0 1 2 3
Kite **(A toy)**	______	______	______	______	______	0 1 2 3
Rainbow **(Produced by nature)**	______	______	______	______	______	0 1 2 3
See-Saw/ Teeter- totter **(Playground equipment)**	______	______	______	______	______	0 1 2 3
Flashlight **(Used to see)**	______	______	______	______	______	0 1 2 3
Witch **(Halloween Character)**	______	______	______	______	______	0 1 2 3
Peacock **(An animal)**	______	______	______	______	______	0 1 2 3

Bird Cage **(A container)**	______ ______ ______ ______ ______	0 1 2 3
Bird's Nest **(Where eggs are kept)**	______ ______ ______ ______ ______	0 1 2 3
Electrical Plug **(Used with an outlet)**	______ ______ ______ ______ ______	0 1 2 3
Wig **(Covering for the head)**	______ ______ ______ ______ ______	0 1 2 3
Snail **(An animal)**	______ ______ ______ ______ ______	0 1 2 3
Whale **(An animal)**	______ ______ ______ ______ ______	0 1 2 3
Screw **(Used for building)**	______ ______ ______ ______ ______	0 1 2 3
Scarf **(A piece of clothing)**	______ ______ ______ ______ ______	0 1 2 3
Well **(Used to access water)**	______ ______ ______ ______ ______	0 1 2 3
Dustpan **(Used to clean)**	______ ______ ______ ______ ______	0 1 2 3
Parachute **(Used for skydiving)**	______ ______ ______ ______ ______	0 1 2 3
Blinds **(Used to block light)**	______ ______ ______ ______ ______	0 1 2 3
Funnel **(Used for pouring)**	______ ______ ______ ______ ______	0 1 2 3
Hinge **(Part of a door)**	______ ______ ______ ______ ______	0 1 2 3
Gauge	______ ______ ______ ______ ______	

(Used for measurement)		0 1 2 3
Porthole **(Type of window)**	______ ______ ______ ______ ______	0 1 2 3
Anvil **(Used to shape metal)**	______ ______ ______ ______ ______	0 1 2 3
Mortar & Pestle **(Used for grinding)**	______ ______ ______ ______ ______	0 1 2 3
Axle **(Allows wheels to spin)**	______ ______ ______ ______ ______	0 1 2 3
		Total: - _________

Number Correct without cue: _____________

Number of semantic cues given: ___________

Number correct with semantic cues: __________

Number of phonemic cues given: __________

Number correct with phonemic cues given:______

Total number correct (with or without cues): __________

0 = incorrect

1 = correct with phonemic cue

2 = correct with semantic cue

3 = correct without a cue

Trial 1, Part 2: Recognition

Target Image	Answers (Circle Answer Given)			Points Earned (1pt per correct answer)	
Axel	1	**<u>2</u>**	3	1pt	0pts
Wig	1	**<u>2</u>**	3	1pt	0pts
Witch	1	**<u>2</u>**	3	1pt	0pts
Blinds	1	**<u>2</u>**	3	1pt	0pts
Porthole	1	2	**<u>3</u>**	1pt	0pts
Snail	1	2	**<u>3</u>**	1pt	0pts
Candle	1	**<u>2</u>**	3	1pt	0pts
Glove	1	**<u>2</u>**	3	1pt	0pts
Bird's Nest	1	**<u>2</u>**	3	1pt	0pts
Well	1	2	**<u>3</u>**	1pt	0pts
Gauge	1	**<u>2</u>**	3	1pt	0pts
Flashlight	1	**<u>2</u>**	3	1pt	0pts
Parachute	1	2	**<u>3</u>**	1pt	0pts
Watch	1	**<u>2</u>**	3	1pt	0pts
Rainbow	1	2	**<u>3</u>**	1pt	0pts

