Assessment of Performance Validity During Neuropsychological Evaluation in Patients with Epilepsy

Nichelle Rothong
Marquette University

ASSESSMENT OF PERFORMANCE VALIDITY DURING NEUROPSYCHOLOGICAL EVALUATION IN PATIENTS WITH EPILEPSY

by

Nichelle D. Rothong, MS

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

Milwaukee, Wisconsin

August 2014

ABSTRACT
ASSESSMENT OF PERFORMANCE VALIDITY DURING
NEUROPSYCHOLOGICAL EVALUATION
IN PATIENTS WITH EPILEPSY

Nichelle D. Rothong, MS

Marquette University, 2014

Patients with epilepsy are considered a motivated population without clear incentive to perform suboptimally on neuropsychological testing. However, in the limited research exploring performance validity testing (PVT) in patients with epilepsy, the base rate of suboptimal performance has ranged from 4 (Hill, Ryan, Kennedy, & Malamut, 2003) to 28% (Loring, Lee, & Meador, 2005). These findings are concerning, as suboptimal PVT scores have been found to be associated with significantly lower neuropsychological performance across most cognitive domains (e.g., Green, Rohling, Lees-Haley, & Allen, 2001).

One possible explanation for the variance in base rate of suboptimal performance is the significant cognitive impairment commonly associated with epilepsy (Bortz, 2003). The present study investigated this unexplored theory by utilizing the Word Memory Test (WMT). The WMT is a PVT that indicates whether scores below failure cutoff likely reflect suboptimal performance or significant cognitive impairment, a determination made by General Memory Impairment Profile (GMIP) analysis. Using WMT normative cutoffs, patients in the current study were categorized into optimal, suboptimal, and GMIP performance groups. Subsequently, differences among groups on a variety of neuropsychological measures were explored. The validity of the GMIP was also examined to provide support for its use with this population.

Findings indicated that 43% of the sample fell into the WMT optimal group, 36% into the suboptimal group, and 21% into the GMIP group. Although WMT performance accounted for 29% of the variance in overall neuropsychological performance, PV did not impact all cognitive domains equally. WMT performance groups scored significantly differently across most neuropsychological measures; patients in the suboptimal and GMIP groups typically obtained significantly lower scores than patients in the optimal group. Results also largely supported the validity of the GMIP in its ability to identify WMT scores below failure cutoff due to borderline memory impairment. Overall, current findings encourage the use and further investigation of the WMT and GMIP analysis in patients with epilepsy.

TABLE OF CONTENTS

LIST OF TABLES iv

CHAPTER

I. INTRODUCTION 1

Background Context 1

Statement of the Problem 3

Purpose of the Study 5

Implications of the Study 6

Research Questions 8

Definition of Terms 8

II. LITERATURE REVIEW 11

Overview of Performance Validity Testing 11

History 11

Methods for Scoring PVTs 14

Neuropsychologists' Current Practices in Assessing PV 16

Types of PVTs 16

Non-Forced-Choice PVTs 16

Forced-Choice PVTs 17

Critical Nature of PVT within Neuropsychological Assessment 19

PVT in Patients with Epilepsy 21

Relationship between PVT Scores and Neuropsychological Test Scores 25

Conclusion 26

III. METHODOLOGY 29

Participants 29

Eligibility Criteria 29

Measures 29

Research Design 57

Consent for Research Participation 57

Human Subjects Consideration 58

Procedures 58

Power Analysis 62

Exploration of Covariates 67

Data Analysis 68

Sample Size 75

Data Screening 76

Accuracy of the Data File 76

Missing and Problematic Data 77

IV. RESULTS 79

Preliminary Analyses 79

Exploring Assumptions of Parametric Tests 79

Approach to Dealing with Non-Normality and Unequal Variances 81

Sample Characteristics 81

Sample Demographics 81

Epilepsy Characteristics 83

Statistical Analyses 84

Research Question 1 84

Research Question 2 86

Research Question 3 92

Research Question 4 126

Research Question 5 131

V. DISCUSSION 144

Base Rates of Optimal, Suboptimal, and GMIP Performance on the WMT 144

WMT Performance and Neuropsychological Test Scores 151

Validity of the GMIP 162

Limitations and Directions for Future Research 170

Conclusion 177

REFERENCES 181

LIST OF TABLES

Table 1 WMT Performance Group Criteria.....60

Table 2 Dummy Coding.....71

Table 3 Sample Demographics.....82

Table 4 Optimal, Suboptimal, and GMIP Groups' Performance on WMT.....85

Table 5 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and VF Scores.....101

Table 6 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and VF Scores.....102

Table 7 Summary of Hierarchical Regression Analysis for Differences in VF Scores Among WMT Performance Groups Accounting for Education Level.....103

Table 8 WMT Performance Groups' Mean VF Scores.....104

Table 9 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and PO Scores.....105

Table 10 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and PO Scores.....106

Table 11 Summary of Hierarchical Regression Analysis for Differences in PO Scores Among WMT Performance Groups Accounting for Education Level.....107

Table 12 WMT Performance Groups' Mean PO Scores.....107

Table 13 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and EF Scores.....108

Table 14 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and EF Scores.....109

Table 15 Summary of Regression Analysis for Differences in EF Scores Among WMT Performance Groups.....110

Table 16 WMT Performance Groups' Mean EF Scores.....110

Table 17 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and LM Scores.....111

Table 18 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and LM Scores....112

Table 19 Summary of Regression Analysis for Differences in LM Scores Among WMT Performance Groups....112

Table 20 WMT Performance Groups' Mean LM Scores....113

Table 21 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and AC Scores....114

Table 22 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and AC Scores....115

Table 23 Summary of Hierarchical Regression Analysis for Differences in AC Scores Among WMT Performance Groups Accounting for Education Level....116

Table 24 WMT Performance Groups' Mean AC Scores....116

Table 25 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and PS Scores....117

Table 26 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and PS Scores....118

Table 27 Summary of Hierarchical Regression Analysis for Differences in PS Scores Among WMT Performance Groups Accounting for Education Level....119

Table 28 WMT Performance Groups' Mean PS Scores....119

Table 29 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and MD Scores....120

Table 30 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and MD Scores....121

Table 31 Summary of Regression Analysis for Differences in MD Scores Among WMT Performance Groups....121

Table 32 WMT Performance Groups' Mean MD Scores....122

Table 33 Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and DTBM Scores....123

Table 34 Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and DTBM Scores....124

Table 35 Summary of Hierarchical Regression Analysis for Differences in DTBM Scores Among WMT Performance Groups Accounting for Education Level.....125

Table 36 WMT Performance Groups' Mean DTBM Scores.....125

Table 37 Correlation Coefficients, Means, and Standard Deviations for GMIP Scores and LM Scores.....129

Table 38 Summary of Regression Analysis of GMIP Scores Predicting LM Scores.....129

Table 39 Correlation Coefficients, Means, and Standard Deviations for GMIP Performance Dummy Variable and LM Scores.....130

Table 40 Summary of Regression Analysis for Differences in LM Scores Among GMIP Performance Groups 131

Table 41 GMIP Performance Groups' Mean LM Scores 131

Table 42 Collinearity Diagnostics for Hierarchical Regression with WMT Subtests as Predictors 133

Table 43 Collinearity Diagnostics for Hierarchical Regression with WMT Composites as Predictors 133

Table 44 Correlation Coefficients, Means, and Standard Deviations for WMT Subtest Scores and GMIP Scores 137

Table 45 Partial and Semipartial Correlation Coefficients for WMT Subtests and GMIP Scores 138

Table 46 Summary of Hierarchical Regression Analysis of the Relationship between WMT Subtests and GMIP Scores.....140

Table 47 Correlation Coefficients, Means, and Standard Deviations for WMT Composites and GMIP Scores 141

Table 48 Partial and Semipartial Correlation Coefficients for WMT Composites and GMIP Scores.....142

Table 49 Summary of Hierarchical Regression Analysis Examining the Relationship between WMT Composites and GMIP Scores 143

CHAPTER I: INTRODUCTION

Background Context

During neuropsychological assessment, neuropsychologists must examine the validity of test performance, determining whether observed low test scores are due to neurological illness or injury, or instead reflective of suboptimal performance (Binder, 1990; Binder, 1993; Green, 2001; Slick, Sherman, & Iverson, 1999). Determination of performance validity (PV) is an important part of neuropsychological assessment as results are used to assist in formulating diagnoses, impressions, and recommendations. Neuropsychological test results may also be used as the main source of evidence to support or refute claims for financial compensation (Guilmette, Hart, & Guiliano, 1993; Slick et al., 1999).

Stand-alone forced-choice performance validity tests (PVTs; Larrabee, 2012) have been found to be the most sensitive and specific method for determining the validity of test scores during neuropsychological evaluation (Heilbronner et al., 2009; Vickery, Berry, Inman, Harris, & Orey, 2001). PVTs are designed to appear more difficult than they actually are (Heilbronner et al., 2009; Inman & Berry, 2002), so much so that individuals with known neurologic, psychiatric, and developmental disorders typically perform normally (Heilbronner et al., 2009; Sweet, 1999). PVTs are designed to be relatively insensitive to genuine cognitive impairment and the effects of psychological illness but sensitive to suboptimal performance (Bianchini, Mathias, & Greve, 2001).

Much of the research investigating PV has been conducted with patients who report cognitive symptoms yet have no apparent deficits on neurological testing (e.g.,

mild traumatic brain injury [mTBI]) (Binder, 1990). Such patients are frequently seeking compensation for an alleged injury or involved in litigation, both of which can serve as external incentives to perform suboptimally on testing. Research examining PVT scores in this population has largely demonstrated that mTBI patients involved in litigation or seeking compensation perform significantly worse on PVTs than patients with moderate to severe head injuries not seeking compensation (e.g., Binder & Kelly, 1996; Boone & Lu, 2007; Green & Iverson, 2001; Greiffenstein, Baker & Gola, 1994; Slick et al., 2003). Overall, the estimate of suboptimal performance during neuropsychological evaluation in the mTBI population is 40% (Larrabee, 2003; Larrabee, 2005). However, poor scores on PVTs are not limited to the mTBI population. High rates of suboptimal performance have also been found in other compensation-seeking patient populations without brain injuries, such as patients with chronic pain (42%; Gervais, Green, Allen, & Iverson, 2001a), fibromyalgia (44%; Gervais et al., 2001b), and toxin exposure (40%; Greve et al., 2006a; 56.7%; Greve et al., 2006b).

Studies have also found evidence of poor PVT scores in patients not seeking compensation or involved in litigation. One such patient group is the epilepsy population. Patients with epilepsy are considered to be motivated for neuropsychological testing with no apparent external reasons (e.g., financial incentives) to underperform. They are usually not being evaluated as a component of a litigation or disability case; in fact, they may already be receiving disability due to their epilepsy. However, research has shown that it cannot be assumed that all patients with epilepsy perform optimally during neuropsychological testing. In the limited amount of research specifically aimed at exploring PVT scores in patients with epilepsy, the base rate of suboptimal performance

has been found to range from 4 (Hill, Ryan, Kennedy, & Malamut, 2003) to 28% (Loring, Lee, & Meador, 2005). Reasons for suboptimal performance in this population are unknown. Therefore, further exploration is called for, which constitutes one of the purposes of the current study. However, regardless of underlying reasons, poor PVT scores are apparent in patients with epilepsy. These poor scores are concerning and require further investigation, as poor PVT scores are associated with significantly lower test scores across most cognitive domains (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Green, Lees-Haley, & Allen, 2002; Green, Rohling, Lees-Haley, & Allen, 2001). This is especially problematic for patients with epilepsy, as results of neuropsychological testing are used to assess seizure lateralization, pre- and post-surgical cognitive functioning, and psychiatric status.

Statement of the Problem

There is limited research specifically examining PVTs in patients with epilepsy. In the research that does exist, the base rate of suboptimal performance on such tests has been found to vary from 4 (Hill et al., 2003) to 28% (Loring et al., 2005). As patients with epilepsy are considered to be motivated for neuropsychological evaluation, this wide range of suboptimal performance on PVTs is unexpected and requires further investigation. Such an investigation would not only help clarify the base rate of suboptimal performance in this population, but may also shed light on possible explanations for this variance in base rate.

Reasons for the variance in base rate of suboptimal performance in patients with epilepsy remain unknown and largely unexplored. One possible explanation for this variance is the significant cognitive impairment commonly associated with epilepsy

(Bortz, 2003), although this theory needs to be further explored. For example, Drane et al. (2006) suggested that profound cognitive impairment likely explained four epilepsy patients' low scores on the Word Memory Test (WMT; Green, Allen, & Astner, 1996), a measure thought to be largely insensitive to significant cognitive impairment (Goodrich-Hunsaker & Hopkins, 2009; Green & Allen, 1999). However, this study was limited due to its small samples size and because WMT Genuine Memory Impairment Profile (GMIP) analysis was not employed. GMIP analysis indicates whether WMT scores below failure cutoff (≤ 82.5% on the Immediate Recognition [IR], Delayed Recognition [DR], or Consistency [CNS] subtests) are likely due to significant cognitive impairment or instead to suboptimal performance. GMIP analysis involves computing the difference between the mean of the WMT easy subtests (IR, DR, and CNS) and the WMT hard subtests (Multiple Choice [MC], Paired Associates [PA], and Free Recall [FR]). A GMIP is defined as at least a 30-point difference between the mean of the WMT easy and hard subtests, and suggests that WMT scores below failure cutoff are likely due to significant cognitive impairment rather than suboptimal performance. Given the limitations of the Drane et al. study, and the absence of other research exploring possible reasons for the varying base rate of suboptimal performance in this population, further investigation is warranted.

Secondly, there is a lack of research examining the relationship between PVT scores and neuropsychological test scores in the epilepsy population. The few studies that have explored this relationship have found that poor scores on PVTs were generally associated with lower scores on neuropsychological tests (Dodrill, 2008; Drane et al., 2006; Locke, Berry, Fakhoury, & Schmitt, 2006; Loring et al., 2005). However, only

Drane et al. excluded likely false positives (patients who likely perform suboptimally on PVTs due to significant cognitive impairment) before conducting their analyses. It is therefore possible that the findings of Dodrill, Locke et al., and Loring et al. are confounded by data from patients with significant cognitive impairment. Thus, conclusions about the relationship between PVT scores and neuropsychological test scores in this population may be inaccurate, and further examination is warranted.

Purpose of the Study

The first purpose of this study is to investigate performance on all subtests of a highly sensitive and specific PVT, the WMT, in patients with epilepsy. Patients will be categorized into one of the following groups based on WMT scores: optimal performance, suboptimal performance, or GMIP. By utilizing GMIP analysis, this study will seek to differentiate patients who score below failure cutoff on the WMT due to suboptimal performance from those who score below failure cutoff due to significant cognitive impairment. The use of GMIP analysis has not yet been explored in patients with epilepsy, and may uncover that a significant number of patients with epilepsy score below failure cutoff on the WMT due to significant cognitive impairment and *not* due to suboptimal performance. Therefore, employing GMIP analysis may help clarify the varying published base rate of poor PVT scores (4 to 28%) in this population. This clarification would be helpful because patients with epilepsy are considered to be motivated for testing, and therefore, not expected to obtain low PVT scores.

The second purpose of this study is to explore the relationship between WMT performance and neuropsychological test scores in patients with epilepsy. This relationship will be explored by examining differences among each WMT group's

neuropsychological test scores. Only four studies have explored the relationship between PVT scores and neuropsychological test scores in patients with epilepsy, two of which (Dodrill, 2008; Drane et al., 2006) only employed portions of the WMT. Using the WMT GMIP, this will be the first study to remove likely false positives (patients who perform suboptimally on the WMT likely due to significant cognitive impairment) from the suboptimal performance group before investigating the relationship between poor WMT scores and neuropsychological test scores. Additionally, this will be the first study to explore the relationship between GMIPs and neuropsychological test scores.

As such, the final purpose of this study is to explore the validity of the GMIP in patients with epilepsy. More specifically, this study seeks to examine whether or not GMIP scores are associated with impaired performance on neuropsychological memory tests. If GMIP scores are associated with memory impairment, this association would provide support for the validity of the GMIP in identifying patients who perform below WMT failure cutoff due to significant cognitive impairment. If no significant relationship is found, questions about the validity of the GMIP in this population will be raised, as it would be anticipated that patients who receive a GMIP score of ≥ 30 would also have impaired memory test scores if the GMIP accurately identifies the presence of significant cognitive impairment. This study will also explore the validity of the GMIP by examining how much each of the WMT subtests, as well as constructed PV and memory composites, explains total GMIP score. Investigations of the validity of the GMIP have yet to be conducted in patients with epilepsy.

Implications of the Study

Significant findings from the present study seek to inform clinical practice and future research. This study aims to clarify the varying base rate of suboptimal performance on PVTs in patients with epilepsy through the use of WMT GMIP analysis. If GMIP analysis reveals that a significant number of patients score below WMT failure cutoff due to significant cognitive impairment, neuropsychologists would be encouraged to administer all WMT subtests so that GMIP scores can be computed. GMIP analysis may therefore aid neuropsychologists with the interpretation of WMT results in this population.

This study also seeks to investigate the relationship between WMT performance and neuropsychological test scores in patients with epilepsy. If results indicate that patients who perform suboptimally on the WMT have significantly lower neuropsychological test scores than patients who perform optimally on the WMT, neuropsychologists would be encouraged to use the WMT with epilepsy patients to identify those who underperform. Significantly lowered test scores due to suboptimal performance are problematic for patients with epilepsy, as such scores are used to inform neuropsychologists' impressions about seizure lateralization, pre- and post-surgical cognitive functioning, and psychiatric status. These impressions may be inaccurate if suboptimal performance is not identified through the use of a PVT like the WMT. Neuropsychologists should mention evidence of suboptimal performance during testing in their reports and note that test results likely underestimate the patient's optimal abilities and should be interpreted with caution.

The present study additionally seeks to investigate the validity of the GMIP in patients with epilepsy. If findings from the current study support the validity of the GMIP,

neuropsychologists would be encouraged to use GMIP analysis in order to differentiate epilepsy patients who score below WMT failure cutoff due to significant cognitive impairment from those who score below WMT failure cutoff due to suboptimal performance. Further, results supporting the validity of the GMIP in patients with epilepsy may lead to future research employing the WMT GMIP with other populations, especially those with significant cognitive impairment. Such studies may help clarify the high false positive rates of other PVTs (e.g., the Test of Memory Malingering [TOMM]; Tombaugh, 1996) in patients with significant cognitive impairment (e.g., dementia).

Research Questions

1. What are the base rates of optimal, suboptimal, and GMIP performance as measured by the WMT?
2. Are there differences on WMT subtest scores among WMT groups (optimal performance, suboptimal performance, and GMIP)?
3. Are there differences in neuropsychological test scores among WMT groups?
4. What is the relationship between GMIP scores and scores on neuropsychological memory tests?
5. How much does each of the WMT subtests explain total GMIP score?

To answer these questions, WMT and neuropsychological test data from patients with epilepsy will be retrospectively analyzed using one-way analyses of variance (ANOVAs), multiple regressions with dummy coding, and simple linear regression.

Definition of Terms

Performance validity (PV): “the validity of actual ability task performance, assessed either by stand-alone tests such as Dot Counting or by atypical performance on neuropsychological tests such as Finger Tapping” (Larrabee, 2012, p. 626). In this study, PV will be assessed by the WMT.

Performance Validity Tests (PVTs): measures that “clarify the extent to which a person’s test performance is or is not an accurate reflection of their (sic) actual level of ability” (Larrabee, 2012, p. 626). PVTs have been commonly referred to as symptom validity tests (SVTs) or effort tests in the literature; however, as effort continues to remain a poorly defined construct and as symptom validity more appropriately describes “the accuracy of symptomatic complaint on self-report measures” (Larrabee, 2012, p. 626), the more accurately descriptive term of PVT will be used in this study.

Sensitivity: the true positive rate (hit rate) for a test; that is, the number of individuals with a condition who have positive test results divided by all individuals with the condition (Hennekens & Buring, 1987). Thus, the sensitivity of a PVT indicates the number of subjects performing suboptimally who are identified as such by the PVT divided by all subjects performing suboptimally. High sensitivity indicates that the majority of subjects performing suboptimally on a PVT are identified as such. Low sensitivity indicates that a certain cut score produces a substantial number of false negative errors, which means that a percentage of subjects performing suboptimally go undetected.

Specificity: the true negative rate; that is, the number of individuals without a condition who have negative test results divided by all individuals without the condition (Hennekens & Buring, 1987). Thus, the specificity of a PVT indicates the number of

subjects performing optimally who are identified as such by the PVT divided by all subjects performing optimally. High specificity indicates that the majority of subjects performing optimally on a PVT are identified as such. Low specificity indicates that a certain cut score produces a substantial number of false positive errors, which means that a percentage of subjects performing optimally are misclassified as performing suboptimally.

Malingering: Slick et al. (1999) defined the malingering of neurocognitive dysfunction as

> The volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility. Substantial material gain includes money, goods, or services of nontrivial value (e.g., financial compensation for personal injury). Formal duties are actions that people are legally obligated to perform (e.g., prison, military, or public service, or child support payments or other financial obligations). Formal responsibilities are those that involve accountability or liability in legal proceedings (e.g. competency to stand trial) (p. 552).

Simulators: “normal,” non-injured subjects, often college students, who have been instructed to simulate (fake) memory impairment or cognitive deficit in anticipation of monetary compensation (Bianchini et al., 2001; Grote & Hook, 2007). Simulators may be coached as to the best way(s) to avoid detection of suboptimal performance or provided with specific instructions designed to maximize their ability to effectively malinger (Grote & Hook, 2007; Inman & Berry, 2002). Simulators may also be uncoached (naïve).

CHAPTER II: LITERATURE REVIEW

Overview of Performance Validity Testing

History

Neuropsychologists must examine the validity of test performance, determining whether observed deficits due to neurological illness or injury, or are instead feigned or exaggerated (Binder, 1990; Binder, 1993; Green, 2001; Slick et al., 1999). This task may be difficult because, at times, individuals may exaggerate or malinger for apparent secondary gain (Bush et al., 2005; Heilbronner et al., 2009). In addition, some individuals may underperform, either intentionally or unintentionally, and may be skilled at preventing detection (Larrabee, 1992; Slick et al., 1999). The reasons for underperformance are unclear and have not been well studied.

Psychologists' subjective assessments of PV and malingering have been found to be inaccurate (Faust, Hart, & Guilmette, 1998; Faust, Hart, Guilmette, & Arkes, 1988; Heaton, Smith, Lehman, & Vogt, 1978). This inaccuracy in subjective assessment is concerning, as high rates of poor scores on objective PV measures have been found in compensation-seeking patients who report subtle cognitive symptoms (e.g., 42% of chronic pain patients seeking disability; Gervais et al., 2001a; 41% of patients with mTBI seeking compensation or involved in litigation; Mittenberg, Patton, Canyock, & Condit, 2002) and in non-litigating patients with significant cognitive impairment (e.g., 28% of epilepsy surgical candidates; Loring et al., 2005). Besides the presence of secondary gains, reasons for poor scores on PVTs have not been studied and remain largely unknown; however, it is clear that a substantial amount of patients perform suboptimally.

As such, the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) advocate for neuropsychologists to move beyond clinical judgment and strongly recommend formal assessment of PV regardless of whether a financial incentive to exaggerate cognitive impairment exists (AACN, 2007; Bianchini et al., 2001; Boone, 2007; Boone, 2009; Bush et al., 2005; Heilbronner et al., 2009; Lynch, 2004; Slick et al., 1999).

The roots of performance validity testing (PVT) lie in the application of operant learning methods in which an individual's behavior is modified by consequences (e.g., reinforcements, punishments, extinction). Brady and Lind (1961) used such methods to detect feigned neurological symptoms in a patient with hysterical blindness. Grosz and Zimmerman (1965) reassessed Brady and Lind's patient three years later using a forced-choice method. Slightly modifying the techniques used by Brady and Lind and Grosz and Zimmerman, Theodor and Mandelcorn (1973) utilized a two-alternative, forced-choice procedure to investigate hysterical blindness. In 1975, Pankratz, Fausti, and Peed applied such methods to investigate hysterical or malingered sensory deficits. Essentially, these early versions of modern-day PVTs began to quantify inconsistencies between ability and performance.

In 1983, Pankratz applied the two-alternative, forced-choice technique to assess the validity of memory deficits in a method he called "symptom validity testing (SVT)." In SVT, a stimulus is presented followed by two alternative stimuli (one target and one foil) from which the patient must select the correct response. The patient has a 50% probability of guessing the correct answer. Patients attempting to exaggerate or feign disability will perceive that a 50% hit rate is too successful (Pankratz, 1983; Pankratz et

al., 1975), and will therefore attempt to appear more impaired than they actually are. In doing so, they will score poorly, and, at times, worse than chance (Grosz & Zimmerman, 1965; Theodor & Mandelcorn, 1973; Pankratz et al., 1975). Below chance level of performance signifies that the patient recognizes correct responses (targets) but instead intentionally chooses incorrect responses (foils) (Binder, 1990; Pankratz, 1983; Vickery et al., 2001). Although Reynolds (1998) argued that worse-than-chance performance is neither random nor based on chance and instead denotes "purposive distortion" (p. 272), it is difficult if not impossible to determine one's intent in performing poorly on a SVT (Boone, 2007).

Pankratz's (1983) SVT procedure has been adapted to assess various sensory deficits (e.g., blindness, color blindness, blurry vision, tunnel vision, deafness) and memory complaints (Pankratz, 1988). In 1989, Hiscock and Hiscock further refined Pankratz's SVT procedure with the advent of what can be considered the first stand-alone forced-choice PVT, the Digit Memory Test (DMT). The DMT consists of 72 trials (three 24-item blocks) in which a five-digit number string is presented followed by a 5, 10, or 15-second delay (delays increase with each block) and then a two-choice recognition trial. Performance is based on scoring above or significantly below chance.

Currently, the design created by Hiscock and Hiscock (1989) generally serves as the foundation of most PVTs used today (Bianchini et al., 2001). The majority of such tests involve the presentation of visual or verbal stimuli followed by forced-choice recognition trails (Bianchini et al., 2001; Bickart, Meyer, & Connell, 1991). Most PVTs rely on a recognition memory format and appear to be assessing memory, but actually

very little memory is required to perform well (Guilmette, Hart, Guiliano, & Leninger, 1994; Heilbronner et al., 2009; Inman & Berry, 2002).

As noted in Chapter I, individuals with known neurologic, psychiatric, and developmental disorders typically perform well on most PVTs (Heilbronner et al., 2009; Sweet, 1999), providing support for their relative insensitivity to significant cognitive impairment and the effects of psychological illness (specificity). In contrast, research has consistently demonstrated that compensation-seeking/litigating patients who report subtle cognitive symptoms (e.g., mTBI) perform significantly worse on PVTs than do patients with similar or worse injuries or disorders who are not seeking compensation or involved in litigation (e.g., Binder & Willis, 1991; Constantinou et al., 2005; Gierok, Dickson, & Cole, 2005; Green, Iverson, & Allen, 1999; Green et al., 2002; Green et al., 2001). These findings support the sensitivity of such tests to suboptimal performance. The underlying basis of PVT is that individuals who perform suboptimally during evaluation will receive an improbably low score on these seemingly difficult tests, when, in fact, they should be performing normally (Bickart et al., 1991).

Methods for scoring PVTs. As previously noted, the probability of a correct response on a two-alternative forced-choice test is 50% (Theodor & Mandelcorn, 1973). The probability of attaining any given score can be determined by referring to a table of binomial probabilities (Bickart et al., 1991). By using this table, scores can be determined to be either significantly above or below chance, and subsequently interpreted as indicative of optimal or suboptimal performance. For example, if a PVT has 50 forced-choice trials, an individual would be expected to obtain a score of approximately 25 correct and 25 incorrect simply by guessing. If this individual were to obtain a score of

17, this would indicate suboptimal performance and would only be expected to occur by chance about 16 out of 1,000 times. However, more recently, research has indicated that the significantly below-chance level of performance has low sensitivity for identifying suboptimal performance or malingering (Binder, 2002; Grote et al., 2000; Guilmette et al., 1993; Hiscock, Branham, & Hiscock, 1994; Martin, Bolter, Todd, Gouvier, & Niccolls, 1993; Martin, Hayes, & Gouvier, 1996). These studies found that a substantial number of feigning individuals, especially those involved in litigation, scored above chance (e.g., >50% correct) on various PVTs, yet significantly below performance expectations given their actual, or, in the case of simulation studies, feigned, injury status. For example, Guilmette et al. found that only 34% of simulators asked to feign memory impairment on the DMT scored significantly below chance. Martin et al. found that only 28% of simulators scored significantly below chance in another study using the same measure.

Findings from these and the additional above-referenced studies led to the creation of two methods for determining above-chance cutoff scores (Bianchini et al., 2001). The first method – the statistical approach – uses a fixed, random, numeric cut score (e.g., 90% correct) to interpret performance. The second method – the normative-based approach – utilizes empirically derived cut scores based on the performance of individuals with documented brain damage without any known motivation or external incentives to malinger or exaggerate difficulties (Binder & Willis, 1991; Guilmette et al., 1993). Overall, below-chance level of performance on PVTs is a relatively rare phenomenon and has low sensitivity for identifying suboptimal performance. As such, over time, scoring procedures have been modified in order to be more sensitive to suboptimal performance and malingering.

Neuropsychologists' Current Practices in Assessing PV

Presently, it appears that neuropsychologists are heeding the advice of the AACN and NAN. Five studies examining neuropsychologists' practices in assessing PV revealed that the majority of respondents, ranging from 56 to 79%, reported using at least one PVT during evaluation (Lally, 2003; Mittenberg et al., 2002; Rabin, Barr, & Burton, 2005; Sharland & Gfeller, 2007; Slick, Tan, Strauss, & Hultsch, 2004). The results of these studies suggest that neuropsychologists most frequently use the TOMM and Rey 15-Item Test (FIT; Rey, 1964), followed by the WMT. Sweet (2011) and others (e.g., Bush et al., 2005; Heilbronner et al., 2009) have recommended that PVT become a standard of practice, and survey results indicate that neuropsychologists are moving in that direction.

Types of PVTs

The two forms of PVT include embedded indices in standard neuropsychological tests (e.g., Reliable Digit Span [RDS]; Greiffenstein et al., 1994; Vocabulary-Digit Span [VDS]; Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995) and stand-alone measures used solely to measure PV (e.g., WMT, DMT). Embedded measures will not be reviewed, as they are not relevant to this study. Generally, stand-alone PVTs have been found to have moderate levels of sensitivity and strong levels of specificity. Stand-alone measures can be subdivided into non-forced-choice and forced-choice.

Non-forced-choice PVTs. Non-forced-choice PVTs are measures that permit a range of responses (Heilbronner et al., 2009). In order to identify suboptimal performance, these tests may assess random responding, extremely slow or incorrect responding, and inconsistency of response patterns (Heilbronner et al., 2009). The FIT is the most

frequently used non-forced-choice PVT (Lees-Haley, Smith, Williams, & Dunn, 1995; Sharland & Gfeller, 2007; Slick et al., 2004). Other non-forced-choice PVTs include the Rey Dot Counting Test (DCT; Rey, 1941), *b* Test (Boone, Lu, & Herzberg, 2002; Boone et al., 2000), and the Rey Word Recognition Test (WRT; Rey, 1941). Research has found that non-forced-choice measures (e.g., FIT) tend to be less sensitive and specific than forced-choice tests (e.g., Reznek, 2005; Vickery et al., 2001).

Forced-choice PVTs. Forced-choice tests are the most common type of PVT employed during neuropsychological assessment (Boone & Lu, 2007; Nitch & Glassmire, 2007). These measures rely on a forced-choice recognition memory format in which subjects are presented with a set of target stimuli (words, numbers, or pictures). They are then shown pairs of stimuli (targets and foils) and must choose the target items (Boone & Lu, 2007; Heilbronner et al., 2009). Forced-choice PVTs include the DMT, WMT, TOMM, Portland Digit Recognition Test (PDRT; Binder, 1993; Binder & Willis 1991), Victoria Symptom Validity Test (VSVT; Slick, Hopp, & Strauss, 1995; Slick, Hopp, Strauss, & Thompson, 1997), Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997; Conder, Allen, & Cox, 1992), Validity Indicator Profile (VIP; Frederick, 1997; Frederick & Crosby, 2000), Medical Symptom Validity Test (MSVT; Green, 2004), and Non-Verbal Medical Symptom Validity Test (NV-MSVT; Green, 2006). Forced-choice PVTs have been validated with various populations, including patients with neurological injuries or disorders, psychiatric disorders, and medical illnesses (e.g., Binder & Willis, 1991; Conder et al., 1992; Frederick & Crosby, 2000; Iverson, Green, & Gervais, 1999; Slick, Hopp, Strauss, & Spellacy, 1996;

Tombaugh, 1996). These measures tend to be more sensitive and specific when compared to non-forced-choice PVTs (Heilbronner et al., 2009; Vickery et al., 2001).

Overall, the WMT has been found to be the most sensitive and specific forced-choice PVT (Drane et al., 2006; Green & Allen, 1999; Green, 2005; Iverson et al., 1999; Tan, Slick, Strauss, & Hultsch, 2002). Through the use of GMIP analysis, it also provides information on whether low scores are likely due to significant cognitive impairment or suboptimal performance. This is a useful feature that addresses weaknesses found in other PVTs, such as the high false positive rates of the TOMM, FIT, and DCT when used with patients with profound cognitive impairment (e.g., dementia and mental retardation). Despite this strength, few studies (Green, Flaro, & Courtney, 2009; Green, Montijo, & Brockhaus, 2011; Henry, Merten, Wolf, & Harth, 2009; Howe, Anderson, Kaufman, Sachs, & Loring, 2007; Howe & Loring, 2009; Singhal, Green, Ashaye, Shankar, & Gill, 2009) have employed GMIP analysis. Overall, results of these studies strongly supported the application of GMIP analysis to WMT, MSVT, and NV-MSVT results in patients with significant cognitive impairment such as that seen in dementia. Specificity rates in the 90s (as high as 98%; Green et al., 2011) have been achieved using GMIP analysis on WMT, MSVT, and NS-MSVT scores in patients with possible mild cognitive impairment, probable dementia, and other neurological disorders associated with significant cognitive impairment (e.g., Henry et al., 2009; Howe et al., 2007; Howe & Loring, 2009). Additional studies utilizing GMIP analysis would be able to improve classification accuracy (i.e., reduce false positives) in other patient populations likely to have significant cognitive impairment (e.g., epilepsy), meaning that such patients would be

identified as scoring below failure cutoff due to significant cognitive impairment and not due to suboptimal performance.

Critical Nature of PVT within Neuropsychological Assessment

As noted in Chapter I, PVTs have become a critical part of neuropsychological assessment. Low scores on these measures call into question the validity of neuropsychological test results (Constantinou et al., 2005; Green et al., 2002, Green et al., 2001) and raise concerns regarding the validity of symptoms being reported. Questions about the validity of test results and reported symptoms impact the neuropsychologist's ability to accurately make diagnoses, prognoses, and appropriate referrals (Constantinou et al., 2005).

A few studies have explored the relationship between PVT scores and neuropsychological test scores. In a study of 904 heterogeneous outpatients who underwent neuropsychological evaluation as part of a compensation claim or litigation, Green et al. (2001) found that a PV composite index (comprised of scores from the WMT, CARB, and California Verbal Learning Test) accounted for 49 to 54% of the variance in overall neuropsychological performance. PV explained more score variance than injury severity, demographic variables, and neuropsychological test scores. Of the measures used in this study, the WMT was the best predictor of overall neuropsychological performance. Patients who did poorly on the WMT, as indicated by scores below failure cutoff, performed significantly worse on neuropsychological tests than those who scored above failure cutoff. WMT scores below failure cutoff also suppressed overall performance 4.5 times more than did moderate-to-severe brain injury. These data suggest

that suboptimal performance on the WMT more strongly predicts neuropsychological test scores than brain injury or neurological disease.

Other studies have found similar results. Using the Green et al. (2001) sample, Green et al. (2002) found that the average of the WMT IR, DR, and CNS scores accounted for 49% of the variance in overall neuropsychological profile (r = .70). Similarly, Rohling, Allen, and Green (2002) found that a PV composite (comprised of the CARB total score and the average of the WMT IR, DR, and CNS scores) accounted for 36 to 45% of the variance in overall neuropsychological profile (r = .67) in 561 patients with various disorders involved in compensation claims. More recently, using the Green et al. (2001) database plus 403 additional cases, Green (2007) found that patients who scored in the bottom WMT performance range (mean of IR, DR, and CNS scores ≤ 50%) scored approximately two standard deviations below those in the top performance range (91-100%) on eight neuropsychological tests. Constantinou et al. (2005) retrospectively analyzed neuropsychological data and TOMM scores from 69 litigants with mTBI. Results indicated that poor performance on the TOMM explained 47% of the variance in the overall neuropsychological deficit score on the Halstead-Reitan Neuropsychological Battery for Adults (HRNB-A; Reitan & Wolfson, 1993). Similar to Green et al.'s (2001) findings, litigants who performed poorly on the TOMM obtained neuropsychological test scores that were much lower than what is typically expected from mTBI patients and significantly lower than the litigants who received high scores on the TOMM. Finally, Stevens, Friedel, Mehren, and Merten (2008) found that poor performance on the WMT and MSVT explained up to 35% of the variance in neuropsychological test scores in a

retrospective analysis of data from 233 Workers' Compensation Boards patients and personal injury litigation claimants.

The above findings highlight the critical nature of PVT during neuropsychological evaluation, as PVT scores have been found to account for up to 54% of the variance in neuropsychological test scores. Further, these findings support Green et al.'s (2002) statement that failure on PVTs most likely leads to suboptimal scores on other tests. The above findings have led to the recommendation that neuropsychologists employ PVT with all patients so that low neuropsychological test scores are not automatically attributed to neurological injuries or disorders, psychiatric disorders, or medical illnesses.

PVT in Patients with Epilepsy

Epilepsy, also known as seizure disorder, is a common neurological disorder that affects individuals cognitively, psychologically, and physically (Epilepsy Foundation of America [EFA], 2010; Schachter, 2009; Sirven, 2002). Research has demonstrated that patients with epilepsy may have impairments in one or more areas of cognitive or motor functioning (Bortz, 2003; Ettinger & Kanner, 2007; Jones-Gotman et al., 2010). Additionally, psychological difficulties (e.g., depression, anxiety, psychoses, attention and impulsivity problems, and personality and behavioral disorders) are present in a large number of patients with epilepsy, sometimes in greater numbers than the general population (Bortz, 2003; Ettinger & Kanner, 2007; Hessen, Lossius, & Gjerstad, 2008; Marcangelo & Ovsiew, 2007; Moore & Baker, 2002; Sadock & Sadock, 2007; Torta & Keller, 1999). Although much research has been devoted to the study of neuropsychological functioning in epilepsy, this research has not typically included PVTs (Locke et al., 2006). As previously noted, if patients do not perform to the best of their

ability during evaluation, test results may be invalid (Green et al., 2002), or, at the very least, underestimate their true abilities. This is especially troublesome for patients with epilepsy, as the results of neuropsychological evaluation are used to help lateralize seizure focus, determine pre- and post-surgical cognitive functioning, and assess psychiatric status.

There is limited research specifically focused on assessing PV within the epilepsy population. Instead, many studies have included patients with epilepsy as neurological controls to demonstrate that PVTs are largely insensitive to significant cognitive impairment. Several studies (Binder & Willis, 1991; Green et al., 2001; Grote et al., 2000; Prigatano, Smason, Lamb, & Bortz, 1997; Rohling et al., 2002; Slick et al., 1996) found that epilepsy patients tended to perform better on PVTs than simulators, compensation-seeking patients with head injury, and patients suspected of malingering. These findings appear to support the contention that PVTs are relatively insensitive to significant cognitive impairment, but, as will be described below, larger studies of patients with epilepsy challenge these conclusions.

Most studies examining PVTs within the epilepsy population have used these measures to aid in the differential diagnosis of patients with epilepsy and patients with psychogenic nonepileptic seizures (PNES), a conversion or somatoform disorder not caused by underlying neuropathology. In such studies, it was presumed that patients with epilepsy would perform better on PVTs than patients with PNES, as they typically lack external incentives to perform poorly and are considered to be motivated for testing. Consistent with this presumption, Binder, Salinsky, and Smith (1994) and Binder, Kindermann, Heaton, and Salinsky (1998) found that patients with epilepsy scored

significantly higher on the PDRT than patients with PNES. Contrary to these findings, however, Hill et al. (2003) found no difference in mean performance across TOMM trials in 48 patients with epilepsy and 57 patients with PNES. Reasons for this discrepancy in findings have not been specifically investigated in patients with epilepsy, but studies with other populations suggest that the TOMM may not be as sensitive or specific as other PVTs, and that the test might not be appropriate for use with individuals with significant cognitive impairment. For example, the WMT, VSVT, and CARB have been found to be more sensitive than the TOMM in simulators and disability claimants (Gervais, Rohling, Green, & Ford, 2004; Tan et al., 2002). Another study (Teichner & Wagner, 2004) found the TOMM to have a high misclassification rate (poor specificity) in patients with dementia. These findings, along with results from Hill et al., suggest that the TOMM may not be appropriate for use with epilepsy patients, many of whom may have significant cognitive impairment.

Other studies of the differential diagnosis of epilepsy and PNES also presumed that patients with epilepsy would perform well on PVTs; but this was not the case. Williamson, Drane, Stroup, Miller, and Holmes (2003) found that 13% of patients with epilepsy scored below failure cutoff on the WMT, compared to 64% of patients with PNES. These results indicated that 13% of patients with epilepsy did not perform optimally on the WMT and, therefore, that their neuropsychological test scores were potentially invalid. More recently, Cragar et al. (2006) investigated the performance of 41 patients with epilepsy, 21 patients with PNES, and 18 patients with epilepsy plus PNES on four PVTs: the DMT, Letter Memory Test (LMT; Inman et al., 1998), TOMM, and PDRT. Results revealed that 22% of patients with epilepsy, 24% of patients with PNES,

and 11% of patients with epilepsy plus PNES scored below cutoffs on at least one PVT. As expected, findings from these studies indicated that patients with PNES performed below cutoffs on PVTs at significantly higher rates than patients with epilepsy. However, patients with epilepsy also scored below cutoffs at higher rates than expected. These findings are surprising since, as previously noted, patients with epilepsy are presumed to do well on PVTs.

Evidence of suboptimal performance on PVTs in the epilepsy population has also been found in two recent studies (Hoskins, Binder, Chaytor, Williamson, & Drane, 2010; Loring et al., 2005) that did not explore differential diagnosis. Hoskins et al. (2010) found evidence of suboptimal performance in patients with epilepsy in an investigation of the oral versus computerized versions of the WMT. Subjects included 67 inpatients at an epilepsy center and 58 forensic and clinical referrals without epilepsy. Results indicated that 21% of patients with epilepsy scored below failure cutoff on the computer version, 14% scored below failure cutoff on the oral version, and 23% scored below failure cutoff on regardless of version. These results indicated that a substantial number of patients with epilepsy performed suboptimally on the WMT regardless of version. Loring et al. (2005) found evidence of suboptimal performance on the VSVT in a retrospective study of 120 non-litigating epilepsy surgical candidates. Findings indicated that 20 patients had questionably valid results and 14 patients had invalid results. Combined, these results indicated that approximately 28% of the sample performed suboptimally on the VSVT. These findings are unexpected, given that previous research assumed that epilepsy surgical candidates put forth optimal performance during testing because they were motivated for surgery. It is noteworthy that Loring et al. and Hoskins et al. found that

some subjects performed poorly on PVTs due to genuine cognitive impairment (e.g., low intellectual function), but this cannot account for all poor performances in their studies. Finally, while Lee, Loring, and Martin's study (1992) did not demonstrate suboptimal performance in their epilepsy patients, they employed the FIT, a PVT that has been found to have a low level of sensitivity (e.g., Reznek, 2005). Therefore, their results should be interpreted cautiously.

Relationship Between PVT Scores and Neuropsychological Test Scores

A few studies have explored differences in neuropsychological test performance between patients with epilepsy and patients with PNES using PVTs to control for suboptimal performance. Drane et al. (2006) found that patients with epilepsy who performed poorly on the WMT, as indicated by scores below failure cutoff, had more neuropsychological test scores below normal limits than patients who scored above failure cutoff. PNES patients with WMT scores below failure cutoff had more impaired test scores than did PNES patients and epilepsy patients who scored above cutoff, as well as epilepsy patients who scored below cutoff. Patients with PNES who scored above WMT failure cutoff displayed the least amount of neuropsychological impairment when compared to all other patient groups. In a similar investigation, Dodrill (2008) found that poor WMT scores in patients with epilepsy and patients with PNES were associated with lower scores on various neuropsychological tests. Locke et al. (2006) also found that poor performance on the TOMM was significantly predictive of lower IQ, memory, language, visuospatial, and motor functioning scores in patients with epilepsy and patients with PNES.

Loring et al. (2005) was the only study identified that that did not examine differences in neuropsychological test performance between patients with epilepsy and PNES. Instead, this study examined VSVT performance of 120 epilepsy surgical candidates. Patients who had valid VSVT scores had the highest performances on a variety of neuropsychological measures. Suboptimal performance on the VSVT was associated with decreased scores on a variety of neuropsychological tests.

Conclusion

Neuropsychologists must examine the validity of test performance during evaluation to help determine whether observed low test scores are due to neurological illness or injury or if instead they reflect suboptimal performance. Forced-choice PVTs have been found to be the most sensitive and specific measures of PV. Their utilization is critical because low PVT scores have been found to account for up to 54% of the variance in neuropsychological test scores (Constantinou et al., 2005; Green, 2007; Green et al., 2002; Green et al., 2001; Rohling et al., 2002; Stevens et al., 2008), and patients with low PVT scores have been found to score significantly lower on neuropsychological tests across cognitive domains.

Most of the research investigating PVTs has been conducted on compensation-seeking/litigating patients who report cognitive symptoms yet fail to show deficits on neurological testing (e.g., mTBI). These patients typically perform poorly on PVTs. Surprisingly, however, low PVT scores have also been found in patients with no apparent external incentives to underperform. One such group is patients with epilepsy, but there is limited research specifically examining PVT with this population.

In the preliminary research available, the base rate of suboptimal performance in patients with epilepsy has been found to range from 4 (Hill et al., 2003) to 28% (Loring et al., 2005). One possible explanation for this variance may be the significant cognitive impairment commonly associated with epilepsy (Bortz, 2003), although this theory has not been directly explored. One study (Drane et al., 2006) suggested that profound cognitive impairment accounted for low scores on the WMT in four patients with epilepsy. However, the false positive rate remains unknown for this study, as GMIP analysis was not employed. Presently, the impact of significant cognitive impairment on PVTs has yet to be fully explored in the epilepsy population. The WMT would be an ideal PVT to employ in future investigations, as the WMT with the GMIP has demonstrated high classification accuracy indices with other patient populations (e.g., Green et al., 2009; Howe et al., 2007; Howe & Loring, 2009). Future research using all subtests of the WMT and GMIP analysis may help identify patients with epilepsy who perform poorly due to significant cognitive impairment and thus help clarify the varying rates of suboptimal performance during neuropsychological evaluation in this population.

Finally, there have been several studies examining the relationship between PVT scores and neuropsychological test scores in the epilepsy population (Dodrill, 2008; Drane et al., 2006; Locke et al., 2006; Loring et al., 2005). Initial findings have generally been consistent with findings from other studies with other patient populations (Constantinou et al., 2005; Green, 2007; Green et al., 2002; Green et al., 2001; Rohling et al., 2002; Stevens et al., 2008), with lower PVT scores associated with significantly lower test scores across most cognitive domains. However, only Drane et al. removed patients with profound cognitive impairment (likely false positives) before conducting analyses.

Patients with significant cognitive impairment need to be identified and eliminated from datasets before examining the relationship between PVT scores and neuropsychological test scores. The WMT with GMIP analysis is best suited for this task. However, as Dodrill, Locke et al., and Loring et al. did not utilize GMIP analysis, it is possible that their findings are confounded by data from patients with significant cognitive impairment. As such, conclusions about the relationship between PVT scores and neuropsychological test scores in this population may be inaccurate. Therefore, further investigation is warranted using GMIP analysis to identify and remove likely false positives before examining the relationship between WMT performance and neuropsychological test scores. Future research should also examine the relationship between GMIP scores and neuropsychological memory test scores in order to establish the validity of the GMIP in the epilepsy population.

CHAPTER III: METHODOLOGY

Participants

The participants in this study were retrospectively identified from the Aurora St. Luke's Regional Epilepsy Center database maintained at Aurora St. Luke's Center for Neuropsychological Services in Milwaukee, WI. Patients were referred for neuropsychological evaluation in order to assess candidacy for epilepsy surgery (pre-surgical) or to assess cognitive functioning and psychological status (non-surgical). All patients had a history of medically intractable seizures, and all pre-surgical patients underwent 24-hour video-EEG monitoring to clarify seizure focus. Patients were diagnosed with epilepsy by a board certified neurologist at Aurora St. Luke's Regional Epilepsy Center.

Eligibility criteria. Patients were eligible for inclusion if the following criteria were met: (1) the patient had a diagnosis of epilepsy; (2) the patient was either a pre-surgical candidate or non-surgical; (3) the patient underwent neuropsychological evaluation and completed the full battery of tests; and (4) the patient was administered the first six subtests of the WMT during neuropsychological evaluation. Exclusion criteria included: (1) the patient underwent previous epilepsy surgery and (2) the patient was not administered the WMT. If a patient was evaluated more than once, either pre-surgically or non-surgically, data from his or her first testing session was used to remove potential practice effects for neuropsychological measures as well as the WMT.

Measures

As part of the standard evaluation process at Aurora St. Luke's Center for Neuropsychological Services, patients were administered an extensive battery of 15 neuropsychological tests and the WMT. The neuropsychological tests were selected to evaluate a variety of cognitive domains. The WMT was selected because it is a highly sensitive and specific PVT that allows for GMIP analysis.

WMT. The WMT is a PVT that consists of six subtests. As defined in Chapter 1, the IR, DR, and CNS subtests are measures of PV (Green, 2005). The multiple choice (MC), paired associates (PA), free recall (FR), and long delayed free recall (LDFR) subtests are measures of memory (Green, 2005). The WMT requires the patient to learn a list of 20 word pairs (e.g., fish-fin, dog-cat) presented twice. The patient then choses the words from the original list from new pairs of words containing both the target word and a foil word (IR). Thirty minutes later, the patient discriminates the original words from a different set of foils (DR). Next, the first word from each pair is presented and the patient selects the word that was paired with it from eight choices (MC). Afterwards, the examiner read the first word in each pair aloud and the patient provides the second word (PA). He or she is then asked to freely recall all of words from the original list in any order (FR). The LDFR (optional subtest) may be given 20 minutes later and consists of the patient freely recalling as many of the original words as possible in any order. This subtest was not administered. The WMT can be administered verbally or self-administered on a computer. The computerized version was administered.

Per the manual, scores above 90% on IR, DR, or CNS indicate a clear pass. Scores at or below 82.5% on IR, DR, or CNS indicate a clear fail. Scores between 83 and 90% are classified as caution. Although MC and PA are classified as memory subtests,

MC scores of 70% or below and PA scores of 50% or below are classified as warning. Such low scores on MC and PA warrant further investigation and are considered suspicious of suboptimal performance when dementia or other profound cognitive impairments have been ruled out (Green, 2005). Additionally, scores on six of the subtests can be used to compute a GMIP score. This score indicates whether poor performance is likely due to significant cognitive impairment (and thus *not* likely attributable to suboptimal performance) or is secondary to suboptimal performance. A GMIP is defined as at least a 30-point difference between the mean of the easy subtests (IR, DR, and CNS) and the mean of the hard subtests (MC, PA, and FR).

Psychometrics. Initial validation studies on the WMT were conducted on data from more than 1,250 consecutive outpatients referred for neuropsychological evaluation over a period of eight years. These studies found the WMT to be a sensitive and specific PVT with various populations, including healthy adults, simulators, patients with neurological disorders, and patients with impaired memory. First, an initial validation study compared WMT scores of 40 healthy adults (Iverson et al., 1999) to those of 57 patients with moderate-to-severe TBI (Green & Allen, 1999). Patients with TBI averaged IR and DR scores above 95% correct and averaged 95.1% correct across all PV subtests, indicating that the WMT is largely insensitive to neurological impairment secondary to serious head injury. Second, patients with neurological disorders have been found to perform above failure cutoff despite having significant impairment (Gorissen, Sanz de la Torre, & Schmand, 2003, cited in Green, 2005; Green & Allen, 1999). For example, in Green and Allen's sample of 40 neurological patients, no significant differences in WMT PV subtest scores were found between patients with impaired and normal scores on the

CLVT, a measure of verbal memory; however, those with impaired CLVT scores scored significantly lower on the WMT memory subtests. Another study found that amnestic patients with hippocampal damage scored above failure cutoff, further supporting the insensitivity of the WMT PV measures to severe cognitive impairment (Goodrich-Hunsaker & Hopkins, 2009). Finally, initial studies conducted on TBI patients have shown that patients with severe brain injuries scored significantly *higher* on WMT PV subtests than those with mild brain injuries (Green, 2005; Green et al., 1999). These results contradict what would be expected if the WMT PV subtests were sensitive to brain injury; if that were the case, it would be expected that those with severe head injuries would fail these subtests. Instead, the mean DR score in those with severe brain injuries (as indicated by abnormal brain scans) was 90.7% correct, which was significantly higher than the mean DR score in those with normal brain scans (82.5% correct) (Green, 2005). Green (2005) concluded that those with mTBIs may be more likely to exaggerate their impairment by putting forth suboptimal performance on the WMT and scoring significantly lower than those with severe TBIs. Overall, the WMT PV subtests have been found to be relatively insensitive to significant cognitive impairment and neurological disorders (Gorissen et al., 2003, cited in Green, 2005; Green, 2005; Green & Allen, 1999) and sensitive to suboptimal performance (Green, 2005; Green et al., 1999). The WMT memory subtests have been found to be sensitive to impaired memory (Green & Allen, 1999).

The WMT has also been found to be a reliable measure of PV with high levels of internal consistency and test-retest reliability. In a sample of 1,207 outpatients, internal consistency was found to be strong between IR and DR (r = .88); MC and PA (r = .90);

and FR and LDFR (r = .86) (Green, 2005). Test-retest reliability of the mean IR and DR scores was found to be very high in a sample of 20 healthy adults (r = .97) (Green, 2005). Test-retest reliability of the mean MC and PA scores was found to be extremely high (r = .99) and very high for the mean of the DR and LDFR scores (r = .92). However, it is important to note that test-retest reliability may not be as strong when examining clinical cases, as clinical patients' performance is more likely to fluctuate over time possibly due to becoming involved in litigation or applying for compensation (Green, 2005). In fact, Green's initial clinical validation sample of over 1,250 outpatients demonstrated poor test-retest reliabilities of .43 for IR scores and .33 for DR scores. The poor test-retest reliabilities of these subtests likely reflected variable engagement during testing and were not likely indicative of a lack of reliability of WMT scores (Green, 2005).

Finally, the WMT has been found to be a highly sensitive and specific PVT. In the initial validation study, 20 clinical patients being evaluated for disability scored above failure cutoff, with an average DR score of 98.2% (Green, 2005). The same patients were then instructed to fake memory impairment. All patients scored below failure cutoff on at least one of the PV measures, resulting in a sensitivity of 100%. Sensitivity was 96% for a group of 25 simulators who scored below failure cutoff on at least one of the PV subtests (Green, 2005). When the two simulator groups were combined, the WMT was found to have a sensitivity of 97.7%. The WMT has also been found to have strong specificity. Data from a study with healthy adult controls (Iverson et al., 1999) and moderate-to-severe TBI patients (Green & Allen, 1999) indicated a specificity of 100%. Results from Tan et al. (2002) also found a WMT specificity of 100%. Overall, the WMT has been found to be a highly sensitive and specific PVT.

Wechsler Adult Intelligence Scale–Third Edition (WAIS–III). The WAIS–III (Wechsler, 1997a) is a clinical instrument for assessing the intellectual ability of individuals aged 16 through 89 years (The Psychological Corporation, 2002). It is comprised of 14 subtests: Picture Completion, Vocabulary, Digit-Symbol Coding, Similarities, Block Design, Arithmetic, Matrix Reasoning, Digit Span, Information, Picture Arrangement, Comprehension, Symbol Search, Letter-Number Sequencing, and Object Assembly. Depending on which subtests are administered, the WAIS–III can provide three IQ scores: Full Scale Intelligence Quotient (FSIQ), Verbal Intelligence Quotient (VIQ), and Performance Intelligence Quotient (PIQ). Four Index scores can also be computed: Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Working Memory Index (WMI), and Processing Speed Index (PSI). The WAIS-III as opposed to the WAIS-IV was administered in this study, as the WAIS-III was part of a previously established neuropsychological test battery administered to all epilepsy patients undergoing evaluation.

WAIS–III indexes and subtests. The VCI is a measure of acquired knowledge and verbal reasoning (The Psychological Corporation, 2002). It is comprised of the Vocabulary, Similarities, and Information subtests. The Vocabulary subtest, a measure of word knowledge (Sattler & Ryan, 2009), requires the patient to orally define a series of visually and orally presented words. The Similarities subtest, which measures the ability to recognize and verbalize relationships between two objects or concepts (Sattler & Ryan, 2009), requires the patient to verbally explain the similarity of orally presented word pairs. The Information subtest, a measure of general fund of knowledge (Sattler & Ryan,

2009), requires the patient to orally answer a series of verbally presented questions that tap his or her knowledge of common events, objects, places, and people.

The POI is a measure of nonverbal, fluid reasoning, attentiveness to detail, and visual-motor integration (The Psychological Corporation, 2002). It is comprised of the Picture Completion, Block Design, and Matrix Reasoning subtests. The Picture Completion subtest, a measure of visual-perceptual reasoning ability (Sattler & Ryan, 2009), requires the patient to identify an important missing part in pictures of common objects and settings. The Block Design subtest, a measure of spatial visualization and perceptual reasoning skills (Sattler & Ryan, 2009), requires the patient to replicate modeled or two-dimensional geometric patterns using two cubes. The Matrix Reasoning subtest, a measure of nonverbal problem-solving skills (Sattler & Ryan, 2009), requires the patient to complete a series of incomplete gridded patterns by pointing to or saying the number of the correct response from five possible choices.

The WMI is a measure of auditory working memory and attention (The Psychological Corporation, 2002). It is comprised of the Arithmetic, Digit Span, and Letter-Number Sequencing subtests. The Arithmetic subtest measures one's ability to carry out mental arithmetic, which involves numerical reasoning abilities, attention, auditory short-term memory, and long-term memory (Sattler & Ryan, 2009). In this subtest, the patient is read a series of arithmetic problems that he or she must solve mentally and respond to orally. The Digit Span subtest, a measure of auditory short-term memory (Sattler & Ryan, 2009), requires the patient to repeat a series of orally presented number sequences. The patient must repeat the sequences verbatim in Digits Forward and backward in Digits Backward. The Letter-Number Sequencing subtest also measures

auditory short-term memory (Sattler & Ryan, 2009). In this subtest, the patient must track and orally repeat sequences of numbers and letters, putting the numbers in ascending order and the letters in alphabetical order.

The PSI is a measure of processing speed (The Psychological Corporation, 2002). It is comprised of the Digit-Symbol Coding and Symbol Search subtests. The Digit-Symbol Coding subtest measures processing speed (Sattler & Ryan, 2009). This subtest consists of a series of numbers that are each paired with their own corresponding symbol in a key. Using the key, the patient writes the symbol corresponding to its number as fast as possible. The Symbol Search subtest measures visual discrimination and visual-perceptual scanning ability (Sattler & Ryan, 2009). This subtest consists of a series of paired groups, with each pair consisting of a target group and a search group. By marking the appropriate box, the patient indicates whether either target symbol appears in the search group.

WAIS–III subtest raw scores are converted to age-corrected scaled scores. Index scores (standard scores) are also computed. Pearson computer scoring was used to score WAIS–III protocols.

Psychometrics. The WAIS–III was standardized on 2,450 healthy individuals aged 16 to 89 years. The sample was separated into 13 age groups and stratified by demographic variables including age, sex, education level, and geographic region of the United States (U.S.) according to U.S. 1995 census data (The Psychological Corporation, 2002). Per the technical manual, average internal consistency reliability coefficients for all of the subtests except Picture Arrangement, Symbol Search, and Object Assembly range from .82 to .93. Picture Arrangement and Object Assembly were found to have

lower average reliability coefficients of .74 and .70, respectively. As Symbol Search and Symbol-Digit Coding are speeded subtests, test-retest reliability coefficients were calculated. Symbol Search and Symbol-Digit Coding were found to have average test-retest reliabilities of .77 and .84, respectively. Average reliability coefficients for the four Indexes range from .88 to .96. Average reliability coefficients for FSIQ, VIQ, and PIQ scores range from .94 to .98.

The manual also provides evidence of convergent and discriminant validity for the WAIS-III. Overall, a high magnitude of intercorrelation was found between most of the subtests, supporting the notion of the presence of a general intelligence factor (*g*). Most of the subtests correlated with each other at a moderate level. All Index subtests were found to have at least moderate intercorrelations, providing evidence of convergent validity. Verbal subtests' intercorrelation coefficients ranged from .70 to .77. Perceptual organization subtests' intercorrelation coefficients ranged from .48 to .60. Working memory subtests' intercorrelation coefficients ranged from .52 to .57. Processing speed subtests' intercorrelation was .65. Subtests were found to strongly correlate with their Indexes. Finally, subtests that measure constructs that are not expected to be strongly correlated with each other (e.g., Vocabulary and Picture Completion) were found to have relatively lower intercorrelations when compared to intercorrelations between measures within domains, providing evidence of discriminant validity.

Further evidence of convergent validity was provided by moderate to high correlations (*r*s ranging from .50 to .91) between most of the WAIS–III IQ and Index scores and other measures of cognitive abilities, including the Wechsler Adult Intelligence Scale–Revised (WAIS–R; Wechsler, 1981), Wechsler Intelligence Scale for

Children–Revised (WISC–R; Wechsler, 1974), Standard Progressive Matrices (Raven, 1976), the Information Processing Accuracy Index of the MicroCog: Assessment of Cognitive Functioning (Powell et al., 1993), and the total score of the Dementia Rating Scale (DRS; Mattis, 1988). WAIS–III VCI and the MicroCog Information Processing Accuracy Index were not strongly correlated (r = .28), providing evidence of discriminant validity for the VCI.