Electrical Plug	1	2	**3**	1pt	0pts
Peacock	1	2	**3**	1pt	0pts
Whale	1	**2**	3	1pt	0pts
See-Saw/ Teeter-Totter	1	**2**	3	1pt	0pts
Anvil	1	**2**	3	1pt	0pts
Butterfly	1	**2**	3	1pt	0pts
Bird Cage	1	2	**3**	1pt	0pts
Funnel	1	2	**3**	1pt	0pts
Dustpan	1	**2**	3	1pt	0pts
Lightbulb	1	**2**	3	1pt	0pts
Clown	1	**2**	3	1pt	0pts
Scarf	1	2	**3**	1pt	0pts
Mortar & Pestle	1	2	**3**	1pt	0pts
Hinge	1	2	**3**	1pt	0pts
Screw	1	2	**3**	1pt	0pts
Kite	1	2	**3**	1pt	0pts

	Total: _______________

Appendix D: N-Tri Response Form (Trial 2)

Trial 2, Part 1: Study Phase (No Response Form)

Trial 2, Part 2: Recognition

Target Image	**Response (Circle Answer Given)**			**Points Earned (1pt per correct answer)**	
Screw	1	**2**	3	1pt	0pts
Lightbulb	1	**2**	3	1pt	0pts
Butterfly	1	**2**	3	1pt	0pts
Clown	1	**2**	3	1pt	0pts
Hinge	1	2	**3**	1pt	0pts
Kite	1	**2**	3	1pt	0pts
Rainbow	1	2	**3**	1pt	0pts
Witch	1	**2**	3	1pt	0pts
Glove	1	2	**3**	1pt	0pts
Mortar & Pestle	1	**2**	3	1pt	0pts
Porthole	1	2	**3**	1pt	0pts
See-Saw/Teeter-Totter	1	**2**	3	1pt	0pts
Flashlight	1	2	**3**	1pt	0pts

Snail	1 **<u>2</u>** 3		1pt	0pts
Whale	1 2 **<u>3</u>**		1pt	0pts
Electrical Plug	1 2 **<u>3</u>**		1pt	0pts
Pressure Gauge	1 **<u>2</u>** 3		1pt	0pts
Scarf	1 **<u>2</u>** 3		1pt	0pts
Bird's Nest	1 2 **<u>3</u>**		1pt	0pts
Well	1 **<u>2</u>** 3		1pt	0pts
Bird Cage	1 **<u>2</u>** 3		1pt	0pts
Anvil	1 **<u>2</u>** 3		1pt	0pts
Dustpan	1 **<u>2</u>** 3		1pt	0pts
Peacock	1 **<u>2</u>** 3		1pt	0pts
Parachute	1 2 **<u>3</u>**		1pt	0pts
Wig	1 2 **<u>3</u>**		1pt	0pts
Funnel	1 **<u>2</u>** 3		1pt	0pts
Axle	1 2 **<u>3</u>**		1pt	0pts
Blinds	1 **<u>2</u>** 3		1pt	0pts

Candle	1	**2**	3		1pt	0pts
Watch	1	2	**3**		1pt	0pts
					Total Score: ____________	

Appendix E: N-Tri Response Form (Trial 3)

Trial 3: Retention Trial

Target Image	Response			Points (1pt per correct answer)	
Scarf	1	2	**<u>3</u>**	1pt	0pts
Peacock	1	2	**<u>3</u>**	1pt	0pts
Candle	1	2	**<u>3</u>**	1pt	0pts
Lightbulb	1	2	**<u>3</u>**	1pt	0pts
Rainbow	1	**<u>2</u>**	3	1pt	0pts
Funnel	1	**<u>2</u>**	3	1pt	0pts
Bird Cage	1	**<u>2</u>**	3	1pt	0pts
Screw	1	**<u>2</u>**	3	1pt	0pts
Blind	1	**<u>2</u>**	3	1pt	0pts
Clowns	1	**<u>2</u>**	3	1pt	0pts
Witch	1	2	**<u>3</u>**	1pt	0pts
Well	1	2	**<u>3</u>**	1pt	0pts
Butterfly	1	**<u>2</u>**	3	1pt	0pts