Wechsler Memory Scale – Third Edition (WMS–III). The WMS–III (Wechsler, 1997b) is a neuropsychological battery of learning, memory, and working memory measures that can be used with individuals aged 16 to 89 years. It is comprised of 11 subtests. The six primary subtests include: Logical Memory I and II (LM II and II), Verbal Paired Associates I and II, Letter Number Sequencing, Faces I and II, Family Pictures I and II, and Spatial Span. The five optional subtests include: Information and Orientation, Word Lists I and II, Mental Control, Digit Span, and Visual Reproduction I and II (VR I and II). Many of the subtests contain an immediate and a delayed condition, which is administered 25 to 35 minutes following the immediate condition. The WMS–III provides eight Primary Index scores: Auditory Immediate, Visual Immediate, Immediate Memory, Auditory Delayed, Visual Delayed, Auditory Recognition Delayed, General Memory, and Working Memory. Four Auditory Process Composites are also computed: Single-Trial Learning, Learning Slope, Retention, and Retrieval. LM I and II and VR I and II will be reviewed below, as they were the only WMS–III subtests that were administered to patients in this study. The WMS-III as opposed to the WMS-IV was administered, as the WMS-III was part of a previously established neuropsychological test battery administered to all epilepsy patients undergoing evaluation.

LM I and II. The LM subtest is a measure of immediate and delayed auditory memory (The Psychological Corporation, 2002). During LM I, two short stories are read to the patient. The second story is read twice. The patient is asked to immediately recall the stories from memory. Twenty-five to 35 minutes later (LM II), the patient is asked to recall both stories again. A recognition trail follows where the patient answers yes/no questions about the stories. Six scores are computed from this subtest: LM I Recall Total, LM II Recall Total, LM I Thematic Total, LM II Thematic Total, LM II % Retention, and LM II Recognition Total. The LM II Recognition Total was not used in analyses, as it cannot be converted from a raw score to a scaled score or percentile. Subtest raw scores are converted to age-corrected scaled scores.

VR I and II. The VR subtest is a measure of immediate and delayed visual memory (The Psychological Corporation, 2002). During VR I, the patient is shown a series of five designs, one at a time, for 10 seconds. After each design is presented, the patient must draw the design from memory. During VR II (25 to 35 minutes later), the patient is asked to draw the designs from memory in any order. A recognition trail follows where the patient is shown a series of 48 designs, one at a time, and has to identify the designs presented in VR I. Six scores are computed from this subtest: VR I Recall Total, VR II Recall Total, VR II Recognition Total, VR % Retention, VR II Copy Total (copy trial not administered), and VR II Discrimination (discrimination trial not administered). Subtest raw scores are converted to age-corrected scaled scores.

Psychometrics. The WMS–III was standardized on 1,250 healthy individuals aged 16 through 89 years. The sample was separated into 13 age groups and stratified by demographic variables including age, sex, education level, and geographic region of the

U.S. according to U.S. 1995 census data (The Psychological Corporation, 2002). The internal consistencies of the LM and VR subtests have been found to be high. Per the technical manual, LM I and II Recall Totals have average internal consistency reliability coefficients of .88 and .79, respectively. LM I and II Thematic Totals have average generalizability reliability coefficients of .77 and .79, respectively. VR I and II Recall Totals have average generalizability reliability coefficients of .79 and .77, respectively. THE VR II Recognition Total has an average generalizability reliability coefficient of .75.

The manual also provides evidence of convergent and discriminant validity for the LM and VR subtests. The intercorrelation coefficients of the LM subtests ranged from .70 to .88, providing support that such subtests are measuring a similar construct. The intercorrelation coefficients for the VR subtests ranged from .27 to .67, with lower correlations found between VR II Copy and all other VR subtests. Overall, both the LM and VR subtests were found to have acceptable discriminant validity as evidenced by weak to moderate intercorrelation coefficients with subtests that measure different types of memory and learning.

Boston Naming Test (BNT). The BNT (Kaplan, Goodglass, & Weintraub, 1983) is a measure of confrontational naming. It was originally published by Kaplan, Goodglass, and Weintraub (1978) as an experimental 85-item version and was later revised to the current 60-item version. The test consists of 60 ink drawings of items ranging in familiarity (e.g., beaver, sphinx) (Lezak, 1995). The patient is shown each drawing and asked to provide the name for each object. The test begins at item 30 for adults. If any of the next eight items are failed, the patient is administered the items backward starting with item 29 until eight consecutive items are passed. Forward testing is then resumed

until six consecutive items are failed. If the patient cannot perceive what the object is or gives an indication that the object has been misperceived, a semantic cue is given (e.g., for beaver, “it’s an animal”). A phonemic cue (e.g., for beaver, “bea”) is given 20 seconds after the original presentation of the drawing or 20 seconds after the sematic cue was given (if it was necessary) if the patient is still unable to correctly name the item. The raw score is the total number of spontaneously correct responses plus the number of correct responses following a stimulus cue between the baseline and ceiling items. This number is then added to the number of items that precede the baseline. For this study, raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton, Grant, and Matthews (1991) external norms, as the normative data provided in the manual are sparse.

Psychometrics. Normative data for the BNT can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, the BNT was standardized on 1,000 individuals with a mean age of 50.3 years old (*SD* = 17.9). Mean level of education was 13.5 years (*SD* = 2.5). Slightly more than half of the sample was male (53.3%). The BNT has been found to have strong reliability and validity. Huff, Collins, Corkin, and Rosen (1986) divided the 85-item version into two equivalent forms. Internal consistency between the forms was found to be high, with a coefficient alpha of .96. Between-forms correlations of .81 in healthy adults and .97 in individuals with Alzheimer’s disease were also reported. Sawrie, Chelune, Naugle, and Luders (1996) found a strong test-retest reliability of .94 after 8 months in a sample of 51 patients with epilepsy. Thompson and Heaton (1989) found correlations to range from .92

to .96 between the 85-item, 60-item, and Huff et al.'s two, non-overlapping 42-item versions in 49 patients, providing further support for the reliability of the measure.

Examining construct validity, Axelrod, Ricker, and Cherry (1994) found that the BNT loaded highly on three major intelligence factors in adults: verbal comprehension, perceptual organization, and freedom from distractibility. Axelrod et al. also reported that the BNT has been found to have acceptable concurrent validity with the Visual Naming Test of the Multilingual Aphasia Examination (MAE; Benton, Hamsher, & Sivan, 1994a). Finally, correlations ranging from .74 to .87 have been found between the BNT and the Gates-McGinite Reading Vocabulary Test across normal and clinical adult populations, providing further evidence for the test's concurrent validity (Hawkins et al., 1993).

MAE Token Test. The Token Test is one of nine subtests of the MAE (Benton et al., 1994a). It is a 22-item oral language comprehension test that is an abbreviated and modified version of De Renzi and Vignolo's (1962) Token Test. There are two forms of this test (A and B) that consist of equivalent items (Benton et al., 1994a). Twenty small and large circles and squares in five colors (red, black, green, yellow, and white) are used to assess the patient's ability to comprehend and carry out simple tasks (Benton et al., 1994a). If the patient does not carry out the task correctly on the first trial, a second trial is implemented. The patient receives 2 points for correct responses on the first trial, and 1 point for correct responses on the second trial. The raw score is converted to a standard score and percentile based on control norms (patients without aphasia) provided in the manual.

Psychometrics. As part of the MAE standardization sample, the Token Test was normed on 360 individuals aged 16 to 69 years whose native language was English

(Benton et al., 1994a). No individual showed evidence or history of hemispheric brain disease. Individuals with a history of mental retardation or a psychiatric hospitalization were excluded from the sample. Reliability data are not presented in the manual or in major neuropsychological texts (e.g., Lezak, 1995; Spreen & Strauss, 1998).

Regarding validity, the Token Test was found to effectively discriminate between 115 control and 48 aphasic subjects (Benton et al., 1994a). Only 3.5% of the control subjects were misclassified by normative cutoffs, and 85.4% of individuals with aphasia were correctly classified, supporting the ability of the test to effectively discriminate between normals and individuals with aphasia. Further supporting the validity of this test, patients with left temporal lobe epilepsy have been found to score lower than patients with right hemisphere impairment (Hermann, Seidenberg, Haltiner, & Wyler, 1992; Hermann & Wyler, 1988). The test has also been shown to be sensitive to delirium in nonaphasic patients (Lee & Hamsher, 1988). Finally, high frequency of naming errors, defective associative word finding, and impaired comprehension on this test was found to correlate with head injury severity in a sample of closed-head injured patients (Levin, Grossman, & Kelly, 1976; Levin, Grossman, Sarwar, & Meyers, 1981).

MAE Sentence Repetition Test. The Sentence Repetition Test is also one of the nine MAE subtests. It consists of 14 sentences of progressively increasing length. The purpose of this test is to assess auditory verbal attention for sentences of increasing length. There are two forms of this test (A and B) that have equivalent difficulty levels (Benton et al., 1994a). The sentences in each form range from three to 18 words long. There are seven grammatical constructions (positive declaration, positive interrogative, imperative, negative declaration, negative interrogative, compound, and complex) in each form. The

examiner reads the sentences aloud to the patient, who then is asked to repeat the sentence verbatim. Sentences cannot be repeated. The test is discontinued after four consecutive incorrect responses. The patient receives one point for each correct response. Total raw scores are corrected for education level and age. Corrected raw scores are then converted into standard scores and percentiles.

Psychometrics. The Sentence Repetition Test was normed on the same sample as the MAE Token Test (see Token Test psychometrics section for details) (Benton et al., 1994a). Reliability data are not presented in the manual or in major neuropsychological texts (e.g, Lezak, 1995; Spreen & Strauss, 1998). Regarding validity, this test was found to effectively discriminate between 115 control and 48 aphasic subjects (Benton et al., 1994a). Only 3.5% of the control subjects were misclassified by normative cutoffs, and 75% of individuals with aphasia were correctly classified, supporting the ability of the test to effectively discriminate between normals and individuals with aphasia. The test has also been found to have acceptable discriminant validity as indicated by a modest correlation with a test that measures visual naming abilities (MAE Visual Naming) (r = .39) (Benton et al., 1994a).

MAE Controlled Oral Word Association (COWA). The COWA is one of nine subtests of the MAE. It is an oral fluency test in which the patient is required to make verbal associations to different letters of the alphabet by saying all the words he or she can think of that begin with a certain letter in 60 seconds (Benton et al., 1994a). The patient is asked to name ordinary words. Proper names (e.g., Bob, Boston), different forms of the same word (e.g., eat, eating), and substantives derived from previously given verbs or adjectives (e.g., fun, funny) are prohibited. The patient is presented with three

letters of progressively increasing associative difficulty (C, F, L). The difficulty level of each letter was defined by the frequency of words beginning with that letter found in standard English dictionaries. There are two forms of this test (A and B). Form A consists of the letters CFL. Form B consists of the letters PRW. The raw score consists of the number of acceptable responses for the three letters. The raw score is converted into an age- and education-corrected standard score and percentile based on control norms (patients without aphasia) provided in the manual.

Psychometrics. The COWA was normed on the MAE standardization sample (see Token Test psychometrics section for details) (Benton et al., 1994a). Reliability data are not presented in the manual or in major neuropsychological texts (e.g., Lezak, 1995; Spreen & Strauss, 1998); however, the test has been found to have adequate validity. For example, the two forms of this test were found to strongly correlate with each other (r = .82) (Benton et al., 1994a). The COWA was found to adequately discriminate between 115 control and 48 aphasic subjects (Benton et al., 1994a). Only 7% of control subjects were misclassified by the normative cutoffs, and 70.8% of individuals with aphasia were correctly classified. A modest correlation (r = .56) was found between COWA and MAE Visual Naming, reflecting the word retrieval aspect of both tasks (Benton et al., 1994a). On the other hand, a weaker correlation between COWA and Sentence Repetition was found (r = .34), indicating that both tests measure different constructs (Benton et al., 1994a).

Judgment of Line Orientation (JLO). The JLO (Benton, Hannay, & Varney, 1975) is a measure of spatial perception. It assesses one's ability to estimate angular relationships between line segments by visually matching angled line pairs to 11

numbered radii forming a semicircle (Lezak, 1995). It consists of 30 items each presenting a different pair of angled lines to be matched to the display cards (Lezak, 1995). The two forms, H and V, consist of the same items presented in a different order. Raw scores are corrected for age and sex. Standard scores and percentiles for score ranges are provided.

Psychometrics. The JLO was normed on 137 normal or control subjects, divided into six age-sex groups (Benton, Varney, & Hamsher, 1978). This test has been found to have acceptable levels of reliability and validity. Corrected split-half reliability of Form H was found to be .94 in a sample of 40 subjects; it was found to be .89 for Form V in a sample of 124 subjects (Benton, Sivan, Hamsher, Varney, & Spreen, 1994b). A test-retest reliability coefficient of .90 was found in a sample of 37 subjects who were administered both forms of the test (Benton et al., 1994b). Studies have found that patients with right hemisphere dysfunction consistently perform worse than normals and patients with left hemisphere impairment (Benton et al., 1994b; Levick, 1982; Trahan, 1991). These findings support the assumption that there is an association between impaired performance and right hemisphere dysfunction (Benton et al., 1994b), and that this test is a valid measure of spatial perception. Evidence of discriminant validity has been demonstrated by a weak partial correlation coefficient (r = .27) between the JLO and Facial Recognition Test (FRT; Benton & Van Allen, 1968).

Facial Recognition Test (FRT). The FRT was developed to examine the ability to recognize and discriminate unfamiliar human faces without involving a memory component (Benton et al., 1994b; Lezak, 1995). There are two forms of this test. The Long Form consists of 54 items. The Short Form (Levin, Hamsher, & Benton, 1975) was

developed to reduce administration time and consists of 27 items. The Short Form was administered to patients in this study. The test consists of three parts. Part one consists of matching of identical front-view photographs. The patient is presented with a single front-view photograph of a face and asked to identify it (by pointing to it or saying its number) in a display of six front-view photographs appearing below the single photograph. Three male and three female faces are presented for matching, requiring a total of six responses. Part two consists of matching front-view with three-quarter view photographs. The patient is presented with a single front-view photograph of a face and asked to locate it three times in a display of six three-quarter views, three being of the presented target face and three being of other faces. In the Short Form, one male and three female faces are presented, requiring a total of 12 responses. Part three consists of matching front-view photographs under different lighting conditions. The patient is presented with a single front-view photograph of a face taken under full lighting conditions and instructed to locate it three times in a display of six front views taken under different lighting conditions; three photographs in the display are of the presented face and three are of other faces. In the Short Form, two male faces and one female face are presented, requiring a total of nine responses. Short Form raw scores are converted to Long Form raw scores, and are then corrected for age and education. Corrected Long Form scores are then converted into standard scores and percentiles.

Psychometrics. The FRT was normed on 286 individuals aged 16 to 74 years (Benton et al., 1994b). One sample consisted of 196 neurological, neurosurgical, and medical patients from the University of Iowa Hospitals. The second sample consisted of 90 normal individuals aged 60 to 74 years who had volunteered to participate in a study

about aging. Differences in age and education level were found to affect test performance to a modest degree, which resulted in age and education score corrections.

Studies have found the FRT to have acceptable levels of reliability and validity. Test-retest reliability after one year in a sample of elderly control subjects was .60 (Levin, Llabre, & Reisman, 1991). Correlations between the Long and Short forms have been found to be strong, ranging from .88 in normals to .92 in patients with brain damage (Benton et al., 1994b). Internal consistency has been found to be moderate, with a coefficient alpha of .57; however, after omitting the first six items of the FRT, it improved to .66 (Hoptman & Davidson, 1993). Studies have found that patients with right parietal lesions performed worse than patients with right temporal lesions (Dricker, Butters, Berman, Samuels, & Carey, 1978; Warrington & James, 1967). Patients with right-hemisphere strokes have been found to score in the lowest percentile range (Egelko et al., 1988). Patients with left hemisphere lesions, who were not aphasic or who were aphasic without comprehension deficits, were have been found to perform similar to normal controls (Hamsher, Levin, & Benton, 1979). These findings suggest that the FRT is a valid visuospatial measure.

Hopkins Verbal Learning Test–Revised (HVLT–R). The HVLT–R (Brandt & Benedict, 2001) is a word-list learning and memory test intended primarily for use with brain-disordered populations. There are six forms of the HVLT–R, each of which consists of a list of 12 nouns. Four items on each list are drawn from three semantic categories that vary across the six forms. The patient is read the word list and then asked to recall as many words as possible, in any order. This process is repeated two more times for a total of three learning trials. After a 20 to 25 minute delay, a delayed recall trial is

administered. A recognition trial follows where the patient is read a list of 24 randomly ordered words (12 targets and 12 foils). The patient is asked to identify as many of the target words as possible with a *yes* response and to respond *no* to foils. Four scores can be computed: Recall Total, Delayed Recall, % Retention, and Recognition/Discrimination. Raw scores are converted into age-corrected T-scores.

Psychometrics. The HVLT–R normative sample consisted of 1,179 healthy individuals aged 16 to 92 years (M = 59 years, SD = 18.62) (Brandt & Benedict, 2001). Years of education ranged from 2 to 20 years (M = 13.47 years, SD = 2.88). The HVLT–R has been found to have acceptable reliability, with test-retest coefficients for the Total Recall and Delayed Recall of .74 and .66, respectively (Benedict, Schretlen, Groninger, & Brandt, 1998). The six forms of the HVLT–R have been found to have similar psychometric properties with respect to recall trials (Benedict et al., 1998; Brandt & Benedict, 2001). Modest differences were present in recognition trials (Benedict et al., 1998). Concurrent validity has been found to be acceptable, with correlations between total learning and delayed recall on the HVLT–R and immediate and delayed recall on the WMS–R (Wechsler, 1987) LM of .75 and .77, respectively (Shapiro, Benedict, Schretlen, & Brandt, 1999). Weaker correlations were found between HVLT–R total learning and delayed recall and WMS–R VR immediate and delayed recall (rs = .54 and .69, respectively), evidencing moderate levels of discriminant validity (Shapiro et al., 1999). Shapiro et al. also found the HVLT–R's measures of new learning and delayed recall to load on a single factor distinct from measures of visual memory and general cognitive function, further supporting the test's discriminant validity.

Category Fluency Test. Animal naming is a frequently used category fluency task. This test is useful to assess fluency in patients who are not able to name many scorable words when administered abstract letter fluency tasks (e.g., COWA) (Lezak, 1995). In this test, the patient is asked to name as many animals as possible in 60 seconds. For this study, raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton et al. (1991) external norms.

Psychometrics. Normative data for the Category Fluency Test can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, the Category Fluency Test was standardized on 1,148 individuals with a mean age of 50 years old (*SD* = 19.2). Mean level of education was 13.8 years (*SD* = 2.5). Slightly more than half of the sample was male (52.4%). Although reliability data could not be found, this test has been found to discriminate between Alzheimer's patients and normal controls (Monsch, Bondi, Butters, & Salmon, 1992), and patients with dementia and those with depression (Hart, Kwentus, Taylor, & Hamer, 1988), better than a letter fluency task. Other studies have found that elderly control patients and patients with dementia named more animals than CFL words, further supporting its validity (Rosen, 1980; Monsch et al., 1992; Monsch et al., 1994).

Trail Making Test (TMT). The TMT (Partington & Leiter, 1949) is a test of visual conceptual and visuomotor tracking that was originally part of the Army Individual Test Battery (1944). It consists of two parts, Part A and Part B. In Part A, the patient is asked to draw lines connecting consecutively numbered circles (numbered 1 through 25). In Part B, the patient is asked to draw lines connecting consecutively numbered and

lettered circles by alternating between the two sequences. The patient must complete these two tasks as quickly as possible without lifting the pencil from the paper. The examiner immediately points out any errors and has the patient proceed from the point the mistake occurred. Trials are discontinued after 300 seconds. Scoring for each part is based on the completion time in seconds. For this study, raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton et al. (1991) external norms.

Psychometrics. Normative data for the TMT can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, the TMT was standardized on 1,212 individuals with a mean age of 46.6 years old (*SD* = 18.1). Mean level of education was 13.6 years (*SD* = 2.8). More than half of the sample was male (56.8%). Reliability coefficients have generally been found to range from .64 to .98 (Spreen & Strauss, 1998). Interrater reliability was found to be .94 for Part A and .90 for Part B (Fals-Stewart, 1991). Lezak (1982) found the reliability of Part A to remain high throughout three administrations to normal controls at 6- and 12-month intervals (*W* = .78), whereas the reliability of Part B was found to be lower (*W* = .67). Parts A and B have been found to moderately correlate with each other (*r* = .49), suggesting that they measure somewhat different constructs (Heilbronner, Henry, Buck, Adams, & Fogle, 1991). The TMT has been found to load on both a rapid visual search and visuospatial sequencing factor (des Rosiers & Kavanagh, 1987; Fossum, Holmberg, & Reinvang, 1992), as well as a cognitive set-shifting factor (Pontius & Yudowitz, 1980). The TMT has been found to correlate with an object-finding test and a hidden pattern

tests (*r*s ranging from .36 to .93) (Ehrenstein, Heister, & Cohen, 1982). Strong correlations were not found between verbal tests (e.g., Token Test), indicating acceptable discriminant validity.

Wisconsin Card Sorting Test (WCST). The WCST (Berg, 1948; Grant & Berg, 1948) is considered to be a measure of executive function that was developed to assess abstract reasoning ability and the ability to shift cognitive set (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The test consists of four stimulus cards and 128 response cards that depict figures of various forms (crosses, circles, triangles, or stars), colors (red, blue, yellow, or green), and numbers (one, two, three, or four). The four stimulus cards are placed before the patient. The patient is handed a deck of 64 response cards and told to match each consecutive card from the deck with one of the four stimulus cards. The patient is not told how to match the cards, but he or she is told each time whether his or her response is right or wrong. Once the patient has matched 10 consecutive cards correctly, the sorting principle is changed without warning. The test proceeds in this manner through a number of set shifts until six successful categories are completed or both decks are exhausted. There is an oral and a computer version of the WCST; the oral version was used in this study. Many scores can be derived from this test. For this study, Perseverative Responses and Total Number of Errors were calculated. Raw scores were converted to age- and education-corrected standard scores. The WCST computer scoring program was used to score all protocols.

Psychometrics. The WCST was normed on a sample of 899 normal individuals whose data was aggregated from six distinct samples (Heaton et al., 1993). The individuals ranged in age from 6 to 89 years. Level of education ranged from

Kindergarten to 20 years. The WSCT has been found to have acceptable reliability and validity. Interscorer and intrascorer reliability were found to be acceptable, with intraclass correlation coefficients ranging from .75 to .97 (Axelrod, Goldman, & Woodard, 1992). In a sample of 46 children and adolescents, generalizability coefficients were found to range from .39 to .72 (*m* = .57) for the WCST scores, indicating largely acceptable scale reliability (Heaton et al., 1993). Factor analytic studies have found evidence of acceptable construct validity. For example, Shute and Huertas (1990) found the perseverative error score to load on the factor defined by a measure of Piagetian formal operations. Another study found that the number of categories achieved loaded on the complex intelligence and planning-organization factors, and the error score loaded on the complex intelligence and planning-flexibility factors (Daigneault, Braun, Gilbert, & Proulx, 1988). This test has been found to be a valid measure of executive functioning in children, adolescents, and adults with neurological impairments. Patients with disorders such as epilepsy, multiple sclerosis, Parkinson's disease, and schizophrenia, as well as patients with structural brain lesions of other etiologies, have been found to perform in the impaired range when compared with normal adults (Heaton et al., 1993).

Grooved Pegboard Test. The Grooved Pegboard Test (Kløve, 1963; Matthews & Kløve, 1964) is a manipulative dexterity test (Lafayette Instrument, 1989). It consists of a small board containing a 5 x 5 set of slotted holes angled in varying directions. Each peg has a key-like ridge at each end requiring it to be rotated into position for correct insertion into a hole. The dominant hand trial is administered first. The patient is told to place the pegs one at a time in the board as fast as possible, using only their dominant hand, going across the board in rows from left to right. The non-dominant trial follows,

with the patient placing the pegs in the board from right to left as fast as possible. The examiner records the completion time in seconds for each trial. Trials may be discontinued after 5 minutes. For this study, two trials per hand were sometimes administered (e.g., if the non-dominant hand was faster than the dominant hand). In these cases, the fastest time for each hand was used as the final score. Raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton et al. (1991) external norms.

Psychometrics. Normative data for the Grooved Pegboard Test can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, the Grooved Pegboard Test was standardized on 1,482 individuals with a mean age of 46 years old (*SD* = 17.1). Mean level of education was 13.5 years (*SD* = 2.8). More than half of the sample was male (60.1%). The test has been found to have acceptable test-retest reliability (*r* = .82) (Kelland, Lewis, & Gurevitch, 1992). It has been found to be sensitive to general slowing due to medication (Lewis & Rennick, 1979) and progression of disease processes (Matthews & Haaland, 1979). It has also been shown to help in the identification of lateralized dysfunction (Haaland & Delaney, 1981).

Grip Strength Test. The Grip Strength Test, also known as Hand Dynamometer, is a test used to measure hand strength (Reitan & Davison, 1974). This test operates under the assumption that lateralized brain impairment may affect the strength of the contralateral hand (Lezak, 1995). For this test, the patient is asked to hold the upper part of the dynamometer in their dominant hand first, palm down. The patient is told to hold his or her arm down by his or her side, away from the body, and asked to squeeze the

dynamometer as hard as possible. Two trials for each hand are performed, alternating between hands. The raw score is the force exerted in kilograms for each hand averaged for the two trials (Lezak, 1995). For this study, raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton et al. (1991) external norms.

Psychometrics. Normative data for the Grip Strength Test can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, this test was standardized on 1,482 individuals with a mean age of 46 years old (*SD* = 17.1). Mean level of education was 13.5 years (*SD* = 2.8). More than half of the sample was male (60.1%). This test has been found to have acceptable reliability, with reliability coefficients ranging from .52 to .98 in normal and neurologically impaired subjects (Spreen & Strauss, 1998). Average test-retest reliability for men and women has been found to be high (*r* for men = .91; *r* for women = .94) (Reddon, Stefanyk, Gill, & Renney, 1985). Kelland et al. (1992) found a test-retest reliability of .98. Differences between hands have not been found to be highly reliable and may be influenced by variable motivation (Provins & Cunliffe, 1972). This measure has been found to be useful in discriminating between patients with epilepsy with left hemisphere dysfunction from those with right hemisphere dysfunction (Strauss & Wada, 1988), in differentiating patients with brain damage from normals (Spreen & Strauss, 1998), and in identifying brain lesion laterality (e.g., Dodrill, 1978).

Finger Tapping Test (FTT). The FTT (Reitan, 1969), also known as the Finger Oscillation Test, is a test of simple motor speed that is part of the HRNB. It is a widely

used test of manual dexterity and may be used to aid in the detection of lateralized impairment (Lezak, 1995). It consists of a tapping key attached to a counter that records the number of taps. The patient is asked to tap as fast as possible with their index finger, not moving their hand or arm. Five 10-second trials with brief rest periods between trials are administered for each hand, dominant hand first (Psychological Assessment Resources [PAR] Staff, 1992). Trials are administered until five consecutive trials produce scores within a five-tap range (PAR Staff, 1993). If this criterion is not met, additional trials are given. A maximum of 10 trials per hand is allowed. The examiner records the number of taps per each 10-second trial. A finger tapping score is computed for each hand. If five consecutive trials produced scores within a five-tap range, the mean number of taps for those trials is calculated (PAR Staff, 1993). If 10 trails were administered because the five-tap criterion was not met, the mean of all 10 trails is calculated (PAR Staff, 1993). For this study, raw scores were converted to age-, education-, sex-, and ethnicity-corrected T-scores using the Heaton et al. (1991) external norms.

Psychometrics. Normative data for the FTT can be found in Heaton et al. (1991). Participants in the Heaton et al. normative project were assessed in several studies over 25 years. Participants were from rural and urban areas across the U.S. As part of this project, the FTT was standardized on 1,212 individuals with a mean age of 46.6 years old (*SD* = 18.1). Mean level of education was 13.6 years (*SD* = 2.8). More than half of the sample was male (56.8%). Spreen and Strauss (1998) reported that reliability coefficients have generally been found to range from .58 to .93 for intraindividual comparisons of same-hand performances. Performance with each hand has been found to be relatively

stable over time, even with intervals up to 2 years between retest sessions (Spreen & Strauss, 1998). Intraindividual differences between dominant and nondominant hands have been found to be more variable (e.g., $r > .70$; Massman & Doody, 1996; Provins & Cunliffe, 1972; $r = .50$; Morrison, Gregory, & Paul, 1979). The FTT has been found to be sensitive to the presence and laterality of brain lesion, with worse performance usually found in the hand contralateral to the lesion (Spreen & Strauss, 1998).

Research Design

This study employed a retrospective, cross-sectional design. A retrospective design was chosen because the nature of this study is exploratory and because it would be difficult to recruit patients with epilepsy to undergo neuropsychological evaluation unless medically necessary. Further, all clinically relevant variables for this study were previously collected at the time of neuropsychological evaluation and available in patient charts. The study consisted of a chart review of pertinent medical history and neuropsychological data for patients with epilepsy who underwent neuropsychological evaluation at the Center for Neuropsychological Services from January 2007 through November 2012.

Consent for Research Participation

A waiver of informed consent was granted by the Institutional Review Boards (IRBs) of Aurora St. Luke's Medical Center and Marquette University for the following reasons. First, this study presented no more than minimal risk to participants: (1) identifiable information was removed from the database after data entry was completed, (2) data (e.g., demographic information, neuropsychological test scores) were already

present in patients' medical records, and (3) no additional evaluations or procedures were performed on participants. Second, due to the nature of the topic, it would have been unadvisable to contact participants to obtain consent. If this were done, participants would have had to be told about PVT and the impact that suboptimal performance can have on test scores. As many of the participants in this study were evaluated years ago, this information may have caused them unnecessary worry about their past performance and neuropsychological findings.

Human Subjects Consideration

All data were collected in accordance with the ethical principles of research outlined by the American Psychological Association, the IRB of Marquette University, and the Research Subject Protection Program (RSPP)/IRB of Aurora St. Luke's Medical Center. Precautions were taken to protect the identity of study participants and the confidentiality of data. Only the minimal demographic, medical, and neuropsychological data necessary for conducting this study were collected. These data were transferred from medical records to data coding packets. Each packet was assigned a unique identifier to protect patient identity and maintain the confidentiality of data. Unique identifiers were connected to patient names in a password-protected document that was deleted upon completion of data entry. Data coding packets were stored in a locked cabinet in the Center for Neuropsychological Services. Data were entered into a password-protected database. Results and documentation of results through manuscripts or presentations will maintain participant confidentiality.

Procedures

Marquette University's IRB and Aurora St. Luke's Medical Center RSPP/IRB granted approval for this study. All patients underwent neuropsychological evaluation at Aurora St. Luke's Center for Neuropsychological Services. January 2007 was chosen as the study's start date because that is when the Center started consistently administering the WMT. Pre-surgical patients had previously undergone video-EEG monitoring at Aurora St. Luke's Regional Epilepsy Center to attempt to classify their seizures. For the neuropsychological evaluation, patients first completed a clinical interview with Dr. Joseph Cunningham, neuropsychologist. They were next administered a comprehensive battery of neuropsychological tests and the WMT. Order of test administration was not controlled. Tests were administered according to standardized instructions per their respective manuals. Patients were administered the computer version of the WMT. The six main WMT subtests (IR, DR, CNS, MC, PA, and FR) were administered. Patients were not administered the LDFR subtest as it is optional. Neuropsychological tests were scored according to standardized procedures outlined in their respective manuals. Raw neuropsychological data were converted into standard, scaled, or T-scores relative to normative data. Scores were converted into percentiles using a psychometric conversion table. Please refer to the Measures section for information about the norms used to score each test.

WMT data. During data collection, a discrepancy in the interpretation of IR, DR, and CNS scores of 90 was discovered: Per the WMT manual, IR, DR, or CNS scores "*above* 90% correct" are interpreted as clear passes (Green, 2005, p. 9). IR, DR, or CNS scores "*between* 83% and 90%" should be interpreted with caution (Green, 2005, p. 9). However, per the WMT computer program printouts, scores of 90 on IR, DR, and CNS

are interpreted as “pass,” and not as “caution” as stated in the manual. Through an e-mail exchange, Dr. Green acknowledged this discrepancy in score interpretation and advised to use the interpretations provided in the manual (P. Green, personal communication, March 8, 2013). Therefore, Green’s (2005) normative recommended cutoffs were used to score the WMT. As described in Table 1, patients were categorized into one of three WMT performance groups:

Table 1
WMT Performance Group Criteria

WMT Performance Groups		
Optimal	*Suboptimal*	*GMIP*
IR, DR, and CNS scores > 90% (clear pass) MC scores >70% (non-warning range) PA scores > 50% (non-warning range)	At least one IR, DR, or CNS score between 83 and 90% (caution range) At least one IR, DR, or CNS score ≤ 82.5% (clear fail) MC scores ≤ 70% (warning range) PA scores ≤ 50% (warning range)	At least one IR, DR, or CNS score ≤ 82.5% (clear fail) and a ≥ 30 point difference between the mean of the easy subtests (IR, DR, and CNS) and the mean of the hard subtests (MC, PA, and FR)

Note. IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; MC = Multiple Choice; PA = Paired Associates; FR = Free Recall; GMIP = General Memory Impairment Profile.

Additionally, GMIP scores were calculated for all patients in order to explore the validity of the GMIP.

Neuropsychological data. For each patient, neuropsychological test scores were converted into *z*-scores based on standard or scaled scores. This was done to convert the scores to a common scale from which they could be compared. Doing so also allowed for the generalizability of results to patients with epilepsy outside of this sample. Unlike in Rohling et al. (2002), *z*-scores relative to patients within the sample were not calculated.

Rohling et al. used relative *z*-scores because their sample was extremely heterogeneous; as this sample was homogeneous, normative *z*-scores were preferred.

Following Rohling's Interpretive Method (RIM; Miller & Rohling, 2001) for analyzing neuropsychological data, each test's *z*-score was assigned to one of the following cognitive domains: (1) Verbal Functions (VF; language abilities, semantic knowledge, crystallized abilities); (2) Perceptual Organizational (PO; visuospatial abilities); (3) Executive Function (EF; cognitive flexibility, abstraction, problem solving); (4) Learning and Memory (LM; immediate, delayed, auditory, visual); (5) Attention and Concentration (AC; focus execute, sustained, span, shift, divided); (6) Processing Speed (PS; psychomotor speed); and (7) Manual Dexterity (MD; motor skills). These domains are commonly used in neuropsychological research (Miller & Rohling, 2001; Zakzanis, Leach, & Kaplan, 1999). Tests and their respective *z*-scores were categorized into domains based on factor analytic studies that have identified which tests load the highest on which domain (e.g., Ardila, Galeano, & Rosselli, 1998; Larrabee & Curtis, 1995; Leonberger, Nicks, Larrabee, & Goldfader, 1992), as well as based on theoretical and practice-related a priori classifications (e.g., Lezak, 1995). Groupings were as follows:

(1) VF (*n* = 6): WAIS–III (Vocabulary, Similarities, Information); BNT; MAE Token Test; MAE Sentence Repetition

(2) PO (*n* = 5): WAIS–III (Picture Completion, Block Design, Matrix Reasoning); JLO; FRT

(3) EF (*n* = 5): MAE COWA; Category Fluency Test; TMT Part B; WCST (Perseverative Responses, Total Number of Errors)

(4) LM (*n* = 13): WMS–III (LM I Recall Total, LM II Recall Total, LM I Thematic Total, LM II Thematic Total, LM II % Retention, VR I Recall Total, VR II Recall Total, VR % Retention, VR II Recognition Total); HVLT–R (Recall Total, Delayed Recall, % Retention, Recognition/Discrimination)

(5) AC (*n* = 4): WAIS–III (Arithmetic, Digit Span, Letter-Number Sequencing); TMT Part A

(6) PS (*n* = 2): WAIS–III (Digit-Symbol Coding and Symbol Search)

(7) MD (*n* = 6): Grooved Pegboard Test (bilateral); Grip Strength Test (bilateral); FTT (bilateral)

Reliability analyses were conducted to explore each domain's internal consistency. Coefficient alphas of each domain should be .70 or higher, indicating that the tests that make up the domain "hang together" (Nunnally, 1978). Results indicated that the domains possessed acceptable levels of reliability, with standardized coefficient alphas ranging from .75 (EF domain) to .93 (LM domain). For each domain, all corrected item-total correlations were greater than .30, indicating that the tests that constructed each domain correlated well with the overall domain (Field, 2009). The deletion of tests from their respective domains did not significantly increase any scale's reliability; therefore, no tests were removed from their original domains.

Next, per RIM, the mean of each domain was calculated, yielding a domain mean *z*-score for each domain. The mean domain *z*-scores were then averaged, yielding a domain test battery mean (DTBM; Miller & Rohling, 2001).

Power Analysis

As inferential statistics were employed (e.g., ANOVA, multiple regression, simple linear regression), power analyses were conducted to determine sample size estimates. A power analysis determines the minimum sample size necessary for a statistical test to find a significant difference when such a difference exists (the probability of rejecting the null hypothesis when it is false) (Cohen, 1988). For this study, power analyses were conducted for each statistical test. A power analysis was not carried our for research question 1, as inferential statistics were not required to answer that question. The following power analyses were carried out for research questions 2 through 5:

Research question 2. Six one-way ANOVAs were conducted. Cohen's *f* was set at .40 to detect a large effect (Cohen, 1992). It is hypothesized that there will be large effect sizes because the patients in each WMT performance group are expected, based on normative data, to have large differences in their scores on WMT subtests. For example, a large effect size is expected between optimal and suboptimal performance groups on the IR and DR subtests, given that a score of 90% on IR or DR (reflective of suboptimal performance) is more than two standard deviations below the normal adult control group mean on IR or DR, and more than one standard deviation below the mean IR or DR score in patients with severe TBI (Green, 2005). Further, a score of 82.5% (failure cutoff) on IR is 5.5 standard deviations below the normal adult control group mean on IR (Green, 2005). A score of 82.5% on DR is 6.7 standard deviations below the normal adult control group mean on DR (Green, 2005). These normative data served as rationale for using a large effect size in the power analysis. Per Bonferroni adjustment results, an alpha value of .008 was used to control for the multiple comparison problem that results from

conducting six ANOVAs. Power was set at the generally accepted level of .80 (Cohen, 1988). Number of groups was entered as three. Power analysis results indicated that a sample size of 96 was required to find statistically significant differences when such differences exist. Although the N was slightly lower (N = 81) than the suggested 96, it was decided that it would be better to have a robust study with data representative of what was necessary to conduct the study rather than a larger sample size with questionable data. Additionally, it would have difficult to recruit additional patients with epilepsy to participate, as they only undergo neuropsychological evaluation when medically necessary.

Research question 3. Multiple regressions using dummy coding were conducted. Cohen's f^2 was set at .35 to detect a large effect (Cohen, 1988). Rationale for the anticipated large effect sizes for each regression is based on existing research examining the relationship between PVT scores and neuropsychological test scores (see Constantinou et al., 2005; Green et al., 2001; Green et al., 2002; Green, 2007; Rohling et al., 2002; Stevens et al., 2008). Results of these studies found that PVT scores accounted for up to 54% of the variance in neuropsychological tests scores, with R^2 values ranging from .35 to .54). As such, it was hypothesized that there would be similarly large effect sizes in each regression and Cohen's f^2 was therefore set at .35. The alpha value was set at .05. Power was set at .80 (Cohen, 1988). For multiple regressions that included number of years of education as a covariate, number of predictors was entered as three. For multiple regressions that did not include the covariate, number of predictors was entered as two. Results indicated that a sample size of 36 was required for multiple regressions

with three predictors, and a sample size of 31 was required for multiple regressions with two predictors. As $N = 81$, there was sufficient power to conduct these analyses.

Research question 4. Two regressions were conducted; therefore, separate power analyses were carried out. First, a simple (bivariate) linear regression was conducted. In a bivariate linear regression the data analyzed are the same as in a correlational analysis; therefore, a bivariate normal model correlation power analysis was conducted. The hypothesis for this research question was one-tailed; therefore, number of tails was entered as one. Correlation ρ H1 was set at .50 to detect a large effect (Cohen, 1988). Rationalization for the large effect size is based on existing research examining the performance of patients with significant cognitive impairment (e.g., dementia) on the WMT, NV-MSVT, and MSVT (see Green, 2005; Green et al., 2009; Green et al., 2011; Henry et al., 2009; Howe et al., 2007; Howe & Loring, 2009; Singhal et al., 2009). Results of these studies indicated that a substantial number of patients who scored below failure cutoff on at least one of the easy PV subtests in the WMT, NV-MSVT, or MSVT, and had significant cognitive impairment as evidenced by neuropsychological test scores and clinical history, demonstrated a GMIP. Thus, for this study, it was hypothesized that if the GMIP indeed signifies the presence of significant cognitive impairment such as that seen in dementia, GMIP scores would predict, and explain a large proportion of the variance in, neuropsychological memory scores. As such, correlation ρ H1 was set at .50 to detect a large effect. The alpha value was set at .05. Power was set at .80 (Cohen, 1988). Correlation ρ H0 was set at 0. Results indicated that a sample size of 23 was required to detect a correlation coefficient of $r = .50$ in the sample. As $N = 81$, there was sufficient power to conduct this analysis.

The second analysis conducted was a multiple regression with a dichotomous predictor variable. Cohen's f^2 was set at .35 to detect a large effect (Cohen, 1988). Rationalization for the anticipated large effect size is based on existing research with the WMT and the MSVT (see Green, 2005; Green et al., 2009; Green et al., 2011; Howe et al., 2007; Howe & Loring, 2009). Results of these studies indicated that patients with severe cognitive impairment (e.g., dementia) tended to score below failure cutoff on at least one PV subtest and had a ≥ 30-point difference between the mean of the easy (PV) and hard (memory) WMT subtests (or a ≥ 20-point difference on the MSVT). Thus, these patients had GMIPs on the WMT or MSVT signifying significant cognitive impairment that was corroborated with clinical history and neuropsychological test performance. In this study, it therefore follows that patients with scores below failure cutoff on at least one PV subtest and GMIP scores ≥ 30 would likely have much lower neuropsychological memory scores than patients with GMIP scores < 30. As such, a large effect size was anticipated and Cohen's f^2 was set to .35 to detect a large effect. The alpha value was set at .05. Power was set at .80 (Cohen, 1988). Number of predictors was entered as two. Results indicated that a sample size of 31 was required. As $N = 81$, there was sufficient power to conduct this analysis.

Research question 5. Two hierarchical regressions were conducted. As the regressions had different numbers of predictors, separate power analyses were conducted. The first multiple regression had six predictors. Cohen's f^2 was set at .35 to detect a large effect (Cohen, 1988). The rationale for the anticipated large effect size is based on the WMT normative research that established the harder subtests (MC, PA, FR) as the memory subtests and the easier subtests (IR, DR, CNS) as the PV subtests (Green, 2005).

Therefore, it is hypothesized that the harder memory subtests will each explain a much larger proportion of the variance in GMIP score than will each of the PV subtests. As such, Cohen's f^2 was set at .35. The alpha value was set at .05. Power was set at .80 (Cohen, 1988). Number of predictors was entered as six. Results indicated that a sample size of 46 was required. As $N = 81$, there was sufficient power to conduct this analysis.

The second hierarchical regression had two predictors. Cohen's f^2 was set at .35 to detect a large effect (Cohen, 1988). It is hypothesized that there will be a large effect size because the subtests that comprise the WMT memory composite have been identified primarily as memory subtests (Green, 2005). If the GMIP is a valid indicator of significant cognitive impairment, then the WMT memory composite would be anticipated to explain a much larger proportion of the variance in GMIP score than would the PV composite. As such, Cohen's f^2 was set to .35. The alpha value was set at .05. Power was set at .80 (Cohen, 1988). Number of predictors was entered as two. Results indicated that a sample size of 31 was required. As $N = 81$, there was sufficient power to conduct this analysis.

Exploration of Covariates

For each research question, potential covariates of sex, race, age, and education level were explored by conducting a correlation analysis between each potential covariate and outcome variable (e.g., WMT subtest scores, GMIP scores, cognitive domain *z*-scores, DTBM). Results indicated significant positive correlations between education level and the VF, PO, AC, and PS cognitive domain *z*-scores, as well as the DTBM. As the majority of data violated the parametric assumption of normality, Spearman's correlation coefficients (r_ss) were computed and ranged from .24 to .30 (all ps < .05).

Thus, when indicated, education level was entered as a covariate in the multiple regressions that explored the relationship between WMT performance and neuropsychological test scores. Education level was centered around its grand mean because, in multiple regression, the intercept represents the value of the outcome when each of the predictors take a value of 0 (Field, 2009). In this case, a value of 0 would have been meaningless, as education level ranged from 8 to 18.

No significant relationships were found between education level and the LM, EF, or MD cognitive domain *z*-scores (all *p*s > .05); therefore, education level was not entered as a covariate in the multiple regressions with these outcome variables. No significant relationships were found between the other potential covariates (sex, race, age) and WMT subtest scores, GMIP scores, cognitive domain *z*-scores, or the DTBM; therefore, these variables were not considered to be covariates and were not included in analyses.

Data Analysis

Research question 1. What are the base rates of optimal, suboptimal, and GMIP performance as measured by the WMT?

Data analysis method. Frequencies were run to calculate the number of patients in each WMT performance group (optimal, suboptimal, and GMIP). Patients were categorized into groups based on the WMT normative criteria described in the Procedures section. Descriptives were run to obtain measures of central tendency for the six WMT subtests (IR, DR, CNS, MC, PA, and FR) in each WMT performance group.

Research question 2. Are there differences on WMT subtest scores among WMT groups?

Null hypothesis. There are no differences on WMT subtest scores among WMT groups.

Alternative hypothesis. Patients in the optimal performance group will have higher WMT subtest scores than patients in the suboptimal performance and GMIP groups.

Data analysis method. Six one-way ANOVAs were conducted to explore the difference in mean WMT subtest scores among the three groups. Multiple t-tests were carried out to explore significant differences among group means. A Bonferroni adjustment was performed to explore whether obtained differences were significant (α .05/6 = .008). Effect sizes (Cohen's *d*) were calculated in order to establish the magnitude of the differences in standard deviation units. Cohen's *d* was selected because it is an appropriate effect size for comparisons between two means (Cohen, 1988), and because it is preferred over Pearson's correlation coefficient *r* as a measure of effect size when group sizes are discrepant (McGrath & Meyer, 2006).

Research question 3. Are there differences in neuropsychological test scores among WMT groups?

Null hypothesis. WMT performance will not predict neuropsychological test scores.

Alternative hypothesis. Patients with the highest WMT performance (optimal performance group) will have higher neuropsychological test scores than patients in the suboptimal performance and GMIP groups.

Data analysis method. Multiple regression using dummy coding was conducted. Multiple regression allows for the examination of the relationships involved when

multiple predictor variables (also known as independent variables) are related to a single outcome variable (also known as a dependent variable) (Cohen & Cohen, 1983). Dummy coding is a procedure that allows for the quantitative representation of a nominal (categorical) predictor variable that has more than two levels (Cohen & Cohen, 1983). Through dummy coding, a nominal predictor variable with multiple levels is transformed into a set of $g - 1$ different predictors (g = some number > 1), each representing one aspect of the distinctions among the g groups (Cohen & Cohen, 1983). Dummy coding uses dichotomous variables (zeros and ones) to convey group membership.

In this study, dummy coding was used to code the WMT performance predictor variable. This variable was a nominal predictor with three levels (GMIP, suboptimal performance, and optimal performance). Each level was transformed into separate predictors, or g groups, to represent distinctions in WMT performance. Each case was assigned to only one of the g groups (g = 3; GMIP, suboptimal performance, or optimal performance). As illustrated in Table 2, dummy code $g - 1$ variables were created (X_1 and X_2). X_1 represented "significant cognitive impairment-ness" and X_2 represented "suboptimalness." A "1" or "0" represented whether a patient was, or was not, a member of each $g - 1$ group. This information was not represented for one of the groups because doing so would have been redundant; a "0" for both X_1 and X_2, indicated that the patient was a member of the remaining optimal performance group, which served as the reference group. This group was chosen as the reference group so that comparisons could be made between the neuropsychological test performance of patients who were considered to have performed optimally on the WMT and those who received suboptimal or GMIP ratings.

Table 2
Dummy Coding

Group	*WMT Performance Level*	X_1	X_2
G_1	GMIP	1	0
G_2	Suboptimal	0	1
G_3	Optimal	0	0

Note. GMIP = Genuine Memory Impairment Profile.

Multiple regression analyses were conducted following dummy coding. Seven multiple regressions with dummy coding were conducted to explore the relationship between WMT performance and each of the seven cognitive domain *z*-scores. Education level was entered as a covariate when indicated per results of the correlational analyses previously described. Each cognitive domain (VF, PO, EF, LM, AC, PS, and MD) *z*-score was entered as the outcome variable for its respective multiple regression. Forced entry (also known as Enter), the regression method in which all predictors are simultaneously forced into the model (Field, 2009), was used for the multiple regressions that did not include the covariate. Hierarchical regressions were conducted for the regressions that included the covariate. This regression method was used because correlational analyses results revealed significant relationships between the covariate and certain outcome variables (VF, PO, AC, and PS cognitive domain *z*-scores). Thus, the covariate was entered in the first block as a predictor and both dummy variables were entered (Forced entry) as predictors in the second block.

Partial regression coefficients (B_i), which indicate the amount and direction of net change in the outcome variable, expressed in units of the outcome variable, of a change in one unit of X_i, will be reported for each dummy variable (Cohen & Cohen, 1983). The effect of the partialling process on a dummy variable is to relate G_i to G_g; for example, to relate the GMIP groups' (G_1) VF scores to the optimal groups' (G_g; reference group) VF

scores. Regression coefficients for the covariate will also be reported when indicated. Zero-order correlation coefficients (*r*; also known as Pearson correlation coefficients), which represent the correlation between each predictor and each outcome variable, will be reported. The sign of *r* indicates the direction of the relationship, and with dummy variables, indicates whether the mean of the outcome variable, for example, in G_1, is larger (positive) or smaller (negative) than the mean of the outcome variable for nonmembers of G_1 (Cohen & Cohen, 1983). The square of each zero-order correlation coefficient (r^2) will also be reported, indicating the proportion of the variance in each outcome variable accounted for by each predictor. Partial correlation coefficients (*pr*) will be reported, representing the correlation between each predictor and outcome variable after common variance with other predictors has been removed from both the outcome and the predictor of interest (Stevens, 2003). In other words, a partial correlation coefficient represents the correlation between the outcome variable and a predictor, when the effects of the other predictors on both the predictor of interest and the outcome are held constant (Cohen & Cohen, 1983; Field, 2009). Squares of semi-partial (part) correlation coefficients (sr^2) will also be reported, indicating the amount by which R^2 (the proportion of the variance in an outcome variable accounted for by the predictor variables) would be reduced if X_i (in this case, either the GMIP or the suboptimal group) were omitted from the predictor variables (Cohen & Cohen, 1983). R^2 values for each multiple regression will be reported. Finally, as R^2 provides the proportion of the outcome variable accounted for in the sample and overestimates that proportion in the population, adjusted R^2 values will also be reported (Cohen & Cohen, 1983).

A multiple regression using dummy coding was next conducted to explore the relationship between WMT performance and the DTBM. The hierarchical method of regression was used as, previously noted, a correlational analysis revealed a significant relationship between the covariate and the DTBM. Thus, the covariate was entered in the first block as a predictor and both dummy variables were entered (Forced entry) as predictors in the second block. The DTBM was entered as the outcome. Regression coefficients will be reported, along with r, r^2, pr, sr^2, R^2, and adjusted R^2 values.

Research question 4. What is the relationship between GMIP scores and scores on neuropsychological memory tests?

Null hypothesis. GMIP scores will not predict memory test scores.

Alternative hypothesis. Patients who score below failure cutoff on at least one PV subtest and have GMIP scores ≥ 30 will have lower memory test scores than patients with GMIP scores < 30.

Data analysis method. Two regressions were conducted. In the first analysis, a simple linear regression was conducted. GMIP score was entered as the predictor and the LM cognitive domain *z*-score (which represents memory tests) was entered as the outcome variable. The regression coefficient will be reported, along with R^2, adjusted R^2, r, and r^2.

In the second analysis, a multiple regression with a dichotomous predictor variable was conducted. The continuous GMIP variable was transformed into a dichotomous categorical variable to define two distinct groups of GMIP performance: patients with GMIP scores < 30 ("Non-GMIP" group; coded as 0) and patients with scores below failure cutoff on at least one PV subtest and GMIP scores ≥ 30 ("GMIP"

group; coded as 1). The LM cognitive domain *z*-score was entered as the outcome variable. Regression coefficients will be reported, along with R^2, adjusted R^2, r, and r^2.

Research question 5: How much does each of the WMT subtests explain total GMIP score?

Null hypothesis. Each WMT subtest score will account for the same amount of variance in GMIP score.

Alternative hypothesis. Memory subtest (MC, PA, FR) scores will account for a greater proportion of the variance in GMIP score than will PV subtest (IR, DR, CNS) scores.

Data analysis method. A hierarchical regression was first conducted to examine how much variance in GMIP score was accounted for by each WMT subtest score. Each WMT subtest score was entered into its own block. As the latter subtests are considered the "memory subtests" of the WMT (Green, 2005), they were entered into the regression model first because they were hypothesized to have greater importance in predicting GMIP scores than were the PV subtests. As such, the FR score was entered into block one, PA score into block two, MC into block three, CNS into block four, DR into block five, and IR into block six. GMIP score was entered as the outcome variable. An R^2 value will be reported. An adjusted R^2 value, which adjusts the coefficient of determination (R^2) to account for the fairly large number of predictor variables in the regression model, will also be reported. Zero-order and partial correlation coefficients will be reported.

The second hierarchical regression examined how much the PV and the memory subtests explained total GMIP score. WMT PV and memory composites were first computed. The WMT PV composite score consisted of the average of the IR, DR, and

CNS scores. The WMT memory composite score consisted of the average of the MC, PA, and FR scores. Reliability analyses were conducted on the composites. Results indicated strong levels of reliability for both the PV composite ($\alpha = .90$) and the memory composite ($\alpha = .93$). For each composite, corrected item-total correlations were greater than .70, indicating that the subtests correlated well with their respective composites. The deletion of subtests from their respective composites did not significantly increase either composite's reliability; therefore, no subtests were removed.

After determining that the composites possessed adequate levels of reliability, a hierarchical regression was conducted. As noted above, the subtests that comprised the WMT memory composite have been identified primarily as memory tests (Green, 2005). Therefore, the WMT memory composite was hypothesized to have more importance in predicting the outcome (GMIP score) and was entered in block one. The WMT PV composite was entered in block two. GMIP score was entered as the outcome variable. Zero-order and partial correlation coefficients will be reported. R^2 and adjusted R^2 values will also be reported.

Sample Size

Review of available medical records identified 133 patients with epilepsy who were referred for neuropsychological evaluation between January 2007 and November 2012. Of these patients, 13 were seen for both pre- and post-surgical evaluations. As the inclusion criteria for this study specified that participants must be pre-surgical candidates or non-surgical, only pre-surgical data for these patients were used. Fourteen patients were excluded from the study because they underwent post-surgical neuropsychological evaluations only. Another fourteen patients were excluded because they were not

administered the WMT. Twenty-two patients were excluded because they were not administered the full battery of neuropsychological tests. One patient was not eligible because, although referred, did not go through with the evaluation. Another patient was not eligible because the WMT was discontinued. The remaining 81 patients were included for analyses.

Although smaller than anticipated, a sample size of 81 was sufficient to achieve the power necessary for the majority of analyses. As power analyses provide *estimates* of sample sizes necessary to detect effects, it was decided that a slightly smaller N sufficiently reflected a robust sample containing the necessary data required for each analysis. Further, considering the retrospective design of this study and that patients with epilepsy undergo neuropsychological evaluation only when medically necessary, recruiting additional patients would have posed both a challenge and a deviation from the study's design.

Data Screening

Accuracy of the data file. Prior to conducting data analyses, data were screened for errors and outliers. Frequencies for all variables were examined to ensure entered values were within appropriate ranges. For continuous variables, the plausibility of means, standard deviations, minimum, and maximums were examined. For categorical variables, responses not part of the item scale (e.g., a response of 6 on a scale that ranged from 1 to 5) were checked against respective patients' data coding packets and medical records. All errors were corrected according to the data in patients' medical records. After correcting data entry errors, values for continuous variables were plausible (e.g., age at time of neuropsychological evaluation ranged from 16 to 70 years old, $M = 39.98$, $SD = 14.28$),

and values for categorical variables fell within predefined ranges. An outlier was identified in the variable examining frequency of seizures per month: One patient reported experiencing 20 absence seizures, brief staring episodes with impairment of awareness and responsiveness (Devinsky, 2004), per day (600 seizures per month). Therefore, the median instead of the mean was reported for this variable, as the median is relatively unaffected by extreme scores at either end of the distribution (Field, 2009).

Missing and problematic data. After IRB approval, an examination of the data revealed that not all patients referred for evaluation between January 2007 and November 2012 were administered the same battery of neuropsychological tests. Two tests were identified as especially problematic for different reasons: The Brief Visuospatial Memory Test–Revised (BVMT–R; Benedict, 1997), a visual memory test, was only administered to 60 of the 81 participants (74%). The Wide Range Achievement Test (WRAT; Wilkinson, 1993; Wilkinson & Robertson, 2006) Word Reading subtest, which measures letter and word decoding through letter identification and word recognition, was administered to all participants; however, 21 patients (26%) were administered the WRAT-3 and 60 patients (74%) were administered the revised WRAT-4. Although the WRAT-3 and WRAT-4 Word Reading subtests inherently aim to measure the construct of verbal functioning, the tests are not identical and therefore theoretically represent the construct in a different way. Rather than dropping the patients who were not administered the BVMT-R and WRAT-4, these two tests were dropped from the study to maintain the study's *N*. Further support for dropping these tests was that the constructs they measure were well represented by other tests completed by all participants.

After dropping the BVMT-R and WRAT, only patients with complete neuropsychological and WMT datasets were included in the analyses. Although missing data is common in research and can occur for a variety of reasons (Field, 2009), methods used to deal with missing data (i.e., inputting the overall sample mean of the item, using a regression model to predict the missing value based on cases with complete data; Little & Rubin, 2001) were not implemented. Missing data imputation methods were not utilized in order to ensure that the data included in the analyses accurately reflected patients' genuine neuropsychological test performance, rather than estimations or predictions. As a result, 22 patients were excluded from the study because they were not administered the full battery of neuropsychological tests, and therefore were missing significant amounts of the neuropsychological data required to conduct the analyses. As previously noted, 14 additional patients were excluded because they were not administered the WMT. One patient was excluded because the WMT was discontinued.

A discrete missing value of "99" was entered to represent missing or unknown demographic data. Non-applicable demographic information was left blank (e.g., if a patient indicated that he/she was not employed, the "Employed" variable was coded "no," and the subsequent "Employment Status" variable was left blank).

CHAPTER IV: RESULTS

Preliminary Analyses

Exploring assumptions of parametric tests. There are four assumptions of parametric tests that were explored prior to conducting analyses: normally distributed data, homogeneity of variance, interval data, and independence (Field, 2009). Although the first assumption can mean different things for different statistical tests, it generally means that the sampling distribution is approximately normal. There are various ways to examine normality, including checking data visually, inspecting skewness and kurtosis values, and comparing the distribution to a normal distribution to see if it differs (Field, 2009).

For data that were used in analyses (education level, neuropsychological data, and WMT data), histograms and probability-probability (P-P) plots, skewness and kurtosis values, and Kolmogorov-Smirnov (K-S) test results of the overall distribution were examined to assess normality. Histograms, P-P plots, and skewness and kurtosis values revealed that the majority of data were not normally distributed. These results were further supported by K-S test results, which indicated that 71% of data used in analyses, *D*s(81) ranging from .10 to .24, *p*s < .05, were significantly non-normal. As research questions 2 and 3 involved comparing WMT performance groups, the K-S tests were re-conducted to investigate whether the distribution of data was normal in each group. Results revealed that data were more normally distributed within the three groups compared to the overall distribution. For the optimal performance group, 45% of data, *D*s(35) ranging from .15 to .37, *p*s < .05, were significantly non-normal. For the suboptimal performance group, 43% of data, *D*s(29) ranging from .16 to .29, *p*s < .05,

were significantly non-normal. For the GMIP group, 29% of data, *D*s(17) ranging from .21 to .39, *p*s < .05, were significantly non-normal. Overall, assessments of normality indicated that much of the data were significantly non-normal and thus violated the assumption of normality.

Homogeneity of variance is the assumption that the variance of one variable is stable (i.e., fairly similar) at all levels of another variable (Field, 2009). This assumption was explored for data in the three WMT performance groups. Levene's tests results revealed that for the majority of neuropsychological data (84%), the variances were equal for patients in the optimal, suboptimal, and GMIP groups, *F*s(2, 78) ranging from .02 to 3.00, *ns*. Variances were also equal across groups for the cognitive domain variables and the DTBM variable, *F*s(2, 78) ranging from .20 to 2.66, *ns*. However, for the WMT IR, DR, CNS, MC, and GMIP variables, variances were significantly different across the three groups, *F*s(2, 78) ranging from 3.53 to 19.82, *p*s < .05. The variances for three neuropsychological variables (Category Fluency Test, HVLT-R Delayed Recall, HVLT-R Recognition/Discrimination) also differed significantly across the three groups, *F*s(2, 78) ranging from 4.31 to 9.57, *p*s < .05. Overall, results indicated that the assumption of homogeneity of variance was not met for all variables across WMT performance groups.

The final two assumptions of parametric tests state that data should be measured at least at the interval level and should be independent (Field, 2009). In this study, data were measured at the interval level. As independence can mean different things depending on the test being conducted, independence as it relates to the assumptions of ANOVA and multiple regression is detailed below.