Wig	1	__**2**__	3	1pt	0pts
Dustpan	1	2	__**3**__	1pt	0pts
Kite	1	__**2**__	3	1pt	0pts
Mortar & Pestle	1	2	__**3**__	1pt	0pts
See-Saw/ Teeter-Totter	1	2	__**3**__	1pt	0pts
Flashlight	1	__**2**__	3	1pt	0pts
Hinge	1	__**2**__	3	1pt	0pts
Pressure Gauge	1	__**2**__	3	1pt	0pts
Whale	1	2	__**3**__	1pt	0pts
Axel	1	__**2**__	3	1pt	0pts
Parachute	1	__**2**__	3	1pt	0pts
Watch	1	2	__**3**__	1pt	0pts
Anvil	1	2	__**3**__	1pt	0pts
Snail	1	__**2**__	3	1pt	0pts

Electrical Plug	1	**<u>2</u>**	3	1pt	0pts
Bird's Nest	1	**<u>2</u>**	3	1pt	0pts
Porthole	1	2	**<u>3</u>**	1pt	0pts
Glove	1	**<u>2</u>**	3	1pt	0pts
				Total: _______________	

Appendix F: Reliable Digit Span

Reliable Digit Span

Instructions: You will press the play button on each page and listen to strings of numbers. Then, you will be asked to recall the numbers in the **same order** that you heard them.

1. 9-7
2. 6-3
3. 5-8-2
4. 6-9-4
5. 7-2-8-6
6. 6-4-3-9
7. 4-2-7-3-1
8. 7-5-8-3-6
9. 3-9-2-4-8-7
10. 6-1-9-4-7-3
11. 4-1-7-9-3-8-6
12. 6-9-1-7-4-2-8
13. 3-8-2-9-6-1-7-4
14. 5-8-1-3-2-6-4-7
15. 2-7-5-8-6-3-1-9-4
16. 7-1-3-9-4-2-5-6-8

Instructions: You will listen to strings of digits. Then, you will be asked to recall the numbers in the **reverse order** that you heard them.
Example: If you hear 7-1, you would type 1-7.
If you hear 3-4, you would type 4-3.

1. 3-1
2. 2-4
3. 4-6
4. 5-7
5. 6-2-9
6. 7-4-5
7. 8-2-7-9
8. 4-9-6-8
9. 6-5-8-4-3
10. 1-5-4-8-6
11. 5-3-7-4-1-8
12. 7-2-4-8-5-6
13. 8-1-4-9-3-6-2
14. 4-7-3-9-6-2-8
15. 9-4-3-7-6-2-1-8
16. 7-2-8-1-5-6-4-3

Appendix G: Portland Digit Recognition Test

Portland Digit Recognition Test

Instructions: (1) You will listen to a string of digits, (2) you will be asked to count backwards out loud, and (3) you will be asked to select which string of digits you were shown from two options.

Block 1	Block 2
1. 78723 or **16920**	1. **16920** or 67745
2. **84288** or 24812	2. 59680 or **84288**
3. 81397 or **42602**	3. **42602** or 52196
4. 82883 or **03206**	4. **03206** or 49254
5. 30628 or **99727**	5. **99727** or 04281
6. **30945** or 19190	6. 99940 or **30945**
7. 06472 or **63015**	7. **63015** or 10567
8. **15241** or 06723	8. 30064 or **15241**
9. **77581** or 00354	9. 07732 or **77581**
10. 64185 or **09028**	10. **09028** or 64982
11. **55343** or 72009	11. 61953 or **55343**
12. 42939 or **71378**	12. **71378** or 94621
13. **29614** or 80604	13. 42189 or **29614**
14. 54987 or **00366**	14. **00366** or 22604
15. 14276 or **60138**	15. **60138** or 03404
16. **24108** or 88004	16. 81929 or **24108**
17. **52237** or 87787	17. 26804 or **52237**
18. **24819** or 97594	18. 96588 or **24819**