Approach to dealing with non-normality and unequal variances. Various transformations were explored to attempt to correct the largely non-normal dataset. Transforming data involves doing something to each piece of data to correct for distribution problems, unequal variances, or outliers (Field, 2009). Specific transformations can be performed to correct for skewness and unequal variances; however, the necessity and usefulness of performing transformations is a complex, debatable issue and depends on the robustness of the statistics performed (Field, 2009). Robustness is the ability of a test estimate statistics that are reliable even when the normal assumptions of the statistic are not met (Field, 2009). For this study, log and square root transformations were explored to correct for the positive skew and unequal variances of some of the variables. The transformations were not helpful, as some data that were initially normal became non-normal and some data remained non-normal despite the transformation. As ANOVA and multiple regression are considered fairly robust tests, data were not transformed.

Sample Characteristics

Sample demographics. Eighty-one patients met inclusion criteria and were included in statistical analyses. As expected, the majority of patients (69%) were pre-surgical, indicating that they underwent neuropsychological evaluation to aid in the determination of their candidacy for epilepsy surgery. Patients ranged in age from 16 to 70 years old (M = 39.98, SD = 14.28). The sample consisted of approximately equal numbers of males (n = 39) and females (n = 42), the majority of whom identified as Caucasian (86%). Most patients were either married (48%) or single (42%). Education level ranged from 8 to 18 years (M = 12.62, SD = 2.15). Slightly more than half of

patients identified as unemployed (53%). It is likely that some of these patients were receiving Social Security Disability Insurance (SSDI) because of their epilepsy; in fact, 32% of patients reported that they were receiving SSDI at the time of the evaluation, while 30% of patients reported that they were not. Slightly more than one-third of patients (38%) did not disclose SSDI status. The majority of patients were right-handed (88%), and 57% denied a family history of sinistrality (left-handedness). Mean WAIS-III FSIQ score was in the high end of the low average range (M = 89.59, SD = 13.87). Mean WAIS-III Verbal IQ was also in the high end of the low average range (M = 89.96, SD = 13.39). Mean WAIS-III Performance IQ was in the low end of the average range (M = 90.67, SD = 14.30) Table 3 provides more detailed demographic information.

Table 3
Sample Demographics (N = 81)

Variable	*n*	*%*
Patient Status		
Pre-surgical	56	69.1
Non-surgical	25	30.9
Gender		
Female	42	51.9
Male	39	48.1
Age at NP Evaluation		
16-19	6	7.4
20-29	18	22.2
30-39	15	18.5
40-49	18	22.2
50-59	17	21.0
60-69	6	7.4
70-79	1	1.2
Race		
Caucasian	70	86.4
African American	5	6.2
Hispanic	5	6.2
Asian	1	1.2
Marital Status		
Married	39	48.1
Never Married	34	42.0
Divorced	7	8.6

Widowed	1	1.2
Employment Status		
Unemployed	43	53.1
Employed	31	38.3
Student	7	8.6
Education Level		
0-8 years	3	3.7
9-11 years	11	13.6
12 years	36	44.4
13-15 years	19	23.5
16-17 years	10	12.3
18-20 years	2	2.5
Handedness		
Right-handed	71	87.7
Left-handed	8	9.9
Ambidextrous	2	2.5

Note. NP = Neuropsychological.

Epilepsy characteristics. Per inclusion criteria, all participants had been diagnosed with epilepsy by a board certified neurologist. Age of seizure onset ranged from birth to 65 years old (M = 22.77, SD = 1.90). Duration of seizure disorder ranged from 0 to 63 years (M = 16.96, SD = 1.75). The median number of seizures experienced by patients per month was 3.50. The median is reported for this variable to account for an outlier (600 absence seizures per month), as the median is relatively unaffected by extreme scores at either end of the distribution (Field, 2009).

Data regarding the side of seizure onset were available for approximately 83% (67/81) of the sample; of those, 26 had seizures arising from the right hemisphere, 32 from the left hemisphere, and 9 had bilateral seizure foci. Approximately 28% of patients were diagnosed with left temporal lobe epilepsy (TLE) and 24% were diagnosed with right TLE. Data regarding number of seizure types were available for approximately 96% (78/81) of the sample; of those, 55% experienced two or three types of seizures, with complex partial/partial generalized/generalized the most frequently endorsed combined

seizure type. Forty-five percent of patients experienced one seizure type – simple partial, complex partial, partial generalized, or generalized. Data regarding most common seizure type were available for 95% (77/81) of the sample. Complex partial seizures were the most common type of partial seizure, experienced by 53% of patients. Tonic-clonic seizures were the most common type of generalized seizure, experienced at least once by 62% of the sample. Most patients (74%) never experienced status epilepticus. The majority of patients (64%) denied a family history of seizures. All but two patients were taking antiepileptic drugs (AEDs) at the time of the evaluation, with the majority (94%) taking one to three AEDs.

Statistical Analyses

Research question 1. What are the base rates of optimal, suboptimal, and GMIP performance as measured by the WMT?

Base rates of WMT performance. Frequencies were conducted to calculate the number of patients in each WMT performance group. Thirty-five patients (43%) were in the optimal performance group and 29 patients (36%) were in the suboptimal performance group. Seventeen (21%) patients were in the GMIP group, indicating that they likely performed poorly on the WMT because of significant cognitive impairment and *not* because of suboptimal performance. Means and standard deviations of the WMT subtests and GMIP score for the WMT performance groups are presented in Table 4.

Table 4
Optimal, Suboptimal, and GMIP Groups' Performance on WMT (*N* = 81)

WMT Group	*IR*	*DR*	*CNS*	*MC*	*PA*	*FR*	*GMIP*
Optimal							
M	98.57	98.57	97.43	92.14	87.14	56.07	19.23
SD	1.94	1.85	2.81	7.50	12.08	13.44	9.13
Suboptimal							
M	90.78	92.33	86.38	72.41	65.69	39.83	32.21
SD	11.61	3.53	10.87	14.12	14.06	12.97	11.50
GMIP							
M	78.68	79.12	72.21	39.71	38.82	20.29	43.47
SD	12.53	7.12	9.10	14.52	10.83	9.01	5.78

Note: IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; MC = Multiple Choice; PA = Paired Associates; FR = Free Recall; GMIP = Genuine Memory Impairment Profile.

As can be seen in Table 4, patients in the WMT optimal performance group scored in the "clear pass" range (IR, DR, and CNS scores > 90%) on all WMT PV subtests, and in the "non-warning range" (MC scores > 70% and PA scores > 50%) on the memory subtests. Patients in this group scored higher across all WMT subtests, and obtained lower GMIP scores (suggesting that WMT performance was not negatively impacted by significant cognitive impairment), than patients in the suboptimal and GMIP groups. Performance in the suboptimal group was, on average, characterized by "cautionary" (at least one IR, DR, or CNS score between 83 and 90%) rather than "failure" (at least one IR, DR, or CNS score ≤ 82.5%) scores across the PV subtests (IR, DR, and CNS). Patients in the suboptimal group also averaged GMIP scores slightly higher than the ≥ 30 cutoff; however, per the WMT manual, such scores would indicate a GMIP *only if* they had *failed* one of the PV subtests. Patients who failed one of the first three PV subtests and had a ≥ 30 point difference between the mean of the easy subtests (IR, DR, and CNS) and the mean of the hard subtests (MC, PA, and FR) were placed into the GMIP group, thereby assuring that each group was mutually exclusive. Patients in the

GMIP group averaged scores in the "failure" range across the PV subtests and in the "warning" range on the memory subtests (MC and PA). Patients in this group also had the highest GMIP scores, suggesting that their performance on the WMT was negatively impacted by significant cognitive impairment.

Research question 2. Are there differences on WMT subtest scores among WMT groups?

***Assumptions of ANOVA*.** The assumptions of ANOVA under which the *F*-statistic is reliable are the same as those previously described for all parametric tests: normally distributed data, homogeneity of variance, interval data, and independence (Field, 2009). For the current analyses, the last two assumptions were met: the outcome variables (WMT subtest scores) were measured on an interval scale and all observations were independent of each other. However, as described below, the first two assumptions were not met.

In ANOVA, normality refers to whether the distribution of data is normal within groups (Field, 2009). K-S tests were conducted to investigate normality. For the optimal performance group, WMT subtest scores, *D*s(35) ranging from .15 to .37, *p*s < .05, were significantly non-normal. For the suboptimal performance group, WMT subtest scores, *D*s(29) ranging from .12 to .29, *p*s < .05, were significantly non-normal. For the GMIP group, the FR score, $D(17) = .24$, $p < .05$, was significantly non-normal; all other WMT subtest scores for this group were normally distributed. Overall, results indicated that WMT subtest score distributions were significantly non-normal in the optimal and suboptimal performance groups, thereby violating the assumption of normality. Except for the FR score, WMT subtest scores were normally distributed in the GMIP group.

Exploratory transformations were conducted to attempt to correct non-normal data. Despite the transformations, WMT scores remained non-normal; therefore, data were not transformed. Although the *F*-statistic in ANOVA can be robust to violations of normality when group sizes are equal (Field, 2009), in this study, group sizes were unequal. Consequently, the accuracy of the *F*-statistic was likely impacted by skew, and the power of *F* may have been affected by non-normality (Field, 2009).

In ANOVA, homogeneity of variance refers to whether the variances in each group are fairly similar (Field, 2009). This assumption was explored for WMT subtest scores in the groups. For IR, DR, CNS, and MC scores, the variances were significantly different for patients in the optimal, suboptimal, and GMIP groups, *F*s(2,78) ranging from 4.46 to 19.82, *p*s < .05. However, for PA and FR scores, the variances were equal across groups, *F*s(2,78) ranging from .95 to .99, *ns*. Overall, results indicated that the variances of the majority of WMT subtest scores were significantly different across groups, and therefore the assumption of homogeneity was largely violated. Exploratory transformations (i.e., log, square root, 1/square root, reciprocal) were conducted to attempt to correct for heterogeneity of variances. The transformations were not particularly helpful; in fact, transformations caused heterogeneity of variances for scores that previously had similar variances (e.g., conducting a reciprocal transformation resulted in the variances for PA and FR scores becoming significantly different across groups). As such, WMT data were not transformed.

While ANOVA is considered robust to violations of homogeneity of variance when samples sizes are equal (Field, 2009), it was not considered robust to such violations in this study because sample sizes were unequal among WMT groups. Since

the optimal group – the group with the largest sample size – had smaller variances in most WMT subtest scores than the other groups with smaller sample sizes, the *F*-ratio may produce a significant result when there is no difference between groups (Field, 2009). To rectify the increased possibility of making a Type I error, Welch's *F*, an alternative *F*-ratio designed to be robust to violations of homogeneity of variance, is reported for the ANOVAs that violated the assumption homogeneity of variance.

Differences on WMT subtest scores among groups. Six one-way ANOVAs were conducted to test for differences in mean WMT subtest scores among the three groups. The first ANOVA indicated that WMT IR scores differed significantly across groups, *Welch's F*(2, 27.99) = 26.55, $p < .001$. Post hoc comparisons using Bonferroni adjusted alpha levels of .008 (.05/6) indicated that the optimal group ($M = 98.57$, $SD = 1.94$) had significantly higher mean IR scores than the suboptimal ($M = 90.78$, $SD = 11.61$), $p = .003$, Cohen's $d = .94$, and GMIP groups ($M = 78.68$, $SD = 12.53$), $p < .001$, Cohen's $d = 2.22$. Post hoc tests also revealed that the suboptimal group ($M = 90.78$, $SD = 11.61$) had significantly higher mean IR scores than the GMIP group ($M = 78.68$, $SD = 12.53$), $p < .001$, Cohen's $d = 1.00$. Results indicated that, in general, patients in the optimal group had significantly higher IR scores than did patients in the suboptimal and GMIP groups. Mean IR scores for the optimal and suboptimal groups differed by nearly one standard deviation, indicating a large difference in means. Mean IR scores for the optimal and GMIP groups differed by more than 2 standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher IR scores than did patients in the GMIP group. Mean IR scores for the suboptimal and GMIP groups differed by one standard deviation, representing a large difference in means.

The second ANOVA showed that WMT DR scores differed significantly across groups, *Welch's F*(2, 31.54) = 89.70, $p < .001$. Post hoc comparisons using the Bonferroni adjustment indicated that the optimal group ($M = 98.57$, $SD = 1.85$) had significantly higher mean DR scores than the suboptimal ($M = 92.33$, $SD = 3.53$), $p < .001$, Cohen's $d = 2.31$, and GMIP groups ($M = 79.12$, $SD = 7.12$), $p < .001$, Cohen's $d = 3.74$. Post hoc tests also revealed that the suboptimal group ($M = 92.33$, $SD = 3.53$) had significantly higher mean DR scores than the GMIP group ($M = 79.12$, $SD = 7.12$), $p < .001$, Cohen's $d = 2.35$. Results indicated that, in general, patients in the optimal group had significantly higher DR scores than did patients in the suboptimal and GMIP groups. Mean DR scores for the optimal and suboptimal groups differed by more than two standard deviations, indicating a large difference in means. Mean DR scores for the optimal and GMIP groups differed by more than three and a half standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher DR scores than did patients in the GMIP group. Mean DR scores for the suboptimal and GMIP groups differed by more than two standard deviations, representing a large difference in means.

The third ANOVA revealed that WMT CNS scores differed significantly across groups, *Welch's F*(2, 29.69) = 72.35, $p < .001$. Post hoc comparisons using the Bonferroni adjustment indicated that the optimal group ($M = 97.43$, $SD = 2.81$) had significantly higher mean CNS scores than the suboptimal ($M = 86.38$, $SD = 10.87$), $p < .001$, Cohen's $d = 1.39$, and GMIP groups ($M = 72.21$, $SD = 9.10$), $p < .001$, Cohen's $d = 3.74$. Post hoc tests also revealed that the suboptimal group ($M = 86.38$, $SD = 10.87$) had significantly higher mean CNS scores than the GMIP group ($M = 72.21$, $SD = 9.10$),

$p < .001$, Cohen's $d = 1.41$. Results indicated that, in general, patients in the optimal group had significantly higher CNS scores than did patients in the suboptimal and GMIP groups. Mean CNS scores for the optimal and suboptimal groups differed by more than one standard deviation, indicating a large difference in means. Mean CNS scores for the optimal and GMIP groups differed by more than three and a half standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher CNS scores than did patients in the GMIP group. Mean CNS scores for the suboptimal and GMIP groups differed by nearly one and a half standard deviations, representing a large difference in means.

The fourth ANOVA indicated that WMT MC scores differed significantly across groups, *Welch's* $F(2, 34.29) = 107.36$, $p < .001$. Post hoc comparisons using the Bonferroni adjustment indicated that the optimal group ($M = 92.14$, $SD = 7.50$) had significantly higher mean MC scores than the suboptimal ($M = 72.41$, $SD = 14.12$), $p < .001$, Cohen's $d = 1.75$, and GMIP groups ($M = 39.71$, $SD = 14.52$), $p < .001$, Cohen's $d = 4.54$. Post hoc tests also revealed that the suboptimal group ($M = 72.41$, $SD = 14.12$) had significantly higher mean MC scores than the GMIP group ($M = 39.71$, $SD = 14.52$), $p < .001$, Cohen's $d = 2.28$. Results indicated that, in general, patients in the optimal group had significantly higher MC scores than did patients in the suboptimal and GMIP groups. Mean MC scores for the optimal and suboptimal groups differed by more than one and a half standard deviations, indicating a large difference in means. Mean MC scores for the optimal and GMIP groups differed by four and a half standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher MC scores than did patients in the GMIP group. Mean MC scores for

the suboptimal and GMIP groups differed by more than two standard deviations, representing a large difference in means.

The fifth ANOVA showed that WMT PA scores differed significantly across groups, $F(2, 78) = 86.06$, $p < .001$. Post hoc comparisons using the Bonferroni adjustment indicated that the optimal group ($M = 87.14$, $SD = 12.08$) had significantly higher mean PA scores than the suboptimal ($M = 65.69$, $SD = 14.06$), $p < .001$, Cohen's $d = 1.64$, and GMIP groups ($M = 38.82$, $SD = 10.83$), $p < .001$, Cohen's $d = 4.21$. Post hoc tests also revealed that the suboptimal group ($M = 65.69$, $SD = 14.06$) had significantly higher mean PA scores than the GMIP group ($M = 38.82$, $SD = 10.83$), $p < .001$, Cohen's $d = 2.14$. Results indicated that, in general, patients in the optimal group had significantly higher PA scores subtest than did patients in the suboptimal and GMIP groups. Mean PA scores for the optimal and suboptimal groups differed by more than one and a half standard deviations, indicating a large difference in means. Mean PA scores for the optimal and GMIP groups differed by more than four standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher PA scores than did patients in the GMIP group. Mean PA scores for the suboptimal and GMIP groups differed by more than two standard deviations, representing a large difference in means.

The final ANOVA revealed that WMT FR scores differed significantly across groups, $F(2, 78) = 48.24$, $p < .001$. Post hoc comparisons using the Bonferroni adjustment indicated that the optimal group ($M = 56.07$, $SD = 13.44$) had significantly higher mean FR scores than the suboptimal ($M = 39.83$, $SD = 12.97$), $p < .001$, Cohen's $d = 1.23$, and GMIP groups ($M = 20.29$, $SD = 9.01$), $p < .001$, Cohen's $d = 3.13$. Post hoc

tests also revealed that the suboptimal group (M = 39.83, SD = 12.97) had significantly higher mean FR scores than the GMIP group (M = 20.29, SD = 9.01), $p < .001$, Cohen's d = 1.75. Results indicated that, in general, patients in the optimal group had significantly higher FR scores than did patients in the suboptimal and GMIP groups. Mean FR scores for the optimal and suboptimal groups differed by more than one standard deviation, indicating a large difference in means. Mean FR scores for the optimal and GMIP groups differed by more than three standard deviations, reflecting a large difference in means. On average, patients in the suboptimal group had significantly higher FR scores than did patients in the GMIP group. Mean FR scores for the suboptimal and GMIP groups differed by more than one and a half standard deviations, representing a large difference in means.

Research question 3. Are there differences in neuropsychological test scores among WMT groups?

Assumptions of multiple regression. To apply a regression model from a sample to a population of interest, several assumptions must be met. Assumption 1 states that predictor variables must be quantitative or categorical, and that the outcome variable must be quantitative, continuous, and unbounded (Field, 2009). Quantitative predictors should be measured at the interval level and categorical variables must have two categories (Field, 2009). In this study, quantitative predictors were measured at the interval level. As the categorical predictor had three categories, dummy coding was used. Outcome variables were unbounded, which means that there were no restrictions on the variability of the outcome (Field, 2009).

Assumption 2 states that the predictors should have some variation in value (Field, 2009). All predictors had variances greater than 0, signifying that Assumption 2 was met.

Assumption 3 indicates that there should be no perfect multicollinearity, which means that there should be no perfect linear relationship between two or more of the predictors (Field, 2009). In other words, the predictors should not correlate too strongly. If predictors correlate too strongly, or perfectly, it becomes nearly impossible to obtain unique estimates of the regression coefficients because there are a limitless number of coefficient combinations that would work equally well (Field, 2009).

Multicollinearity can be identified through different methods. One method is to scan a correlation matrix of the predictor variables and see if any correlate very highly ($\geq .80$) (Field, 2009). Scanning a correlation matrix of predictors revealed correlation coefficients ranging from -.39 to .04, indicating that there were no strong correlations between predictors. As this method may miss more subtle forms of multicollinearity, two collinearity diagnostic statistics should also be checked: the variance inflation factor (VIF) and the tolerance statistic (Field, 2009). The VIF shows whether a predictor has a strong linear relationship with the other predictor(s) (Field, 2009). A VIF greater than 10 is problematic (Myers, 1990), and an average VIF of all the predictors of considerably greater than 1 indicates that the regression may be biased (Bowerman & O'Connell, 1990). No regressions had VIFs greater than 10. Average VIFs were 1.13 for the regressions with the covariate and 1.17 for the regressions without the covariate, confirming that collinearity was not a problem for the regression models.

The tolerance statistic is the reciprocal of the VIF (Field, 2009). Tolerance values below .1 reflect a serious problem, and values below .2 are cause for concern (Menard,

1995). Tolerance values ranged from .83 to .98 for the regressions with the covariate, and were .85 for the regressions without the covariate, providing further support for collinearity not being problematic for the regression models.

Another way to investigate collinearity is to check the collinearity diagnostics, which provide eigenvalues of the scaled, uncentered cross-products matrix, condition indexes, and variance proportions (Field, 2009). For the regressions with the covariate, the largest difference between eigenvalues was 1.52. This difference was fairly small, indicating that the eigenvalues were relatively similar and that the regression models were likely unchanged by small changes in the measured variables. The condition indexes are another way of expressing the eigenvalues and symbolize the square root of the ratio of the biggest eigenvalue to the eigenvalue of interest (Field, 2009). There are no cutoffs for how much larger a condition index needs to be to reflect a problem with collinearity. In this case, condition indexes did not vary too much from 1 (2.70 was the largest value), suggesting that collinearity was not a problem.

The final way to check for collinearity is to look for predictors that have large variance proportions on the same small eigenvalue, as this indicates that their regression coefficients are dependent (Field, 2009). This was first explored for regression models that included the covariate. For the covariate (education level), 0% of the variance of the regression coefficient was associated with eigenvalue number 1, 44% was associated with eigenvalue number 2, 54% was associated with eigenvalue number 3, and 2% was associated with eigenvalue number 4. For the first dummy variable representing "GMIPness," 7% of the variance of the regression coefficient was associated with eigenvalue number 1, 20% was associated with eigenvalue number 2, 24% was

associated with eigenvalue number 3, and 49% was associated with eigenvalue number 4. For the second dummy variable representing “suboptimalness,” 10% of the variance of the regression coefficient was associated with eigenvalue number 1, 9% was associated with eigenvalue number 2, 11% was associated with eigenvalue number 3, and 71% was associated with eigenvalue number 4. In these regression models, the two dummy variables had substantial variance proportions on eigenvalue number 4, suggesting dependency between these two variables. However, conducting a Pearson correlation revealed a moderate, negative relationship between these variables (r = -.39, p < .001), suggesting that strong collinearity was not present.

Variance proportions were also checked for the regression models without the covariate. For the first dummy variable, 7% of the variance of the regression coefficient was associated with eigenvalue number 1, 42% was associated with eigenvalue number 2, and 50% was associated with eigenvalue number 3. For the second dummy variable, 10% of the variance of the regression coefficient was associated with eigenvalue number 1, 20% was associated with eigenvalue number 2, and 70% was associated with eigenvalue number 3. Similar to the regression models with the covariate, the dummy variables in the regression models without the covariate had considerable variance proportions on the same small eigenvalue, suggesting some dependency between them. However, as these variables were not found to highly correlate with each other (r = -.39, p < .001), strong collinearity was likely not problematic.

Assumption 4 states that predictors should not be correlated to external variables, which are variables not included in the regression model that influence the outcome variable (Field, 2009). This assumption was tested by exploring the relationships between

the predictors and the potential covariates of sex, race, age, and education level. As described in Chapter 3, no significant correlations were found, indicating that Assumption 4 was met.

Assumption 5 deals with homoscedasticity and states that at each level of the predictor variables, the variance of the residual terms should be constant (Field, 2009). In other words, at each level of the predictors, the residuals should have the same variance (homoscedasticity). To check this assumption for each regression model, the standardized residuals (standardized differences between the observed data and the values predicted by the model) were plotted against the standardized predicted values of the outcome variable based on the model. If the graph reflects homoscedasticity, data should appear to be evenly distributed around zero (Field, 2009). If the data funnel out, then the graph likely indicates heteroscedasticity, or that the variances are very unequal (Field, 2009). A curved graph indicates that the assumption of linearity has likely been broken (Field, 2009). Graphs funneled out for the majority of the regression models (models with the DTBM and the VF, PO, AC, and PS domains as outcome variables), indicating that there was heteroscedasticity in the data. For the EF, LM, and MD regressions, residuals were spread across three separate vertical lines. Residuals had unequal variances at various levels of the predictor, indicating heteroscedasticity was also a problem for these regression models. Although the residuals were clustered into three separate vertical lines for the LM and MD regressions, the residuals were respectively relatively symmetrically distributed around 0 at each level of the predictor, indicating that there was a linear relationship between WMT performance and these respective domain scores. However, for the EF regression, residuals were mostly negative when the predicted value of Y was

approximately -1.5, mostly positive when the predicted value of Y was slightly smaller than 0, and mostly negative when the predicted value of Y was 1. These results suggested that the relationship between WMT performance and EF scores was not linear. Overall, the assumption of homoscedasticity was violated for all of the regression models. As previously described, transforming the data did not improve non-normality or heterogeneity of variance, and thus, data remained untransformed. Failure to meet this assumption means that findings cannot be generalized beyond this sample.

Assumption 6 states that residual terms should be uncorrelated for any two observations (Field, 2009). The Durbin-Watson test is used to test this assumption. Test statistic values can range from 0 to 4 with a value of 2 signifying that the residuals are uncorrelated (Field, 2009). Durbin-Watson values ranged from 1.68 to 2.17 for the regression models, suggesting that the residuals were largely uncorrelated.

Assumption 7 states that the residuals in each regression model are random, normally distributed variables with a mean of, or very close to, zero (Field, 2009). To check the normality of residuals, histograms and normal P-P plots were examined for each model. For the VF model, the histogram distribution appeared slightly non-normal with slight deviations from normality also evidenced on the normal P-P plot. For the DTBM and the PO, EF, LM, MD, AC, and PS models, histogram distributions were largely normal though normal P-P plots evidenced slight deviations from normality.

Assumption 8 states that all of the values of the outcome variable are independent (Field, 2009). This assumption was met, as all values of each outcome variable came from a different patient.

The final assumption states that the relationship being modeled is a linear one. As noted above, a residual versus predicted scatterplot revealed a nonlinear relationship between WMT performance and EF scores. As such, the generalizability of findings from this regression is extremely limited. All other relationships modeled were linear.

Assessing the regression models: Diagnostics. Outliers and influential cases should be explored to see whether the model fits the observed data well (Field, 2009). Outliers are cases that substantially differ from the majority of observed cases, and can add bias to a regression model because they impact estimated regression coefficient values (Field, 2009). Outliers can be identified by checking standardized residuals. Standardized residuals with absolute values greater than 3.29 are likely problematic (Field, 2009). If more than 1% of the sample has standardized residuals with absolute values greater than 2.58, or, if more than 5% of the sample has standardized residuals with absolute values greater than 2, the model may be an inaccurate representation of the sample data (Field, 2009). Standardized residuals were checked for each model. No model had cases with standardized residuals with absolute values greater than 3.29. For three regression models (EF, MD, and DTBM), more than 1% of their samples had standardized residuals with absolute values greater than 2.58: the EF and DTBM regression models each had 1 case (1.23%) outside the limit, and the MD regression model had 2 cases (2.47%) outside the limit. As absolute values of standardized residuals for these regression samples were between 1 to 2% of what was expected, it was concluded that the samples largely appeared to conform to what would be expected for fairly accurate models. Only the LM regression model had slightly more than 5% of cases (6.17%) that had standardized residuals with absolute values greater than 2. However, as

this sample was within 1.17% of what was expected, it was concluded that the sample appeared to conform to what would be expected for a fairly accurate model.

Regression models were also checked for influential cases. Adjusted predicted values were compared to predicted values to ensure that cases did not have large influences over the model. For each model, all adjusted predicted values were very similar to predicted values, suggesting that the models were stable. Cook's Distances, which are measures of the influence of a case on the model (Cook & Weisberg, 1982), were examined. Cook's values greater than 1 indicate a possible problem (Cook & Weisberg, 1982). No model contained cases with Cook's values close to 1, suggesting that there were no cases that greatly influenced each model's ability to predict all cases. Average leverage values, which are measures of the effect of the observed value of the outcome over the predicted values (Field, 2009), were calculated for each regression. Following recommendations from Hoaglin and Welsch (1978) and Stevens (2002), cases were examined to check for values two to three times greater than the average leverage value for each model. For all models, all cases were within the boundary of three times the average leverage; however, five models (VF, PO, AC, PS, and DTBM) had one case each that was slightly greater than two times the average leverage. As the number of cases outside the smaller of the recommended average leverage values was very small, no cases were considered to have undue influences over the model. Mahalanobis distances, which indicate the distance of cases from the means of the predictors (Field, 2009), were examined for high values. For smaller sample sizes (N = 100) with around 3 predictors, Mahalanobis distances greater than 15 are troublesome (Field, 2009). Mahalanobis distances ranged from 1.30 to 9.32 across models, indicating that values were well within

suggested parameters. Standardized DFBeta values, which indicate differences between parameters of regression models estimated using all cases and estimated with one case excluded, were investigated. Standardized BFBetas with absolute values greater than 1 identify cases that markedly affect the model parameters (Field, 2009). All standardized DFBeta absolute values were less than 1. Finally, covariance ratios (CVRs), which are measures of whether a case impacts the variance of the regression parameter (Field, 2009), were examined. Per recommendations from Belsey, Kuh, and Welsch (1980), cases were examined for CRV values greater than 1 plus three times the leverage and for CRV values less than 1 minus three times the leverage. All CVRs fell within, or just outside, recommended ranges. Overall, examination of these values suggested that no influential cases were present in the regression models.

Exploration of the covariate. The independence of the covariate (education level) and treatment effect (WMT performance) was explored prior to conducting regression analyses. Levene's test indicated that for education level, the variances were equal for patients in each WMT group, $F(2, 78) = 2.03$, *ns*. Thus, the experimental effect was not confounded with the effect of the covariate.

Differences in VF test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in VF scores among WMT performance groups after controlling for education level. Table 5 summarizes the descriptive statistics and correlational analyses results. As can be seen, there was a significant positive correlation between education level and VF scores, $r_{COV}(79) = .23$, $p < .05$, indicating that patients with more years of education had higher VF scores. Education level accounted for 5% of the variance in VF scores, $r^2_{COV} = .05$, $p < .05$.

GMIPness was significantly negatively correlated with VF scores, $r_{X1}(79) = -.41$, $p < .001$, indicating that the mean of the VF scores in the GMIP group was smaller than the mean of the VF scores for non-GMIP group members. GMIPness accounted for 17% of the variance in VF scores, $r^2_{X1} = .17$, $p < .001$. Suboptimalness was not significantly correlated with VF scores, $r_{X2}(79) = -.01$, $p = ns$, indicating that the mean of the VF scores in the Suboptimal group was not significantly different than the mean of the VF scores for non-Suboptimal group members.

Table 5
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and VF Scores

	r				
Variables	*Y*	*COV*	X_1	X_2	r^2_{Yi}
Y	1.00	.23*	-.41***	-.01	–
COV	.23*	1.00	.04	-.16	.05*
X_1	-.41***	.04	1.00	-.39***	.17***
X_2	-.01	-.16	-.39***	1.00	.00
M	-.54	.00	.21	.36	
SD	.66	2.15	.41	.48	

Note. Y = VF scores; COV = education level; X_1 = GMIP group; X_2 = suboptimal group; *r* = zero-order (Pearson) correlation coefficient.
*$p < .05$. ***$p < .001$.

Table 6 displays partial and semipartial (part) correlation coefficients between the predictors and VF scores. As can be seen, there was a significant positive relationship between education level and VF scores after common variance with the dummy variables was removed from both the education level covariate (residualized predictor) and VF score outcome variable (residualized outcome), $pr_{COV}(77) = .25$, $p < .05$. There was also a significant positive relationship between education level and VF scores after removing variance that the education level covariate had in common with the dummy variables,

$sr_{COV}(77) = .22$, $p < .05$. Education level uniquely accounted for 5% of the variance in VF scores, $sr^2_{COV} = .05$, $p < .05$.

There was a significant negative difference, in correlational terms, in VF scores between the GMIP and optimal (reference) groups holding constant the effects of education level and Suboptimalness on both the GMIP group and VF scores, $pr_{X1}(77) = -.46$, $p < .001$. Nineteen percent of the variance in VF scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .19$, $p < .001$. In other words, 19% of the variance in VF scores was explained by the fact that the GMIP group averaged different VF scores than the optimal group. Additionally, sr^2 indicates the amount by which R^2 would be reduced if X_i were omitted from the predictor variables (Cohen & Cohen, 1983). This means that the loss of the distinction between the GMIP and optimal groups would result in a loss of 19% of the variance accounted for in VF scores, or, that R^2 would drop from .25 to .06.

There was not a significant difference, in correlational terms, in VF scores between the suboptimal and optimal groups when the effects of education level and GMIPness on both the suboptimal group and VF scores were held constant, $pr_{X2}(77) = -.17$, $p = ns$. A significant portion of the variance in VF scores was not accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .02$, $p = ns$.

Table 6
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and VF Scores

Variables	pr_i	sr_i	sr^2_i
COV	.25*	.22*	.05*
X_1	-.46***	-.44***	.19***
X_2	-.17	-.14	.02

Note. pr_i = partial correlation coefficient; sr_i = semipartial (part) correlation coefficient. *$p < .05$. ***$p < .001$.

The overall regression model was statistically significant, $F(3, 77) = 8.58$, $p < .001$. As shown in Table 7, education level alone accounted for 5% of the variance in VF scores, $F(1, 79) = 4.38$, $p < .05$, and the WMT performance dummy variables explained an additional 20% of the variance, $F_{change}(2, 77) = 10.16$, $p < .001$. Thus, a total of 25% of the variance in VF scores was explained by education level and WMT performance. More specifically, holding WMT performance constant, each additional year of education was associated with a .07 increase in VF scores. Controlling for the effects of education level and suboptimalness, there was a significant difference between the mean VF scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, VF scores decreased by .77 points. In other words, patients in the GMIP group averaged VF scores .77 points lower than patients in the optimal group. Mean VF scores between patients in the suboptimal and optimal groups were not found to significantly differ, indicating that VF scores were relatively similar whether patients performed optimally or suboptimally on the WMT. Mean VF scores for each WMT performance group are presented in Table 8.

Table 7
Summary of Hierarchical Regression Analysis for Differences in VF Scores Among WMT Performance Groups Accounting for Education Level ($N = 81$)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	-.54	.07				
COV	.07	.03	.23*			
Constant				-.30	.10	
COV				.07	.03	.22*
X_1				-.77	.17	-.48***
X_2				-.22	.15	-.16

Note. *B* = unstandardized regression coefficient; *SE B* = standard error of the regression coefficient; *Beta* = standardized regression coefficient; Model 1: $R^2 = .05$ ($p < .05$), adjusted $R^2 = .04$; Model 2: $R^2 = .25$ ($p < .001$), adjusted $R^2 = .22$, $\Delta R^2 = .20$ ($p < .001$).
*$p < .05$. ***$p < .001$.

Table 8
WMT Performance Groups' Mean VF Scores

WMT Performance Group	*Mean VF Scores*
Optimal	-.30
GMIP	-1.07
Suboptimal	-.52

Note. Scores are presented as *z*-scores.

Differences in PO test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in PO scores among WMT performance groups after controlling for education level. Table 9 summarizes the descriptive statistics and correlational analyses results. As can be seen, there was a significant positive correlation between education level and PO scores, $r_{COV}(79) = .31$, $p < .01$, indicating that patients with more years of education had higher PO scores. Education level accounted for 10% of the variance in PO scores, $r^2_{COV} = .10$, $p < .01$. GMIPness was significantly negatively correlated with PO scores, $r_{X1}(79) = -.40$, $p < .001$, indicating that the mean of the PO scores in the GMIP group was smaller than the mean of the PO scores for non-GMIP group members. GMIPness accounted for 16% of the variance in PO scores, $r^2_{X1} = .16$, $p < .001$. Suboptimalness was not significantly correlated with PO scores, $r_{X2}(79) = -.06$, $p = ns$, indicating that the mean of the PO scores in the suboptimal group was not significantly different than the mean of the PO scores for non-suboptimal group members.

Table 9
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and PO Scores

	r				
Variables	*Y*	*COV*	X_1	X_2	r^2_{Yi}
Y	1.00	.31**	-.40***	.06	–
COV	.31**	1.00	.04	-.16	.10**
X_1	-.40***	.04	1.00	-.39***	.16***
X_2	-.06	-.16	-.39***	1.00	.00
M	-.27	.00	.21	.36	
SD	.76	2.15	.41	.48	

Note. Y = PO scores. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 10 displays partial and semipartial correlation coefficients between the predictors and PO scores. As can be seen, there was a significant positive relationship between education level and PO scores after common variance with the dummy variables was removed from both the education level covariate and the PO outcome variable, $pr_{COV}(77) = .34$, $p < .01$. There was also a significant positive relationship between education level and PO scores after removing variance that the education level covariate had in common with the dummy variables, $sr_{COV}(77) = .31$, $p < .01$. Education level uniquely accounted for 10% of the variance in PO scores, $sr^2_{COV} = .10$, $p < .01$.

There was a significant negative difference, in correlational terms, in PO scores between the GMIP and optimal groups holding constant the effects of education level and suboptimalness on both the GMIP group and PO scores, $pr_{X1}(77) = -.42$, $p < .001$. Fifteen percent of the variance in PO scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .15$, $p < .001$. The loss of the distinction between the GMIP and optimal groups would result in a loss of 15% of the variance accounted for in PO scores, or, that R^2 would drop from .26 to .11.

There was not a significant difference, in correlational terms, in PO scores between the suboptimal and optimal groups when the effects of education level and GMIPness on both the suboptimal group and PO scores were held constant, $pr_{X2}(77) = -.06$, $p = ns$. A significant portion of the variance in PO scores was not accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .00$, $p = ns$.

Table 10
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and PO Scores

Variables	pr_i	sr_i	sr^2_i
COV	.34**	.31**	.10**
X_1	-.42***	-.39***	.15***
X_2	-.06	-.05	.00

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

The overall regression model was statistically significant, $F(3, 77) = 9.07$, $p < .001$. As shown in Table 11, education level alone accounted for 9% of the variance in PO scores, $F(1, 79) = 8.21$, $p < .01$, and the WMT performance dummy variables explained an additional 17% of the variance, $F_{\text{change}}(2, 77) = 8.70$, $p < .001$. Thus, a total of 26% of the variance in PO scores was explained by education level and WMT performance. More specifically, holding WMT performance constant, each additional year of education was associated with a .11 increase in PO scores. Controlling for the effects of education level and suboptimalness, there was a significant difference between the mean PO scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, PO scores decreased by .80 points. Stated differently, patients in the GMIP group averaged PO scores .80 points lower than patients in the optimal group. Mean PO scores between patients in the suboptimal and optimal groups were not found to significantly differ, indicating that PO

scores were relatively similar whether patients performed optimally or suboptimally on the WMT. Mean PO scores for each WMT performance group are presented in Table 12.

Table 11
Summary of Hierarchical Regression Analysis for Differences in PO Scores Among WMT Performance Groups Accounting for Education Level (N = 81)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	-.27	.08				
COV	.11	.04	.31**			
Constant				-.07	.11	
COV				.11	.04	.31**
X_1				-.80	.20	-.43***
X_2				-.08	.17	-.05

Note. Model 1: R^2 = .09 (p < .01), adjusted R^2 = .08; Model 2: R^2 = .26 (p < .001), adjusted R^2 = .23, ΔR^2 = .17 (p < .001). **p < .01. ***p < .001.

Table 12
WMT Performance Groups' Mean PO Scores

WMT Performance Group	*Mean PO Scores*
Optimal	-.07
GMIP	-.87
Suboptimal	-.15

Note. Scores are presented as *z*-scores.

Differences in EF test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in EF scores among WMT performance groups. Table 13 summarizes the descriptive statistics and correlational analyses results. As can be seen, GMIPness was significantly negatively correlated with EF scores, $r_{X1}(79) = -.35$, $p < .01$, indicating that the mean of the EF scores in the GMIP group was smaller than the mean of the EF scores for non-GMIP group members. GMIPness accounted for 12% of the variance in EF scores, $r^2_{X1} = .12$, $p < .01$. Suboptimalness was not significantly correlated with EF scores, $r_{X2}(79) = -.11$, p =

ns, indicating that the mean of the EF scores in the suboptimal group was not significantly different than the mean of the EF scores for non-suboptimal group members.

Table 13
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and EF Scores

	r			
Variables	*Y*	X_1	X_2	r^2_{Yi}
Y	1.00	-.35**	-.11	–
X_1	-.35**	1.00	-.39***	.12**
X_2	-.11	-.39***	1.00	.01
M	-.75	.21	.36	
SD	.88	.41	.48	

Note. Y = EF scores. ** $p < .01$. *** $p < .001$.

Table 14 displays partial and semipartial correlation coefficients between the predictors and EF scores. As is evident, there was a significant negative difference, in correlational terms, in EF scores between the GMIP and optimal groups holding constant the effect of suboptimalness on both the GMIP group and EF scores, $pr_{X1}(78) = -.42$, $p < .001$. Eighteen percent of the variance in EF scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .18$, $p < .001$. The loss of the distinction between the GMIP and optimal groups would result in a loss of 18% of the variance accounted for in EF scores, or, that R^2 would drop from .19 to .01.

There was also a significant negative difference, in correlational terms, in EF scores between the suboptimal and optimal groups when the effect of GMIPness on both the suboptimal group and EF scores was held constant, $pr_{X2}(78) = -.28$, $p < .05$. Seven percent of the variance in EF scores was accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .07$, $p < .05$. The loss of the

distinction between the suboptimal and optimal groups would result in a loss of 7% of the variance accounted for in EF scores, or, that R^2 would drop from .19 to .12.

Table 14
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and EF Scores

Variables	pr_i	sr_i	sr^2_i
X_1	-.42***	-.42***	.18***
X_2	-.28*	-.26*	.07*

Note. *$p < .05$. ***$p < .001$.

The regression model was statistically significant, $F(2, 78) = 9.02$, $p < .001$, and accounted for 19% of the variance in EF scores ($R^2 = .19$, Adjusted $R^2 = .17$). As shown in Table 15, controlling for the effect of suboptimalness, there was a significant difference between the mean EF scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, EF scores decreased by .98 points. In other words, patients in the GMIP group averaged EF scores .98 points lower than patients in the optimal group. Holding constant the effect of GMIPness, there was also a significant difference between the mean EF scores of suboptimal and optimal group members. As a patient changed from performing optimally to suboptimally on the WMT, EF scores decreased by .51 points, indicating that patients in the suboptimal group averaged EF scores .51 points lower than patients in the optimal group. Mean EF scores for each WMT performance group are presented in Table 16.

Table 15
Summary of Regression Analysis for Differences in EF Scores Among WMT Performance Groups (*N* = 81)

Variables	*B*	*SE B*	*Beta*
Constant	-.36	.14	
X_1	-.98	.24	-.46***
X_2	-.51	.20	-.28*

Note. R^2 = .19 (p < .001), adjusted R^2 = .17. *p < .05. ***p < .001.

Table 16
WMT Performance Groups' Mean EF Scores

WMT Performance Group	*Mean EF Scores*
Optimal	-.36
GMIP	-1.34
Suboptimal	-.87

Note. Scores are presented as *z*-scores.

Differences in LM test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in LM scores among WMT performance groups. Table 17 summarizes the descriptive statistics and correlational analyses results. As can be seen, GMIPness was significantly negatively correlated with LM scores, $r_{X1}(79) = -.54$, $p < .001$, indicating that the mean of the LM scores in the GMIP group was smaller than the mean of the LM scores for non-GMIP group members. GMIPness accounted for 29% of the variance in LM scores, $r^2_{X1} = .29$, $p < .001$. Suboptimalness was not significantly correlated with LM scores, $r_{X2}(79) = -.09$, $p = ns$, indicating that the mean of the LM scores in the suboptimal group was not significantly different than the mean of the LM scores for non-suboptimal group members.

Table 17
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and LM Scores

	r			
Variables	*Y*	X_1	X_2	r^2_{Yi}
Y	1.00	-.54***	-.09	–
X_1	-.54***	1.00	-.39***	.29***
X_2	-.09	-.39***	1.00	.01
M	-.73	.21	.36	
SD	.88	.41	.48	

Note. Y = LM scores. *** $p < .001$.

Table 18 displays partial and semipartial correlation coefficients between the predictors and LM scores. As can be seen, there was a significant negative difference, in correlational terms, in LM scores between the GMIP and optimal groups holding constant the effect of suboptimalness on both the GMIP group and LM scores, $pr_{X1}(78) = -.63$, $p < .001$. Thirty eight percent of the variance in LM scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .38$, $p < .001$. The loss of the distinction between the GMIP and optimal groups would result in a loss of 38% of the variance accounted for in LM scores, or, that R^2 would drop from .40 to .02.

There was also a significant negative difference, in correlational terms, in LM scores between the suboptimal and optimal groups when the effect of GMIPness on both the suboptimal group and LM scores was held constant, $pr_{X2}(78) = -.38$, $p < .01$. Ten percent of the variance in LM scores was accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .10$, $p < .01$. The loss of the distinction between the suboptimal and optimal groups would result in a loss of 10% of the variance accounted for in LM scores, or, that R^2 would drop from .40 to .30.

Table 18
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and LM Scores

Variables	pr_i	sr_i	sr^2_i
X_1	-.63***	-.62***	.38***
X_2	-.38**	-.32**	.10**

Note. **$p < .01$. ***$p < .001$.

The regression model was statistically significant, $F(2, 78) = 25.68$, $p < .001$, and accounted for 40% of the variance in LM scores ($R^2 = .40$, Adjusted $R^2 = .38$). As displayed in Table 19, controlling for the effect of suboptimalness, there was a significant difference between the mean LM scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, LM scores decreased by 1.45 points. In other words, patients in the GMIP group averaged LM scores 1.45 points lower than patients in the optimal group. Holding constant the effect of GMIPness, there was also a significant difference between the mean LM scores of suboptimal and optimal group members. As a patient changed from performing optimally to suboptimally on the WMT, LM scores decreased by .63 points, indicating that patients in the suboptimal group averaged LM scores .63 points lower than patients in the optimal group. Mean LM scores for each WMT performance group are presented in Table 20.

Table 19
Summary of Regression Analysis for Differences in LM Scores Among WMT Performance Groups ($N = 81$)

Variables	*B*	*SE B*	*Beta*
Constant	-.20	.12	
X_1	-1.45	.20	-.68***
X_2	-.63	.17	-.35**

Note. $R^2 = .40$ ($p < .001$), adjusted $R^2 = .38$. **$p < .01$. ***$p < .001$.

Table 20
WMT Performance Groups' Mean LM Scores

WMT Performance Group	*Mean LM Scores*
Optimal	-.20
GMIP	-1.65
Suboptimal	-.83

Note. Scores are presented as *z*-scores.

Differences in AC test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in AC scores among WMT performance groups after controlling for education level. Table 21 summarizes the descriptive statistics and correlational analyses results. As can be seen, there was a significant positive correlation between education level and AC scores, $r_{COV}(79) = .25$, $p < .05$, indicating that patients with more years of education had higher AC scores. Education level accounted for 6% of the variance in AC scores, $r^2_{COV} = .06$, $p < .05$. GMIPness was significantly negatively correlated with AC scores, $r_{X1}(79) = -.37$, $p < .001$, indicating that the mean of the AC scores in the GMIP group was smaller than the mean of the AC scores for non-GMIP group members. GMIPness accounted for 14% of the variance in AC scores, $r^2_{X1} = .14$, $p < .001$. Suboptimalness was not significantly correlated with AC scores, $r_{X2}(79) = -.12$, $p = ns$, indicating that the mean of the AC scores in the suboptimal group was not significantly different than the mean of the AC scores for non-suboptimal group members.

Table 21
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and AC Scores

	r				
Variables	*Y*	*COV*	X_1	X_2	r^2_{Yi}
Y	1.00	.25*	-.37***	-.12	–
COV	.25*	1.00	.04	-.16	.06*
X_1	-.37***	.04	1.00	-.39***	.14***
X_2	-.12	-.16	-.39***	1.00	.01
M	-.67	.00	.21	.36	
SD	.86	2.15	.41	.48	

Note. Y = AC scores. $^{*}p < .05$. $^{***}p < .001$.

Table 22 displays partial and semipartial correlation coefficients between the predictors and AC scores. As can be seen, there was a significant positive relationship between education level and AC scores after common variance with the dummy variables was removed from both the education level covariate and the AC outcome variable, $pr_{COV}(77) = .25$, $p < .05$. There was also a significant positive relationship between education level and AC scores after removing variance that the education level covariate had in common with the dummy variables, $sr_{COV}(77) = .22$, $p < .05$. Education level uniquely accounted for 5% of the variance in AC scores, $sr^2_{COV} = .05$, $p < .05$.

There was a significant negative difference, in correlational terms, in AC scores between the GMIP and the optimal group holding constant the effects of education level and suboptimalness on both the GMIP group and AC scores, $pr_{X1}(77) = -.46$, $p < .001$. Nineteen percent of the variance in AC scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .19$, $p < .001$. The loss of the distinction between the GMIP group and the optimal group would result in a loss of 19% of the variance accounted for in AC scores, or, that R^2 would drop from .26 to .07.

There was also a significant negative difference, in correlational terms, in AC scores between the suboptimal and optimal groups when the effects of education level and GMIPness on both the suboptimal group and AC scores were held constant, $pr_{X2}(77) = -.28$, $p < .05$. Six percent of the variance in AC scores was accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .06$, $p < .05$. The loss of the distinction between the suboptimal and optimal groups would result in a loss of 5% of the variance accounted for in AC scores, or, that R^2 would drop from .26 to .21.

Table 22
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and AC Scores

Variables	pr_i	sr_i	sr^2_i
COV	.25*	.22*	.05*
X_1	-.46***	-.44***	.19***
X_2	-.28*	-.25*	.06*

Note. *$p < .05$. ***$p < .001$.

The overall regression model was statistically significant, $F(3, 77) = 9.15$, $p < .001$. As shown in Table 23, education level alone accounted for 6% of the variance in AC scores, $F(1, 79) = 5.20$, $p < .05$, and the WMT performance dummy variables explained an additional 20% of the variance, $F_{change}(2, 77) = 10.50$, $p < .001$. Thus, a total of 26% of the variance in AC scores was explained by education level and WMT performance. More specifically, holding WMT performance constant, each additional year of education was associated with a .09 increase in AC scores. Controlling for the effects of education level and suboptimalness, there was a significant difference between the mean AC scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, AC scores decreased by 1 point. Stated differently, patients in the GMIP group averaged AC scores 1 point

lower than patients in the optimal group. Holding constant the effect of GMIPness, there was also a significant difference between the mean AC scores of suboptimal and optimal group members. As a patient changed from performing optimally to suboptimally on the WMT, AC scores decreased by .48 points, indicating that patients in the suboptimal group averaged AC scores .48 points lower than patients in the optimal group. Mean AC scores for each WMT performance group are presented in Table 24.

Table 23
Summary of Hierarchical Regression Analysis for Differences in AC Scores Among WMT Performance Groups Accounting for Education Level (N = 81)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	-.67	.09				
COV	.10	.04	.25*			
Constant				-.29	.13	
COV				.09	.04	.22*
X_1				-1.00	.22	-.48***
X_2				-.48	.19	-.27*

Note. Model 1: R^2 = .06 (p < .05), adjusted R^2 = .05; Model 2: R^2 = .26 (p < .001), adjusted R^2 = .23, ΔR^2 = .20 (p < .001). *p < .05. ***p < .001.

Table 24
WMT Performance Groups' Mean AC Scores

WMT Performance Group	*Mean AC Scores*
Optimal	-.29
GMIP	-1.29
Suboptimal	-.77

Note. Scores are presented as *z*-scores.

Differences in PS test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in PS scores among WMT performance groups after controlling for education level. Table 25 summarizes the descriptive statistics and correlational analyses results. As can be seen, there was a significant positive correlation between education level and PS scores, $r_{COV}(79) = .30$, p

< .01, indicating that patients with more years of education had higher PS scores. Education level accounted for 9% of the variance in PS scores, $r^2_{COV} = .09$, $p < .01$. GMIPness was significantly negatively correlated with PS scores, $r_{X1}(79) = -.39$, $p < .001$, indicating that the mean of the PS scores in the GMIP group was smaller than the mean of the PS scores for non-GMIP group members. GMIPness accounted for 15% of the variance in PS scores, $r^2_{X1} = .15$, $p < .001$. Suboptimalness was not significantly correlated with PS scores, $r_{X2}(79) = -.01$, $p = ns$, indicating that the mean of the PS scores in the suboptimal group was not significantly different than the mean of the PS scores for non-suboptimal group members.

Table 25
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and PS Scores

	r				
Variables	*Y*	*COV*	X_1	X_2	r^2_{Yi}
Y	1.00	.30**	-.39***	-.01	–
COV	.30**	1.00	.04	-.16	.09**
X_1	-.39***	.04	1.00	-.39***	.15***
X_2	-.01	-.16	-.39***	1.00	.00
M	-.77	.00	.21	.36	
SD	.86	2.15	.41	.48	

Note. Y = PS scores. ** $p < .01$. *** $p < .001$.

Table 26 displays partial and semipartial correlation coefficients between the predictors and PS scores. As can be seen, there was a significant positive relationship between education level and PS scores after common variance with the dummy variables was removed from both the education level covariate and the PS outcome variable, $pr_{COV}(77) = .32$, $p < .01$. There was also a significant positive relationship between education level and PS scores after removing variance that the education level covariate

had in common with the dummy variables, sr_{COV} (77) = .29, $p < .01$. Education level uniquely accounted for 8% of the variance in PS scores, $sr^2_{COV} = .08$, $p < .01$.

There was a significant negative difference, in correlational terms, in PS scores between the GMIP and the optimal group holding constant the effects of education level and suboptimalness on both the GMIP group and PS scores, $pr_{X1}(77) = -.44$, $p < .001$. Eighteen percent of the variance in PS scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .18$, $p < .001$. The loss of the distinction between the GMIP group and the optimal group would result in a loss of 18% of the variance accounted for in PS scores, or, that R^2 would drop from .27 to .09.

There was not a significant difference, in correlational terms, in PS scores between the suboptimal and optimal groups when the effects of education level and GMIPness on both the suboptimal group and PS scores were held constant, $pr_{X2}(77) = -.15$, $p = ns$. A significant portion of the variance in PS scores was not accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .02$, $p = ns$.

Table 26
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and PS Scores

Variables	pr_i	sr_i	sr^2_i
COV	.32**	.29**	.08**
X_1	-.44***	-.42***	.18***
X_2	-.15	-.13	.02

Note. **$p < .01$. ***$p < .001$.

The overall regression model was statistically significant, $F(3, 77) = 9.55$, $p < .001$. As shown in Table 27, education level alone accounted for 9% of the variance in PS scores, $F(1, 79) = 7.94$, $p < .01$, and the WMT performance dummy variables explained an additional 18% of the variance, $F_{\text{change}}(2, 77) = 9.50$, $p < .001$. Thus, a total

of 27% of the variance in PS scores was explained by education level and WMT performance. More specifically, holding WMT performance constant, each additional year of education was associated with a .12 increase in PS scores. Controlling for the effects of education level and suboptimalness, there was a significant difference between the mean PS scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, PS scores decreased by .96 points. In other words, patients in the GMIP group averaged PS scores .96 points lower than patients in the optimal group. Mean PS scores between patients in the suboptimal and optimal groups were not found to significantly differ, indicating that PS scores were relatively similar whether patients performed optimally or suboptimally on the WMT. Mean PS scores for each WMT performance group are presented in Table 28.

Table 27
Summary of Hierarchical Regression Analysis for Differences in PS Scores Among WMT Performance Groups Accounting for Education Level (N = 81)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	-.77	.09				
COV	.12	.04	.30**			
Constant				-.48	.13	
COV				.12	.04	.30**
X_1				-.96	.22	-.46***
X_2				-.25	.19	-.14

Note. Model 1: R^2 = .09 (p < .01), adjusted R^2 = .08; Model 2: R^2 = .27 (p < .001), adjusted R^2 = .24, ΔR^2 = .18 (p < .001). **p < .01. ***p < .001.

Table 28
WMT Performance Groups' Mean PS Scores

WMT Performance Group	*Mean PS Scores*
Optimal	-.48
GMIP	-1.44
Suboptimal	-.73

Note. Scores are presented as *z*-scores.

Differences in MD test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in MD scores among WMT performance groups. Table 29 summarizes the descriptive statistics and correlational analyses results. As can be seen, GMIPness was not significantly correlated with MD scores, $r_{X1}(79) = -.14$, $p = ns$, indicating that the mean of the MD scores in the GMIP group was not significantly different than the mean of the MD scores for non-GMIP group members. Suboptimalness was also not significantly correlated with MD scores, $r_{X2}(79) = .04$, $p = ns$, indicating that the mean of the MD scores in the suboptimal group was not significantly different than the mean of the MD scores for non-suboptimal group members.

Table 29
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables and MD Scores

	r			
Variables	*Y*	X_1	X_2	r^2_{Yi}
Y	1.00	-.14	.04	–
X_1	-.14	1.00	-.39***	.02
X_2	.04	-.39***	1.00	.00
M	-.91	.21	.36	
SD	.81	.41	.48	

Note. Y =MD scores. *** $p < .001$.

Table 30 displays partial and semipartial correlation coefficients between the predictors and MD scores. As is evident, there was not a significant difference, in correlational terms, in MD scores between the GMIP and optimal groups when the effect suboptimalness on both the GMIP group and MD scores was held constant, $pr_{X1}(78) = -.14$, $p = ns$. A significant portion of the variance in MD scores was not accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .02$, $p = ns$.

There also was not a significant difference, in correlational terms, in MD scores between the suboptimal and optimal groups when the effect GMIPness on both the suboptimal group and MD scores was held constant, $pr_{X2}(78) = -.02$, $p = ns$. A significant portion of the variance in MD scores was not accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .00$, $p = ns$.

Table 30
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables and MD Scores

Variables	pr_i	sr_i	sr^2_i
X_1	-.14	-.14	.02
X_2	-.02	-.02	.00

Note. All correlation coefficients were *ns*.

The regression model was not statistically significant, $F(2, 78) = .80$, $p = ns$, and therefore did not account for any of the variance in MD scores. As can be seen in Table 31, controlling for the effect of suboptimalness, there was not a significant difference between the mean MD scores of GMIP and optimal group members. Similarly, controlling for the effect of GMIPness, there was not a significant difference between the mean MD scores of suboptimal and optimal group members. Mean MD scores for each WMT performance group are presented in Table 32.

Table 31
Summary of Regression Analysis for Differences in MD Scores Among WMT Performance Groups ($N = 81$)

Variables	*B*	*SE B*	*Beta*
Constant	-.84	.14	
X_1	-.29	.24	-.15
X_2	-.03	.20	-.02

Note. $R^2 = .02$ ($p = ns$), adjusted $R^2 = -.01$.

Table 32
WMT Performance Groups' Mean MD Scores

WMT Performance Group	*Mean MD Scores*
Optimal	-.84
GMIP	-1.13
Suboptimal	-.87

Note. Scores are presented as *z*-scores.

Differences in DTBM test scores among WMT performance groups. Correlation and multiple regression analyses were conducted to examine differences in DTBM scores among WMT performance groups after controlling for education level. Table 33 summarizes the descriptive statistics and correlational analyses results. As can be seen, there was a significant positive correlation between education level and DTBM scores, $r_{COV}(79) = .25$, $p < .05$, indicating that patients with more years of education had higher DTBM scores. Education level accounted for 6% of the variance in DTBM scores, $r^2_{COV} = .06$, $p < .05$. GMIPness was significantly negatively correlated with DTBM scores, $r_{X1}(79) = -.48$, $p < .001$, indicating that the mean of the DTBM scores in the GMIP group was smaller than the mean of the DTBM scores for non-GMIP group members. GMIPness accounted for 23% of the variance in DTBM scores, $r^2_{X1} = .23$, $p < .001$. Suboptimalness was not significantly correlated with DTBM scores, $r_{X2}(79) = -.05$, $p = ns$, indicating that the mean of the DTBM scores in the Suboptimal group was not significantly different than the mean of the DTBM scores for non-suboptimal group members.

Table 33
Correlation Coefficients, Means, and Standard Deviations for WMT Performance Dummy Variables, Education Level Covariate, and DTBM Scores

	r				
Variables	*Y*	*COV*	X_1	X_2	r^2_{Yi}
Y	1.00	.25*	-.48***	-.05	–
COV	.25*	1.00	.04	-.16	.06*
X_1	-.48***	.04	1.00	-.39***	.23***
X_2	-.05	-.16	-.39***	1.00	.00
M	-.66	.00	.21	.36	
SD	.63	2.15	.41	.48	

Note. Y = DTBM scores. * $p < .05$. *** $p < .001$.

Table 34 displays partial and semipartial correlation coefficients between the predictors and DTBM scores. As can be seen, there was a significant positive relationship between education level and DTBM scores after common variance with the dummy variables was removed from both the education level covariate and the DTBM outcome variable, $pr_{COV}(77) = .28$, $p < .05$. There was also a significant positive relationship between education level and DTBM scores after removing variance that the education level covariate had in common with the dummy variables, $sr_{COV}(77) = .23$, $p < .05$. Education level uniquely accounted for 5% of the variance in DTBM scores, $sr^2_{COV} = .05$, $p < .05$.

There was a significant negative difference, in correlational terms, in DTBM scores between the GMIP and the optimal group holding constant the effects of education level and suboptimalness on both the GMIP group and DTBM scores, $pr_{X1}(77) = -.55$, $p < .001$. Twenty-nine percent of the variance in DTBM scores was accounted for by defining GMIP as a category distinct from the optimal category, $sr^2_{X1} = .29$, $p < .001$. The loss of the distinction between the GMIP group and the optimal group would result in a

loss of 29% of the variance accounted for in DTBM scores, or, that R^2 would drop from .35 to .06.

There was also a significant negative difference, in correlational terms, in DTBM scores between the suboptimal and optimal groups when the effects of education level and GMIPness on both the suboptimal group and DTBM scores were held constant, $pr_{X2}(77) = -.26$, $p < .05$. Four percent of the variance in DTBM scores was accounted for by defining suboptimal as a category distinct from the optimal category, $sr^2_{X2} = .04$, $p < .05$. The loss of the distinction between the suboptimal and optimal groups would result in a loss of 4% of the variance accounted for in DTBM scores, or, that R^2 would drop from .35 to .31.

Table 34
Partial and Semipartial Correlation Coefficients for WMT Performance Dummy Variables, Education Level Covariate, and DTBM Scores

Variables	pr_i	sr_i	sr^2_i
COV	.28*	.23*	.05*
X_1	-.55***	-.54***	.29***
X_2	-.26*	-.21*	.04 *

Note. *$p < .05$. ***$p < .001$.

The overall regression model was statistically significant, $F(3, 77) = 13.81$, $p < .001$. As shown in Table 35, education level alone accounted for 6% of the variance in DTBM scores, $F(1, 79) = 5.31$, $p < .05$, and the WMT performance dummy variables explained an additional 29% of the variance, $F_{\text{change}}(2, 77) = 16.99$, $p < .001$. Thus, a total of 35% of the variance in DTBM scores was explained by education level and WMT performance. More specifically, holding WMT performance constant, each additional year of education was associated with a .07 increase in DTBM scores. Controlling for the effects of education level and suboptimalness, there was a significant difference between

the mean DTBM scores of GMIP and optimal group members. As a patient changed from performing optimally on the WMT to having a GMIP performance, DTBM scores decreased by .89 points. Stated differently, patients in the GMIP group averaged DTBM scores .89 points lower than patients in the optimal group. Holding constant the effects of education level and GMIPness, there was also a significant difference between the mean DTBM scores of suboptimal and optimal group members. As a patient changed from performing optimally to suboptimally on the WMT, DTBM scores decreased by .31 points, indicating that patients in the suboptimal group averaged DTBM scores .31 points lower than patients in the optimal group. Mean DTBM scores for each WMT performance group are presented in Table 36.

Table 35
Summary of Hierarchical Regression Analysis for Differences in DTBM Scores Among WMT Performance Groups Accounting for Education Level (*N* = 81)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	-.66	.07				
COV	.07	.03	.25*			
Constant				-.37	.09	
COV				.07	.03	.24*
X_1				-.89	.15	-.58***
X_2				-.31	.13	-.23*

Note. Model 1: R^2 = .06 (p < .05), adjusted R^2 = .05; Model 2: R^2 = .35 (p < .001), adjusted R^2 = .33, ΔR^2 = .29 (p < .001). *p < .05. ***p < .001.

Table 36
WMT Performance Groups' Mean DTBM Scores

WMT Performance Group	*Mean DTBM Scores*
Optimal	-.37
GMIP	-1.26
Suboptimal	-.68

Note. Scores are presented as *z*-scores.

Research question 4. What is the relationship between GMIP scores and scores on neuropsychological memory tests?

Assumptions of regression. For the linear and multiple regressions that were conducted for research question 4, Assumptions 1 (variable types) and 2 (non-zero variance) were met: quantitative predictors were measured at the interval level; categorical variables had two categories; outcome variables were quantitative, measured at the interval level, and unbounded; and all predictors had variances greater than zero. The assumption of multicollinearity (Assumption 3) was not applicable to either regression as the first regression had one continuous predictor and the second regression had one dichotomous categorical predictor. Predictors were not significantly correlated with external variables, which were previously identified as the potential covariates of sex, race, education, or age, indicating that Assumption 4 was met.

The assumption of homoscedasticity (Assumption 5) was not met for either regression. In the linear regression, the data were observed to funnel out. In the multiple regression, the variance of residual terms was different for each level of the predictor. As previously described, transforming the data did not improve heteroscedasticity, and thus, data remained untransformed. Failure to meet the assumption of homoscedasticity means that findings cannot be generalized beyond this sample.

The assumption of independent errors (Assumption 6) was met for both regressions. Durbin-Watson values were 1.57 for the linear regression and 1.67 for the multiple regression. To check the normality of residuals for Assumption 7, histograms and normal P-P plots were examined. For the linear regression, the histogram distribution appeared largely normal though the normal P-P plot demonstrated slight deviations from

normality. For the multiple regression, the histogram distribution was slightly negatively skewed; slight deviations from normality were also evident on the normal P-P plot.

The assumption of independence (Assumption 8) was met for both regressions, as all values of each outcome variable came from a different patient. Finally, the assumption of linearity was met for both regressions.

Assessing the regression models: Diagnostics. Outliers and influential cases were explored to see whether the models fit the observed data well. To detect outliers, standardized residuals were checked for each regression. No regression model had cases with standardized residuals with absolute values greater than 3.29. The linear regression model had 1 case (1.23% of the sample) that had a standardized residual with an absolute value greater than 2.58 and 2 cases (2.47% of the sample) that had standardized residuals with absolute values greater than 2. The multiple regression model had 2 cases (2.47% of the sample) that had standardized residuals with absolute values greater than 2.58 and 3 cases (3.70% of the sample) that had standardized residuals with absolute values greater than 2. As absolute values of standardized residuals for the first regression model were within 1% of what was expected, it was concluded that the sample largely appeared to conform to what would be expected for a fairly accurate model. The second regression model had slightly more than 5% of cases (6.17%) that had standardized residuals with absolute values greater than 2. However, as this sample was within 1.17% of what was expected, it was concluded that the sample appeared to conform to what would be expected for a fairly accurate model.

Regression models were also checked for influential cases. Adjusted predicted values were compared to predicted values to ensure that cases did not have large

influences over the model. For both models, adjusted predicted values were very similar to predicted values, suggesting that the models were stable. Cook's Distances were examined next. Neither regression model contained cases with Cook's values close to 1, suggesting that no cases greatly influenced either model's ability to predict all cases. Average leverage values were calculated for each regression. All cases in both regression models were within the boundary of two times their respective average leverage values, suggesting that no cases had undue influence over the models. Mahalanobis distances were examined for high values. Mahalanobis distances ranged from .01 to 3.80 across models, indicating that values were well within suggested parameters previously described. Standardized DFBeta values were investigated. All standardized DFBeta absolute values were less than 1, signifying that no cases substantially influenced the models' parameters. Finally, CVRs were examined. For the linear regression model, one case fell well below the bottom limit; however, the ratio was still close to 1 therefore indicating that the case had very little influence on the variances of the model parameters. For the multiple regression model, 2 cases fell slightly below the bottom limit; however, the ratios were close to 1, signifying that the cases had little influence on the variances of the model parameters. Overall, examination of all of these values suggested that no influential cases were present in the regression models.

Relationship between GMIP scores and scores on neuropsychological memory tests. GMIP scores were used in a simple linear regression analysis to predict LM scores. The correlations of the variables are shown in Table 37. As can be seen, there was a significant negative correlation between GMIP scores and LM scores, $r_{Xi}(79) = -.74$, $p < .001$, indicating that patients with higher GMIP scores had lower LM scores. As a

reminder, the higher the GMIP score, the *more likely* the patient performed poorly on the WMT due to significant cognitive impairment. GMIP score accounted for 55% of the variance in LM scores, $r^2_{Xi} = .55$, $p < .001$.

Table 37
Correlation Coefficients, Means, and Standard Deviations for GMIP Scores and LM Scores

	r		
Variables	*Y*	X_i	r^2_{Yi}
Y	1.00	-.74***	–
X_i	-.74***	1.00	.55***
M	-.73	28.96	
SD	.88	13.36	

Note. Y = LM scores; X_i = GMIP scores.
***$p < .001$.

The regression model was statistically significant, $F(1, 79) = 98.24$, $p < .001$, and accounted for 55% of the variance in LM scores ($R^2 = .55$, Adjusted $R^2 = .55$). As can be seen in Table 38, when GMIP scores were zero, the model predicted average LM scores of .69. A one-point increase in GMIP scores predicted that LM scores would decrease by .05 points. The standardized beta value indicated that as GMIP scores increased by one standard deviation, LM scores decreased by .74 standard deviations.

Table 38
Summary of Regression Analysis of GMIP Scores Predicting LM Scores ($N = 81$)

Variables	*B*	*SE B*	*Beta*
Constant	.69	.16	
X_i	-.05	.01	-.74***

Note. $R^2 = .55$ ($p < .001$), adjusted $R^2 = .55$. ***$p < .001$.

Correlation and multiple regression analyses were next conducted to examine differences in LM scores among GMIP performance groups. Patients with GMIP scores < 30 were categorized into the non-GMIP group (reference) and patients with GMIP scores

≥ 30 who failed at least one PV subtest were categorized into the GMIP group. As a reminder, GMIP scores < 30 are *not* suggestive of a GMIP and instead indicate that performance on the WMT was *not* likely influenced by significant cognitive impairment. Failure on at least one PV subtest and GMIP scores ≥ 30 suggest a GMIP, indicating that a patient likely performed poorly on the WMT because of significant cognitive impairment. Descriptive statistics and correlations are shown in Table 39. As can be seen, GMIPness was significantly negatively correlated with LM scores, $r_{Xi}(79) = -.54$, $p < .001$, indicating that the mean of the LM scores in the GMIP group was smaller than the mean of the LM scores from non-GMIP group members. GMIPness accounted for 29% of the variance in LM scores, $r^2_{Xi} = .29$, $p < .001$.

Table 39
Correlation Coefficients, Means, and Standard Deviations for GMIP Performance Dummy Variable and LM Scores

	r		
Variables	*Y*	X_i	r^2_{Yi}
Y	1.00	-.54***	–
X_i	-.54***	1.00	.29***
M	-.73	.21	
SD	.88	.41	

Note. Y = LM scores; X_i = GMIP group. ***$p < .001$.

The regression model was statistically significant, $F(1, 79) = 33.09$, $p < .001$, and accounted for 30% of the variance in LM scores ($R^2 = .30$, Adjusted $R^2 = .29$). As can be seen in Table 40, there was a significant difference between the mean LM scores of GMIP and non-GMIP group members. As a patient changed from not having a GMIP to having a GMIP, LM scores decreased by 1.16 points. In other words, patients in the

GMIP group averaged LM scores 1.16 points lower than patients in the non-GMIP group. Mean LM scores for each GMIP performance group are presented in Table 41.

Table 40
Summary of Regression Analysis for Differences in LM Scores Among GMIP Performance Groups (N = 81)

Variables	*B*	*SE B*	*Beta*
Constant	-.48	.09	
X_i	-1.16	.20	-.54***

Note. R^2 = .30 (p < .001), adjusted R^2 = .29. *** p < .001.

Table 41
GMIP Performance Groups' Mean LM Scores

GMIP Performance Group	*Mean LM Scores*
Non-GMIP	-.48
GMIP	-1.64

Note. Scores are presented as *z*-scores.

Research question 5. How much does each of the WMT subtests explain total GMIP score?

Assumptions of multiple regression. In the hierarchical regressions that were conducted, predictor variables were quantitative and measured at the interval level, and outcome variables were quantitative, measured at the interval level, and unbounded, indicating that Assumption 1 was met. Assumption 2 was met, as all predictors had variances greater than zero.

The assumption of multicollinearity was not met in either regression model. The correlation matrix of predictors for the first model revealed correlation coefficients ranging from .48 to .91. Two pairs of predictors (IR and CNS; MC and PA) correlated very highly (r = .88 and r = .91, respectively). As such, their *b*-values were less trustworthy and the size of R may have been limited. Further, high levels of collinearity between these two pairs of predictors made it difficult to assess the respective importance

of each predictor. There was not as high of a correlation between the two predictors in the second regression model (r = .73); however, as will be described below, multicollinearity was evidenced elsewhere.

VIFs and tolerance statistics were also checked to identify more subtle forms of multicollinearity. No regression model had VIFs greater than 10; however, both models had average VIFs greater than 1 (5.53 and 2.12, respectively), signifying the presence of bias in each model. Tolerance values ranged from .13 to .34 for the first regression model. Four values were below .2, evidencing more cause for concern of collinearity in the model. In the second regression model, tolerance values for both predictors were .47. As those values were greater than .2, they did not reflect a cause for concern of collinearity.

Collinearity diagnostics were also checked for each hierarchical regression model. For the first multiple regression model, the largest difference between eigenvalues was 6.82. This was a large difference and indicated that the solutions of the regression parameters may have been affected by small changes in the predictors or outcome. The condition indexes for the first model also varied greatly from 1 to 81.88, indicating that collinearity was a problem. Finally, eigenvalue variance proportions were checked. As can be seen in Table 42, collinearity was evident between two pairs of predictors in the first multiple regression model. The IR and CNS predictors had substantial variance proportions on eigenvalue 6, and the MC and PA predictors had substantial variance proportions on eigenvalue 5. These sets of predictors were also found to have strong correlations (r = .86 and r = .91, respectively), providing further support for collinearity. As can be seen in Table 43, evidence of strong collinearity was not present when eigenvalue variance proportions were examined for the second regression model.

Table 42
Collinearity Diagnostics for Hierarchical Regression with WMT Subtests as Predictors

	Variance Proportions						
Dimension	*(Constant)*	*WMT IR*	*WMT DR*	*WMT CNS*	*WMT MC*	*WMT PA*	*WMT FR*
1	.00	.00	.00	.00	.00	.00	.00
2	.00	.00	.00	.00	.01	.01	.17
3	.00	.00	.00	.00	.10	.12	.81
4	.11	.07	.01	.08	.04	.15	.00
5	.01	.02	.01	.02	.62	.72	.01
6	.03	.82	.06	.63	.01	.00	.01
7	.84	.09	.92	.27	.22	.00	.00

Table 43
Collinearity Diagnostics for Hierarchical Regression with WMT Composites as Predictors

	Variance Proportions		
Dimension	*(Constant)*	*WMT Memory Composite*	*WMT PV Composite*
1	.00	.00	.00
2	.05	.54	.01
3	.94	.45	.99

Assumption 4 states that predictors should not correlate with external variables that influence the outcome. This assumption was met for both of the regression models, as predictors were not significantly correlated with the potential covariates of sex, race, education, or age.

The assumption of homoscedasticity was not met for either regression model. Examination of standardized residuals versus standardized predicted values plots revealed heteroscedasticity and non-linearity in both regression models. As previously noted, data remained untransformed, as transformations did not improve non-normality or heterogeneity of variance. Failure to meet the assumption of homoscedasticity means that the findings cannot be generalized beyond this sample.

The assumption of independent errors was met for both regression models. Durbin-Watson values were 1.80 for the first model and 1.81 for the second model, suggesting that each models' respective residuals were uncorrelated.

Histograms and normal P-P plots were next examined to check the normality of residuals for Assumption 7. For the first hierarchical regression model, the histogram distribution evidenced a few outliers but otherwise appeared fairly normal; however, evidence of non-normality was apparent on the normal P-P plot. The histogram distribution of the second hierarchical regression model also contained a few outliers but otherwise appeared normal; however, deviations from normality were present on the normal P-P plot.

The assumption of independence was met for both regression models, as all values of each outcome variable came from a different patient. Finally, as noted above, the assumption of linearity was not met for either regression model. As such, the generalizability of findings from this regression is extremely limited.

Assessing the regression models: Diagnostics. Outliers and influential cases were explored to see whether the models fit the observed data well. To detect outliers, standardized residuals were checked. Both regression models had one case with a standardized residual with an absolute value greater than 3.29: the first model had a case with a standardized residual of 4.32 and the second model had a case with a standardized residual of 6.02. As such, these cases were identified as outliers and added to the error level in each model. Additionally, the first regression model had 2.47% of its sample (2 cases) with standardized residuals with absolute values greater than 2.58. Although this value was within 1 to 2% of what was expected, it provided additional evidence that the

model may have been an inaccurate representation of the sample data. Neither regression model had 5% of cases with standardized residuals with absolute values greater than 2. Overall, examination of standardized residuals identified error in both models, thereby suggesting that the models may not have accurately represented the data.

Regression models were also checked for influential cases. Adjusted predicted values were compared to predicted values to ensure that cases did not have large influences over the models. Both models had the same case with a large difference between its adjusted predicted and predicted values, as well as a large Studentized deleted residuals. These values suggested that this case exerted a large influence over the parameters of the respective models.

Cook's distances were next examined. The same case as previously mentioned was again troublesome: its Cook's distance was 22.81 in the first model and 23.84 in the second model. With Cook's distances much larger than 1, this case was considered to have greatly influenced each model as a whole.

Average leverage values were calculated for the regression models. The first model had one case that was greater than two times the average leverage and three cases that were greater than three times the average leverage. The second model had two cases that were greater than two times the average leverage and one case that was greater than three times the average leverage. These findings further supported previously described evidence of the presence of cases with excessive influence over both models.

Mahalanobis distances were examined for high values. Mahalanobis distances ranged from .03 to 55.97 across models, indicating that both models had some cases that exerted undue influence on their respective models.

Standardized DFBeta values were investigated. The first regression model had three cases that had standardized DFBeta values with absolute values greater than 1. The second model had one case with a standardized DFBeta value above the cutoff. These values provide additional support that both regression models contained cases that substantially influenced model parameters.

Finally, CVRs were examined. The first regression model had five cases above and three cases below recommended cutoffs. The second regression model had two cases that fell below recommended cutoffs. These results suggested the presence of cases in both models that influenced the variance of the regression parameters.

Overall, examination of these diagnostic values suggested that influential cases were present in both regression models. Therefore, the models may not have been accurate representations of the sample data.

Relationship between WMT subtests and GMIP score. Correlation and hierarchical regression analyses were conducted to examine how much each of the WMT subtest scores explained GMIP score. Descriptive statistics and correlations are shown in Table 44. As can be seen, there were significant negative correlations between each WMT subtest score and GMIP scores (all *p*s < .001). Patients with higher scores on any of the WMT subtests had lower GMIP scores. FR, PA, and MC scores accounted for substantial levels of variance in GMIP scores. As previously noted, since these (and other) predictors were highly correlated, it was difficult to assess their individual importance in the model.

Table 44
Correlation Coefficients, Means, and Standard Deviations for WMT Subtest Scores and GMIP Scores

	r							
Variables	*Y*	*FR*	*PA*	*MC*	*CNS*	*DR*	*IR*	r^2
Y	1.00	-.83***	-.91***	-.86***	-.47***	-.63***	-.39***	–
FR	-.83***	1.00	.80***	.78***	.56***	.72***	.48***	.69***
PA	-.91***	.80***	1.00	.91***	.62***	.81***	.54***	.83***
MC	-.86***	.78***	.91***	1.00	.65***	.86***	.59***	.74***
CNS	-.47***	.56***	.62***	.65***	1.00	.76***	.88***	.22***
DR	-.63***	.72***	.81***	.86***	.75***	1.00	.66***	.40***
IR	-.39***	.48***	.54***	.59***	.88***	.66***	1.00	.15***
M	28.96	42.75	69.32	74.07	88.18	92.25	91.60	
SD	13.36	18.43	22.27	23.04	12.41	8.37	11.71	

Note. Y = GMIP scores; FR = Free Recall; PA = Paired Associates; MC = Multiple Choice; CNS = Consistency; DR = Delayed Recall; IR = Immediate Recall. ***$p < .001$.

Table 45 displays partial and semipartial correlation coefficients between the predictors and GMIP scores. As can be seen, there were significant relationships between all WMT subtest scores, except for CNS and IR, and GMIP scores after common variance with other predictors was removed from the predictor of interest and the outcome (all *p*s < .001). There were also significant relationships between all WMT subtests scores, except for CNS and IR, and GMIP scores after removing variance that each predictor of interest had in common with other predictors (all *p*s < .001). FR scores uniquely accounted for 40% of the variance in GMIP scores, $sr^2_{FR} = .04$, $p < .001$. PA scores uniquely accounted for 60% of the variance in GMIP scores, $sr^2_{PA} = .06$, $p < .001$. MC scores uniquely accounted for 40% of the variance in GMIP scores, $sr^2_{MC} = .04$, $p < .001$. DR scores uniquely accounted for 50% of the variance in GMIP scores, $sr^2_{DR} = .05$, $p < .001$. CNS and IR scores did not account for any of the variance in GMIP scores, $sr^2_{CNS} = 0.00$, $p = ns$ and $sr^2_{IR} = 0.00$, $p = ns$.

Table 45
Partial and Semipartial Correlation Coefficients for WMT Subtests and GMIP Scores

Variables	pr_i	sr_i	sr^2_i
FR	-.61***	-.19***	.04***
PA	-.70***	-.24***	.06***
MC	-.62***	-.19***	.04***
CNS	-.05	-.01	.00
DR	.69***	.23***	.05***
IR	.15	.04	.00

Note. *** $p < .001$.

The overall regression model was statistically significant, $F(6, 74) = 191.74$, $p < .001$. As shown in Table 46, FR scores alone accounted for 69% of the variance in GMIP scores, $F(1, 79) = 177.85$, $p < .001$. PA scores explained an additional 16% of the variance in GMIP scores, $F_{change}(1, 78) = 85.86$, $p < .001$. CNS scores explained an additional 3% of the variance in GMIP scores, $F_{change}(1, 76) = 17.69$, $p < .001$. Finally, DR scores explained an additional 5% of the variance in GMIP scores, $F_{change}(1, 75) = 65.79$, $p < .001$. MC and IR scores did not add to the explanation of the variance in GMIP scores, $F_{change}(1, 77) = 1.75$, $p = ns$ and $F_{change}(1, 74) = 1.75$, $p = ns$, respectively. Overall, a total of 94% of the variance in GMIP scores was explained by WMT subtest scores. More specifically, holding other predictors constant, as FR scores increased by one point, GMIP scores decreased by .24 points. Standardized beta values indicated that, holding other predictors constant, as FR scores increased by one standard deviation, GMIP scores decreased by .33 standard deviations. Holding other predictors constant, as PA scores increased by one point, GMIP scores decreased by .36 points. Standardized beta values indicated that, holding other predictors constant, as PA scores increased by one standard deviation, GMIP scores decreased by .61 standard deviations. Holding other predictors constant, as MC scores increased by one point, GMIP scores decreased by .32 points.

Standardized beta values indicated that, holding other predictors constant, as MC scores increased by one standard deviation, GMIP scores decreased by .54 standard deviations. Holding other predictors constant, the DR predictor had a significant positive weight (*b* opposite in sign from its correlation with GMIP scores), indicating that higher DR scores predicted higher GMIP scores (suppressor effect). Holding other predictors constant, as DR scores increased by one point, GMIP scores increased by .86 points. Standardized beta values indicated that, holding other predictors constant, as DR scores increased by one standard deviation, GMIP scores increased by .54 standard deviations. With the other predictors held constant, increases in CNS or IR scores did not cause GMIP scores to increase or decrease by significant amounts.

Table 46

Summary of Hierarchical Regression Analysis of the Relationship between WMT Subtests and GMIP Scores ($N = 81$)

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	54.74	2.10		65.97	1.90		66.74	1.98		52.15	3.91		5.24	6.46		3.31	6.59	
FR	-.60	.05	-.83***	-.22	.05	-.30***	-.20	.05	-.28***	-.22	.05	-.30***	-.24	.04	-.33***	-.24	.04	-.33***
PA				-.40	.04	-.67***	-.34	.07	-.56***	-.35	.06	-.58***	-.37	.04	-.61***	-.36	.04	-.61***
MC							-.08	.06	-.14	-.14	.06	-.25*	-.31	.05	-.53***	-.32	.05	-.54***
CNS										.24	.06	.22***	.04	.05	.04	-.03	.08	-.03
DR													.85	.11	.53***	.86	.10	.54***
IR																.09	.07	.08

Note. Model 1: $R^2 = .69$ ($p < .001$), adjusted $R^2 = .69$; Model 2: $R^2 = .85$ ($p < .001$), adjusted $R^2 = .85$, $\Delta R^2 = .16$ ($p < .001$); Model 3: $R^2 = .86$ ($p < .001$), adjusted $R^2 = .85$, $\Delta R^2 = .00$ ($p = ns$); Model 4: $R^2 = .88$ ($p < .001$), adjusted $R^2 = .88$, $\Delta R^2 = .03$ ($p < .001$); Model 5: $R^2 = .94$ ($p < .001$), adjusted $R^2 = .93$, $\Delta R^2 = .05$ ($p < .001$); Model 6: $R^2 = .94$ ($p < .001$), adjusted $R^2 = .94$, $\Delta R^2 = .00$ ($p = ns$). $^{*}p < .05$. $^{***}p < .001$.

Correlation and hierarchical regression analyses were then conducted to examine how much the WMT memory and PV composites explained total GMIP score. Descriptive statistics and correlations are shown in Table 47. As can be seen, there was a significant negative correlation between the WMT memory composite and GMIP scores, $r_{MEM}(79) = -.92$, $p < .001$, indicating that patients with higher WMT memory composite scores had lower GMIP scores. The memory composite accounted for 85% of the variance in GMIP scores, $r^2_{MEM} = .85$, $p < .001$. There was also a significant negative correlation between the WMT PV composite and GMIP scores, $r_{PV}(79) = -.52$, $p < .001$, indicating that patients with higher PV scores had lower GMIP scores. However, the PV composite only accounted for 27% of the variance in GMIP scores, $r^2_{PV} = .27$, $p < .001$.

Table 47
Correlation Coefficients, Means, and Standard Deviations for WMT Composites and GMIP Scores

	r			
Variables	*GMIP*	*Memory Composite*	*PV Composite*	r^2
GMIP	1.00	-.92***	-.52***	
Memory Composite	-.92***	1.00	.73***	.85***
PV Composite	-.52***	.73***	1.00	.27***
M	28.96	62.05	90.68	
SD	13.36	20.05	10.03	

Note. *** $p < .001$.

Table 48 displays partial and semipartial correlation coefficients between the predictors and GMIP scores. As can be seen, there was a significant negative relationship between the WMT memory composite and GMIP scores after common variance with the PV composite was removed from both the memory composite and GMIP scores, $pr_{MEM}(78) = -.93$, $p < .001$. There was also a significant negative relationship between the memory composite and GMIP scores after removing variance that the memory composite

had in common with the PV composite, sr_{MEM} (78) = -.79, $p < .001$. The memory composite uniquely accounted for 62% of the variance in GMIP scores.

There was a significant positive relationship between the WMT PV composite and GMIP scores after common variance with the memory composite was removed from both the PV composite and GMIP scores, $pr_{PV}(78) = .56$, $p < .001$. There was also a significant positive relationship between the PV composite and GMIP scores after removing variance that the PV composite had in common with the memory composite, sr_{PV} (78) = .22, $p < .001$. The PV composite uniquely accounted for 5% of the variance in GMIP scores.

Table 48
Partial and Semipartial Correlation Coefficients for WMT Composites and GMIP Scores

Variables	pr_i	sr_i	sr^2_i
Memory Composite	-.93***	-.79***	.62***
PV Composite	.56***	.22***	.05***

Note. *** $p < .001$.

The overall regression model was statistically significant, $F(2, 78) = 332.91$, $p < .001$. As displayed in Table 49, the WMT memory composite alone accounted for 85% of the variance in GMIP scores, $F(1, 79) = 441.99$, $p < .001$, and the WMT PV composite explained an additional 5% of the variance, $F_{change}(1, 78) = 34.79$, $p < .001$. Thus, a total of 90% of the variance in GMIP scores was explained by the WMT memory and PV composites. Holding the PV composite constant, as memory composite scores increased by one point, GMIP scores decreased by .77 points. Standardized beta values indicated that, holding the other predictor constant, as memory composite scores increased by one standard deviation, GMIP scores decreased by 1.15 standard deviations. Holding the memory composite constant, the PV composite had a significant positive weight (*b*

opposite in sign from its correlation with GMIP scores), indicating that higher PV composite scores predicted higher GMIP scores (a suppressor effect). Holding the memory composite constant, as PV composite scores increased by one point, GMIP scores increased by .42 points. Standardized beta values indicated that, holding the memory composite constant, as PV composite scores increased by one standard deviation, GMIP scores increased by .31 standard deviations.

Table 49

Summary of Hierarchical Regression Analysis Examining the Relationship between WMT Composites and GMIP Scores (*N* = 81)

	Model 1			Model 2		
Variables	*B*	*SE B*	*Beta*	*B*	*SE B*	*Beta*
Constant	67.03	1.90				
Memory Composite	-.61	.03	-.92***			
Constant				38.49	5.09	
Memory Composite				-.77	.04	-1.15***
PV Composite				.42	.07	.31***

Note. Model 1: R^2 = .85 (p < .001), adjusted R^2 = .85; Model 2: R^2 = .90 (p < .001), adjusted R^2 = .89, ΔR^2 = .05 (p < .001). ***p < .001.

CHAPTER V: DISCUSSION

The present study set forth to investigate the base rate of suboptimal performance on the WMT, the relationship between WMT performance and neuropsychological test scores, and the validity of the GMIP in patients with epilepsy. To the author's knowledge, this is the first study to use WMT normative cut scores and GMIP analysis to classify patients into optimal, suboptimal, and GMIP groups, and subsequently examine how such groups performed across a variety of neuropsychological measures. This also appears to be the first study to explore the validity of the GMIP in patients with epilepsy, an essential undertaking if GMIP analysis is to be employed and interpreted with this population. Findings of this study shed light on previously explored (e.g., base rates of suboptimal performance on PVTs) and unexplored (e.g., validity of the GMIP) areas of PV assessment with the epilepsy population.

Base Rates of Optimal, Suboptimal, and GMIP Performance on the WMT

Base rates of suboptimal performance on PVTs in patients with epilepsy have been reported to range from 4 (Hill et al., 2003) to 28% (Loring et al., 2005). This wide range of suboptimal performance on PVTs is unexpected because patients with epilepsy, especially pre-surgical candidates, are presumed to be motivated for neuropsychological evaluation with no apparent external incentives to underperform. Reasons for the variance in base rate of suboptimal performance on PVTs in this population remain unknown and largely unexplored; however, one possible explanation may be the significant cognitive impairment commonly associated with epilepsy (Bortz, 2003).

Non-epileptic patient groups with significant cognitive impairment (e.g., moderate to severe TBI, mental retardation, developmental disorders) typically perform well on PVTs (Heilbronner et al., 2009; Sweet, 1999). However, certain PVTs (e.g., TOMM, FIT, DCT) have been found to have low levels of specificity (high false positive rates) when used with patients with severe cognitive impairment (e.g., dementia and mental retardation) (Boone et al., 2002; Goldberg & Miller, 1986; Philpott, 1992; Schretlen, Brandt, Krafft, & van Gorp, 1991; Spiegel, 2006; Teichner & Wagner, 2004). As patients with epilepsy may have significant cognitive impairment, it is reasonable to consider that such impairment may affect their ability to perform optimally on PVTs. In turn, base rate estimates of suboptimal performance on PVTs in this population may vary considerably because patients have been misclassified as performing suboptimally, when, in fact, they scored below failure cutoffs due to significant cognitive impairment. The present study attempted to investigate this possibility by utilizing the WMT, a highly sensitive and specific PVT. Through GMIP analysis, the WMT indicates whether scores below failure cutoff likely reflect suboptimal performance or significant cognitive impairment.

Using Green's (2005) normative cutoffs, patients were categorized into one of three WMT performance groups: optimal, suboptimal, and GMIP. Results indicated that 43% of the sample (n = 35: 21 pre-surgical, 14 non-surgical) fell into the optimal group; 36% (n = 29: 22 pre-surgical, 7 non-surgical) into the suboptimal group; and 21% (n = 17: 13 pre-surgical, 4 non-surgical) into the GMIP group. The base rate of suboptimal performance attained in this study is higher than previously reported base rates of suboptimal performance on PVTs in the epilepsy population (e.g., 22%; Cragar et al.,

2006; 23%; Hoskins et al., 2010; 28%; Loring et al., 2005), and as discussed below, likely reflects how patients were sorted into WMT performance groups.

Unlike extant research that has employed the WMT in patients with epilepsy (e.g., Drane et al., 2006; Hoskins et al., 2010), this study utilized all WMT subtest scores to categorize patients into performance groups. There were two intentions behind using all subtest scores to sort patients into groups. The first intention was to be able to employ GMIP analysis, a computation that requires all WMT subtest scores, so that patients could be sorted into a GMIP group. The second intention was to be able to use MC and PA scores when determining how patients were categorized into groups. Sorting patients into three performance groups as opposed to the two (pass and fail) typically constructed in PVT studies meant that patients who scored in the caution range on IR, DR, or CNS subtests and those who scored in the warning range on MC and PA subtests were placed into a performance group along with patients who scored in the failure range on IR, DR, or CNS subtests but did not have a GMIP. This group was labeled the suboptimal performance group.

The utilization of all WMT subtest scores to facilitate this type of performance classification system has not yet been carried out in extant research. Instead, studies have sorted patients into pass or fail groups, and patients with WMT scores in the caution range have been placed into the pass group because their IR, DR, or CNS scores were above the 82.5% failure cutoff (Drane et al., 2006; Hoskins et al., 2010). In the present study, considering patients with scores in the caution range as passing the WMT seemed problematic, especially since a seemingly high score of 90% (caution range) on IR or DR is more than two standard deviations below the normal adult mean (Green, 2005).

Additionally, unlike in the current study, previous PVT studies with patients with epilepsy have not included MC and PA scores in the interpretation of WMT performance. The most likely reason why these scores have not been included in interpretation is because they are considered to be memory subtests of the WMT (Green, 2005). However, MC scores ≤ 70% and PA scores ≤ 50% receive a warning rating and are considered suspicious of suboptimal performance when dementia or other profound cognitive impairments have been ruled out (Green, 2005). Moreover, MC scores of 75% and PA scores of 64%, both of which are *above* warning cutoffs, are three standard deviations below the normal adult mean (Green, 2005), providing further support for considering these scores in WMT interpretation. Therefore, using Green's (2005) findings as justification, the current study used MC and PA scores when sorting patients into performance groups. Patients with MC and PA scores in the warning range who did not have GMIPs were placed into the suboptimal group.

Categorizing patients with WMT scores in the caution and warning ranges into the suboptimal instead of optimal group accounted for the high base rate of suboptimal performance achieved in the current study. Indeed, results indicated that, on average, scores in the caution rather than failure range characterized performance across PV subtests in the suboptimal group (IR: $M = 90.78$, $SD = 11.61$; DR: $M = 92.33$, $SD = 3.53$; CNS: $M = 86.38$, $SD = 10.87$). Interestingly, mean MC and PA scores in the suboptimal group were above warning cutoffs (MC: $M = 72.41$, $SD = 14.12$; PA: $M = 65.69$, $SD = 14.06$), signifying that, on average, MC and PA scores were not the scores that drove most patients into this performance group. As the PV subtests are considered easy, and are usually passed by patients with various disorders (e.g., moderate to severe TBI; Green

& Allen, 1999; Green et al., 1999; Green, 2005; neurological disorders; Gorissen, Sanz de la Torre, & Schmand, 2003, cited in Green, 2005), it is likely that many patients in the suboptimal group did not perform to the best of their ability on the WMT.

For example, one patient obtained the following WMT scores: IR = 95, DR = 90, CNS = 90, MC = 90, PA = 70, and FR = 65. GMIP score was 17. The patient's DTBM was in the average range. The PO domain score was in the high average range. VF, LM, AC, and MD domain scores were in the average range and the EF domain score was in the borderline range. The patient was pre-surgical, had been experiencing mainly complex partial seizures for four years, and was taking three AEDs at the time of the evaluation. He/she had 16 years of education, was working full-time, and was not receiving SSDI. As such, history, current level of functioning, and neuropsychological performance did not offer any explanations as to why the patient obtained suboptimal WMT DR and CNS scores. Therefore, it appears as though he/she likely underperformed on the WMT (that is, was a true positive for suboptimal performance on the WMT), and subsequently, likely underperformed on other neuropsychological measures during testing. However, as discussed in the next section, suboptimal WMT performance may not impact all cognitive domains equally.

It is also possible that some patients in the suboptimal group were false positives, that is, identified by the WMT as performing suboptimally, when, in fact, behavioral observations and performance on neuropsychological tests suggested that they performed to the best of their ability during the evaluation. However, upon examination of the data, case examples of potential WMT false positives were unable to be identified. This was due to the fact that patients in the suboptimal performance group averaged DTBM scores

in the average range (albeit in the low end of the average range), thus making it difficult to determine whether such scores were in fact reflective of actual ability level (which would indicate a false positive on the WMT for suboptimal performance) or instead lowered due to suboptimal performance (which would indicate a true positive on the WMT for suboptimal performance, similar to the case example previously described).

The case example above that described a likely true positive for suboptimal performance, as well as the inability to identify a case example demonstrating a clear potential false positive for suboptimal performance, both emphasize the importance of interpreting WMT scores within the context of clinical history and neuropsychological performance, and also encourage neuropsychologists to explore potential explanations for suboptimal WMT scores. The present study did not examine reasons for suboptimal performance; however, as 38% of the sample did not disclose SSDI status, it is possible that some patients may have been applying for disability during the time of the evaluation. Though possible, this scenario is unlikely given the mean seizure duration of 17 years and that patients with intractable epilepsy would likely qualify for disability due to seizure severity alone. Further, most pre-surgical epilepsy patients are considered motivated for surgery, and consequently, assumed motivated for neuropsychological testing. It is even more staggering, then, that when only considering pre-surgical patients, the base rate of suboptimal performance was 39%, with average IR and CNS scores in the caution range and all other WMT scores in the pass range. Thus, the overall high, and slightly higher pre-surgical, base rate of suboptimal performance in the current study is surprising and calls into question how important – or unimportant – it is to differentiate between optimal and suboptimal performance on the WMT in patients with epilepsy.

The arguable importance and necessity of differentiating between optimal and suboptimal WMT performance in patients with epilepsy will be discussed in the next section; however, it is important to note that both groups scored significantly different on all WMT subtests. As hypothesized, large effect sizes were found between optimal and suboptimal groups on all WMT subtests. Cohen's *d* values ranged from .94 to 2.31, indicating that patients in the optimal group scored a minimum of nearly one standard deviation higher on all subtests than patients in the suboptimal group. Also as hypothesized, large effect sizes were found across WMT subtests when comparing patients in the optimal and GMIP groups. Cohen's *d* values ranged from 2.22 to 4.54, signifying that patients in the optimal group scored a minimum of slightly more than two standard deviations higher on subtests than patients in the GMIP group. Finally, again as anticipated, large effect sizes were also observed when comparing patients in the suboptimal and GMIP groups. Cohen's *d* values ranged from 1.00 to 2.35, indicating that patients in the suboptimal group scored a minimum of one standard deviation higher across subtests than patients in the GMIP group. Overall, the presence of such large effect sizes between performance groups across all WMT subtests provided additional support for the categorization of patients into optimal, suboptimal, and GMIP groups.

Finally, results also showed that 21% of the sample was classified into the GMIP group, suggesting that these patients scored below failure cutoff not because of suboptimal performance but because of significant cognitive impairment. The attained base rate of GMIP performance was higher than the presumed 10% false-positive rate reported by Drane et al. (2006), although it should be noted that Drane et al. did not employ GMIP analysis. In the current study, WMT performance in the GMIP group was,

on average, characterized by scores below failure cutoff on all PV subtests (IR: *M* = 78.68, *SD* = 12.53; DR: *M* = 79.12, *SD* = 7.12; CNS: *M* = 72.12, *SD* = 9.10) and warning scores on the memory subtests (MC: *M* = 39.71, *SD* = 14.52; PA: *M* = 38.82, *SD* = 10.83). The mean GMIP score in this group was 43.47 (*SD* = 5.78), which was substantially greater than the minimum 30-point normative GMIP inclusion criteria. Given that 52% of patients were diagnosed with TLE (28% left TLE; 24% right TLE), a seizure disorder associated with impairments in memory and various other areas of cognitive functioning (Grote, Smith, & Ruth, 2001; Hermann et al., 2006; Hermann, Seidenberg, Lee, Chan, & Rutecki, 2007; Keary, Frazier, Busch, Kubu, & Iampietro, 2007), significant cognitive impairment may have negatively impacted some patients' abilities to perform above failure cutoff on the WMT. As will be discussed below, this would only hold true if patients in the GMIP group displayed significant cognitive impairment on neuropsychological measures, and if the GMIP was found to be a valid indicator of significant cognitive impairment in patients with epilepsy.

WMT Performance and Neuropsychological Test Scores

PVT scores have been found to account for approximately 50% of the variance in overall neuropsychological performance in non-epileptic populations, with lower PVT scores typically associated with significantly lower test scores across most cognitive domains (e.g., Constantinou et al., 2005; Green et al., 2001; Green et al., 2002; Rohling et al., 2002). Within the epilepsy population, although the research is more limited, suboptimal PVT scores have also been associated with significantly lower performance across a variety of neuropsychological measures (Drane et al., 2006; Locke et al., 2006; Loring et al., 2005). The current study is the first of its kind to investigate differences in

neuropsychological test scores among WMT optimal, suboptimal, and GMIP performance groups.

Results of the present study are largely consistent with findings from related extant literature. Generally, support was found for WMT performance accounting for variance in overall neuropsychological performance, albeit not in as high a percentage as previously reported values (e.g., 49-54%; Green et al., 2001). Results suggested that PV does not affect cognitive domains equally, a finding that is consistent with and expands upon results from existing studies examining PV in the epilepsy population. Findings also indicated that patients in the suboptimal and GMIP performance groups performed significantly lower on most, but not all, neuropsychological measures than patients in the optimal group, adding further support to the notion that PVT scores impact neuropsychological performance.

First, results of the present study indicated that WMT performance and education level accounted for 35% of the variance in overall neuropsychological performance as measured by the DTBM, a composite of cognitive domain scores. Education level alone explained 6% of the variance in DTBM scores, which is relatively consistent with results from Green et al. (2001) indicating that education level explained 11% of the variance in OTBM scores, which represented average performance across a variety of neuropsychological measures. Current results revealed that WMT performance accounted for an additional 29% of the variance in DTBM scores. Although these findings support well-established notions that PV accounts for a certain percentage of the variance in neuropsychological test scores, the achieved rate was considerably lower than some of the higher rates reported in studies with litigant and compensation claimant populations

(e.g., 49-54%; Green et al., 2001; 49%; Green et al., 2002; 47%; Constantinou et al., 2005; 36-45%; Rohling et al., 2002; 35%; Stevens et al., 2008). Instead, current findings more closely resemble results of Bowden, Shores, and Mathias (2006) and Rohling and Demakis (2010), which collectively demonstrated that PV explained 13-25% of the variance in neuropsychological performance in TBI litigants.

It remains unclear why the current study did not attain a rate as high as those found in previous studies (e.g., Constantinou et al., 2005; Green et al., 2001; Green et al., 2002; Rohling et al., 2002; Stevens et al., 2008); however, differences in PV measures, study design, and patient populations might have contributed to varying results. For example, Green et al. (2001) constructed a PV composite from the three blocks of the CARB; WMT IR, DR, and CNS scores; and the CVLT Logit formula. Green et al. (2002) used the same PV composite as Green et al (2001). Constantinou et al. used TOMM Trial 2 scores as their main PV measure. Stevens et al. used WMT IR, DR, and CNS scores, and MSVT IR and DR scores as their PV measures. Bowden et al. (2006) and Rohling and Demakis (2010) both used the WMT IR score as their main indicator of PV. As such, the PV measures employed in these studies might be considered "purer" measures of PV when compared to the measure that was utilized in this study – the entire WMT. As previously described, using all WMT subtests meant that both PV and memory subtest scores were used when classifying WMT performance. Current findings may therefore reflect that WMT scores used in analyses measured PV *and* memory, and indicate that overall, the WMT was not as "clean" a measure of PV as the PVTs used in other studies.

Differences in study design may have also impacted results. For example, Green et al. (2001), Green et al. (2002), and Rohling and Demakis (2010) examined the impact

of PV on the OTBM, a composite of average performance across a variety of neuropsychological measures. Other studies explored the impact of PV on summary indices or individual subtests of Wechsler IQ and memory tests (e.g., Bowden et al., 2006; Constantinou et al., 2005; Stevens et al., 2008), the overall neuropsychological deficit score on the HRNB-A (Constantinou et al., 2005), or on stand-alone neuropsychological tests (e.g., TMT; Stevens et al., 2008). The present study used the DTBM and cognitive domain scores as outcome measures. Thus, the use of different outcome variables, which measure neuropsychological constructs in slightly different ways, may have accounted for some of the discrepancy found in the rates of variance in neuropsychological performance accounted for by PV.

Additionally, differences in patient population may have influenced results. As noted above, most studies that have examined the relationship between PVT and neuropsychological test scores have done so with litigant and claimant TBI patients. Though relatively unexplored, PV may impact neuropsychological performance differently in other patient populations that have seemingly little or no motivation to underperform. Although a small number of studies have explored the relationship between PVT scores and neuropsychological performance in patients with epilepsy (e.g., Drane et al., 2006; Locke et al., 2006; Loring et al., 2005), none did so utilizing a design similar to that of the current study. Thus, estimates of the amount of variance in overall neuropsychological performance attributable to PV have not been provided for the epilepsy population. As will be discussed below, Locke et al. came the closest to providing such estimates by reporting how much TOMM performance explained variance in individual cognitive domains, but did not construct an overall measure of

neuropsychological performance such as the DTBM or OTBM. It was therefore difficult to compare this study's overall variance accounted for rate of 29% with findings from Locke et al. or studies conducted with other patient populations.

Another major, and arguably more intriguing, finding was that WMT performance did not affect each cognitive domain equally. Similar to previous findings (e.g., Green et al., 2001; Green et al., 2002), WMT performance had the most powerful effect on the LM domain, which represented scores on learning and memory tests. Specifically, WMT performance accounted for 40% of the variance in LM scores. This rate was much higher than that attained by Locke et al. (2006), who found that TOMM scores explained 7% of the variance in their Memory Functioning domain. However, the LM composite in the current study was comprised of 13 measures of verbal and visual memory, and Locke et al.'s Memory Functioning composite contained only the WMS-III Immediate Memory and General Memory indices. Thus, the LM composite likely represented a wider range of learning and memory functioning than the composite used by Locke et al., which may help explain the higher rate of LM variance accounted for by PV attained in this study. Additionally, it is of note that the LM domain in the current study consisted mostly of verbal memory tests, and that the WMT is a PVT that relies on verbal memory, particularly during its more challenging memory subtests. It is therefore reasonable to suggest that the WMT accounted for such a substantial proportion of the variance in LM scores because some of its subtests (e.g., MC, PA, FR) are also measures of verbal memory.

WMT performance was also found to have a significant, but less powerful, impact on the VF, AC, EF, PS, and PO domains. Holding education level constant, results

indicated that WMT performance explained 20% of the variance in the VF and AC domains, which represented scores on verbal functioning and attention and concentration tests, respectively. WMT performance accounted for 19% of the variance in the EF domain, which represented scores on executive functioning measures; however, these results must be interpreted with caution and are not generalizable, as the relationship between WMT performance and EF scores was not linear. Again holding education level constant, WMT performance accounted for 18% of the variance in the PS domain, which represented scores on processing speed tests, and 17% of the variance in the PO domain, which represented scores on tests of perceptual organization and reasoning. Contrary to previously reported findings (e.g., Green et al., 2001; Green et al., 2002; Locke et al., 2006), WMT performance did not significantly impact motor functioning scores. Education level accounted for 5-9% of the variance in cognitive domain scores, which is relatively consistent with the 11% rate found by Green et al. (2001). Overall, these results revealed that WMT performance explained more variance in most cognitive domains than what has been previously reported in patients with epilepsy. Locke et al. is the only other study to report such rates in patients with epilepsy and found that TOMM performance accounted for 4 to 9% of the variance in various cognitive domains.

As demonstrated by the current findings and those of Locke et al. (2006), although PVT performance accounted for some of the variance in neuropsychological performance, the majority of this variance remained unexplained. Locke et al. explored more potential contributors than the present study and found that individually, medical variables (e.g., current number of AEDs, duration of seizure disorder), diagnosis (e.g., epilepsy vs. PNES), neuropathology, psychopathology, and PV explained relatively small

amounts of variance in several cognitive domains. For example, psychopathology explained 3% of the variance in memory scores. However, when considered together, the combination of predictors explained a significant percentage of variance in different cognitive domains. For example, the combination of predictors described above explained 32% of the variance in memory functioning. Similar to current results though, the majority of the variance in cognitive domain scores still remained unaccounted for in Locke et al.'s study.

The current study did not replicate Locke et al.'s (2006) design; that is, other possible explanatory variables related to seizures (e.g., recent seizure activity, seizure location) or other areas of functioning (e.g., psychopathology) were not included in current regression models. Such variables were not included because they were considered to fall outside of the main focus area of this study, which was exploring the relationship between WMT performance and neuropsychological test scores in an epilepsy population. Therefore, it remains unknown how incorporating such potential explanatory variables into regression models may have impacted results, and, similar to Locke et al.'s findings, the majority of the variance in cognitive domain scores remains unaccounted for. This unexplained variance should be further investigated in future studies that include a wide range of possible explanatory variables, such as those relating to seizure disorder, medical status, AEDs and other medications, neuropathology, psychopathology, PVT performance, disability status, and demographics.

Finally, perhaps the most noteworthy results of this section are related to differences observed in cognitive domain scores and overall neuropsychological performance among WMT performance groups. Results generally supported the

hypothesis that patients in the optimal performance group would have significantly higher neuropsychological test scores than patients in the suboptimal and GMIP groups. However, as will be discussed, this hypothesis was not supported in all cognitive domains, and an unanticipated subtlety emerged when examining DTBM scores.

First, holding constant education level when necessary, patients in the GMIP group averaged significantly lower scores on the DTBM and all cognitive domains except for the MD domain (the MD regression model was not statistically significant) than patients in the optimal group. At face value, this finding appears to add support to the validity of the GMIP in identifying patients who scored below WMT failure cutoff due to significant global cognitive impairment. Upon closer examination, it became apparent that although there were significant differences on various indicators of neuropsychological performance between GMIP and optimal groups, those differences were not consistently practically significant, i.e., clinically meaningful. For example, patients in the optimal group averaged VF, PO, EF, AC, and DTBM scores in the average range, whereas GMIP patients' average scores were in the low average range. These results suggested that patients in the GMIP group were not markedly impaired on most neuropsychological tests, the significance of which will be discussed below. Significant and clinically meaningful differences were observed between the optimal and GMIP groups on LM and PS domain scores: on both of these measures, the optimal group averaged scores in the average range and the GMIP group averaged scores in the borderline range.

These results provided initial support for the validity of the GMIP in identifying patients who scored below WMT failure cutoff due to significant impairment on tests of

learning and memory, as defined by LM scores in the borderline range. These results also provided additional support for previously discussed findings that indicated that PV does not impact cognitive domains equally, even in patients who demonstrate some level of impairment on testing.

Rather unexpectedly, these results also indicated that GMIP patients did not demonstrate significant cognitive impairment across all neuropsychological measures. In fact, on average, patients in the GMIP group averaged DTBM scores in the low average range. As will be discussed below, this was surprising given that approximately half of the sample was diagnosed with TLE and 69% of patients were pre-surgical, both of which typically indicate the presence of significant cognitive impairment that likely impacts aspects of functioning.

It remains unclear why substantial clinically meaningful differences were not consistently found between patients in the optimal and GMIP groups on any cognitive domains except for LM and PS. The most likely explanation for the lack of consistent practically significant results may have to do with sample characteristics. Results suggested that, on average, participants did not demonstrate marked cognitive impairment in most areas of testing. In fact, mean overall neuropsychological performance was in the average range for optimal and suboptimal performance groups (DTBM *M* *z*-scores = -.37 and -.68, respectively), and in the low average range for the GMIP group (DTBM *M* *z*-score = -1.26).

The lack of a considerably cognitively impaired sample was surprising, given that approximately half the sample (52%) was diagnosed with TLE, which has been shown to be associated with impairments in all areas of cognitive functioning (e.g.,

memory, attention, language, intelligence, executive functioning; Hermann et al., 2006; Hermann et al., 2007; Keary et al., 2007), as well as slow but continuous cognitive deterioration (Jokeit & Ebner, 1999). Further, 69% of the sample was pre-surgical, signifying that the majority of patients had intractable epilepsy that may have had a negative impact on various areas of functioning (e.g., cognitive, affective, behavioral, activities of daily living). However, since the sample was, on average, relatively young (M = 39.98, SD = 14.28) and, on average, had not been experiencing seizures for the majority of their lives (M = 16.96, SD = 1.75), it is likely that most patients were not yet displaying severe global impairment on testing.

Findings from Jokeit and Ebner (1999) support such a possibility, as results of their study showed that patients with a longer duration of refractory TLE (30+ years) exhibited more severe cognitive impairment on the WAIS-R than patients who had had the disorder for shorter durations (<15 and 15-30 years). Additionally, a large percentage of the current sample (47%) was employed, offering further evidence that many patients did not demonstrate the type of marked cognitive impairment that would likely impede involvement in the workforce.

Moving on, examining differences in cognitive domain and DTBM scores between optimal and suboptimal performance groups revealed mixed results. Overall, results suggested that suboptimal performance on the WMT did not uniformly impact all cognitive domains or the DTBM. Specifically, significant differences were not found between these two groups on VF, PO, PS, and MD domain scores, indicating that such scores were relatively similar despite optimal or suboptimal WMT performance.

Significant differences were found between optimal and suboptimal groups on EF, LM, and AC scores: patients in the suboptimal group averaged significantly lower scores than patients in the optimal group. More specifically, patients in the suboptimal group averaged scores in the low average range on these outcome variables, whereas patients in the optimal group averaged scores in the average range. The only exception to this trend was observed in DTBM scores: though scores between the two groups were significantly different, these differences were of little practical significance as both groups scored in the average range. Overall, in the absence of marked cognitive impairment on neuropsychological testing, results suggested that patients in the suboptimal performance group did not perform to the best of their ability on certain measures of neuropsychological functioning (e.g., measures of executive functioning, learning and memory, attention/concentration), and thus, that their scores in these domains underestimated actual ability levels.

These findings are consistent with those found in Loring et al. (2005) and Keary et al. (2013), the only studies to stratify patients with epilepsy into valid, questionable, and invalid PVT groups based on normative VSVT cut scores. Similar to current results, Loring et al. and Keary et al. found that patients in the questionable group scored significantly lower on a variety of WAIS-III and WMS-III indices than patients in the valid group. Also similar to current results, Loring et al. and Keary et al. found that patients in both groups almost always scored in different ranges. For example, in the Loring et al. study, patients in the valid group scored in the low average range on the FSIQ, VIQ, and PIQ indices, whereas patients in the questionable group scored in the borderline range. Keary et al. found that patients in the valid group scored in the average

range on three WMS-III indices whereas patients in the questionable group scored in the low average range, a finding consistent with current results. Overall, consistency between results of Loring et al., Keary et al., and the current study provides additional support for the way in which patients were categorized into WMT performance groups. Similarities between results of those studies and the current one also provide further support for the well-established notion that poor PVT scores are associated with significantly lower neuropsychological test scores across a range of cognitive domains. .

All things considered, two potential reasons emerged for the lack of consistent, practically significant differences in neuropsychological test scores among WMT performance groups in the current study. First, consistent with extant research, suboptimal WMT scores did not impact performance on all cognitive domains equally (e.g., suboptimal WMT scores were associated with significantly lower learning and memory scores but not with significantly lower verbal functioning scores). Further, the impact that suboptimal performance had on the cognitive domains it affected was arguably clinically negligible (e.g., lowered scores from the average range into the low average range), though knowing that such scores were lowered due to suboptimal PV would be useful when interpreting lower-than-expected test scores in those domains. Second, the lack of a significantly globally impaired sample likely limited the extent to which substantial, clinically meaningful differences in test scores could be found among groups. As such, further research replicating the present study's design with a more clinically impaired sample of patients with epilepsy may help expand on current results.

Validity of the GMIP

Using GMIP analysis, the WMT has been found to have a high level of specificity (e.g., 98%; Green et al., 2011) in patients with severe memory impairment (e.g., dementia) (Green et al., 2003; Green et al., 2011). Similarly, the GMIP of the MSVT, a shorter version of the WMT, has been shown to have a high level of sensitivity (84%) in patients with dementia, indicating that the GMIP correctly classified patients with dementia as having dementia 84% of the time (Howe & Loring, 2009). These studies support the validity of the GMIP in accurately differentiating between WMT and MVST scores below failure cutoff due to significant cognitive impairment and those below failure cutoff due to suboptimal performance in patients with dementia; however it remains unknown if the GMIP can accurately do so in other patient populations that may have significant cognitive impairment. The current study was the first to explore the validity of the GMIP in patients with epilepsy by investigating the relationship between GMIP scores and scores on neuropsychological memory tests. The present study also examined how much each WMT subtest and constructed WMT memory and PV composites explained GMIP scores to further examine the validity of the GMIP. Overall, results largely provided support for the validity of the GMIP in patients with epilepsy, though a slight threat was identified.

Overall, results supported the hypotheses that GMIP scores would predict LM scores, and, more specifically, that patients with GMIPs would have significantly lower LM scores than patients without GMIPs. In the first regression model, GMIP scores were used in a simple linear regression analysis to predict LM scores. GMIP scores accounted for 55% of the variance in LM scores, indicating that GMIP scores were fairly strong predictors of LM scores. As anticipated, there was a strong negative relationship between

GMIP scores and LM scores (r = -.74, p < .001), signifying that patients with higher GMIP scores had lower LM scores. These results suggested that the higher the GMIP score, the more likely a patient scored below WMT failure cutoff because of significant memory impairment. This is what would be expected if the GMIP were a valid indicator of WMT scores below failure cutoff due to significant memory impairment.

Results of the first regression model also indicated that 45% of the variance in LM scores was *not* explained by GMIP scores. Such a high variance unaccounted for rate suggested that other potential contributory factors that were not included in the model (e.g., seizure data, neuropathology, psychopathology) might have explained nearly as much variance in LM scores as GMIP scores did. As previously noted, such potential explanatory variables were not included in regression models because they were considered to fall outside of the main focus of this study, which was exploring the relationship between WMT performance and neuropsychological test scores in patients with epilepsy. Future studies are encouraged to replicate the current design and include other potential contributory variables to allow for a more thorough investigation of the validity of the GMIP in the epilepsy population.

The second regression model explored differences in LM scores between GMIP performance groups. Based on normative cut scores, patients who did not meet GMIP criteria were categorized into the non-GMIP group and those who met GMIP criteria (failed at least one PV subtest and had GMIP scores ≥ 30) were categorized into the GMIP group. Results revealed that GMIP performance explained 30% of the variance in LM scores, signifying that 70% of the variance in LM scores remained unaccounted for. Again, it remains unknown what other factors might have contributed to the unexplained

variance in LM scores, as additional potential explanatory variables were not included in the model for reasons previously described. Patients in the GMIP group averaged significantly lower LM scores than patients in the non-GMIP group. Specifically, patients in the GMIP group averaged LM scores 1.16 points lower than patients in the non-GMIP group. This resulted in patients in the non-GMIP group averaging LM *z*-scores of -.48, which were in the average range, and patients in the GMIP group averaging LM *z*-scores of -1.64, which were in the borderline range. These findings were consistent with previously described regression results that found that GMIP patients averaged LM scores in the borderline range.

These findings provide support for the sensitivity of the GMIP to significant memory impairment (as defined by LM scores in the borderline range) in patients with epilepsy. More specifically, these results suggest that GMIP scores are able to validly distinguish between patients who score in the average and borderline ranges on LM measures. This differentiation is of practical significance in interpreting WMT performance because it provides a possible explanation for WMT scores below failure cutoff.

Although results provided initial support for the sensitivity of the GMIP to significant memory impairment, examination of the data also revealed the potential for GMIP true and false positives similar to those identified when discussing suboptimal scores. A GMIP true positive would indicate that the GMIP was an accurate reflection of WMT scores below failure cutoff due to significant memory impairment. On the other hand, a GMIP false positive would suggest that the GMIP was invalid because of suboptimal performance throughout the evaluation including during the WMT (so much

so that at least one PV subtest was failed and a ≥ 30-point difference between the mean of the easy and hard subtests was achieved). Two case examples illuminate these potential classification inaccuracies.

In the first case example, a patient obtained the following WMT scores: IR = 58, DR = 73, CNS = 65, MC = 40, PA = 30, and FR = 10. GMIP score was 38. This patient's DTBM was in the extremely low range. The majority of his/her cognitive domain scores were also in the extremely low range, though PO and MD domain scores were in the borderline range. The patient had been experiencing seizures for 29 years and was pre-surgical. He/she was averaging 20 complex partial seizures per month and taking one AED during the time of the evaluation. He/she had 11 years of education, was unable to work because of his/her epilepsy, and was receiving SSDI during the time of the evaluation, thereby seemingly indicting a lack of an external financial incentive to underperform. Taking into consideration clinical history and overall neuropsychological performance, this patient's GMIP score of 38 was likely an accurate indication of WMT scores below failure cutoff due to significant memory impairment as well as likely substantial global cognitive impairment.

A second case example emphasizes the need for interpreting GMIP scores within the context of overall neuropsychological profile and clinical history. The second patient was also pre-surgical and had been experiencing partial seizures for 36 years. Seizure frequency was not noted in the records, but the patient was taking one AED at the time of the evaluation. He/she had 12 years of education, was employed full-time, and did not disclose SSDI status. As SSDI status was unknown, it was possible that an external incentive to underperform on testing was present. The patient scored in the average range

on the DTBM and all cognitive domains except for LM and PS, where scores were in the low average range. WMT scores were as follows: IR = 80, DR = 75, CNS = 60, MC = 40, PA = 35, and FR = 23. GMIP score was 39. Only considering WMT scores, it appeared as though the patient may have scored below failure cutoffs due to cognitive impairment. However, the patient's clinical history and largely average range performance on various measures of neuropsychological functioning suggested that his/her WMT performance could more accurately be described as suboptimal. Further, considering WMT scores within the context of his/her low average learning and memory scores provides additional evidence of suboptimal WMT performance, as someone with low average memory performance should be able to complete the WMT without difficulty. In this case, although greater than the ≥ 30 cutoff, the GMIP score of 39 was a false positive for a GMIP, and instead indicated a false negative for the WMT. That is, this patient was identified by the WMT as scoring below failure cutoff because of the possibility of significant cognitive impairment, whereas significant cognitive impairment was not evident on testing or in daily functioning. Therefore, the patient's WMT performance was more accurately described as suboptimal and not reflective of significant cognitive impairment.

The two cases described above serve as reminders that although the specificity of the GMIP has been found to be quite high, i.e., in the 90s, in various patient populations (e.g., Green et al., 2011; Henry et al., 2009; Howe et al., 2007; Howe & Loring, 2009), the GMIP will not 100% accurately identify every patient who scores below WMT failure cutoff due to significant memory impairment. As such, it remains good clinical practice to interpret GMIP scores within the context of overall neuropsychological

performance and clinical history. This is especially true considering that, similar to results on any neuropsychological test, true and false positives for the GMIP are bound to emerge.

Moving on, results of the final two regression models largely provided support for the validity of the GMIP, though a slight threat was identified. It should be noted that these models violated many of the assumptions of multiple regression, including that of linearity, and contained influential cases. Therefore, the models may not have accurately represented the sample data and results are not generalizable.

Results of the first regression model revealed that 94% of the variance in GMIP score was explained by WMT subtest scores, and mostly supported the hypothesis that WMT memory subtest (MC, PA, FR) scores would account for a greater proportion of the variance in GMIP scores than would PV subtest (IR, DR, CNS) scores. More specifically, FR scores accounted for the majority (69%) of the variance in GMIP scores, with PA scores explaining an additional 16% of the variance. CNS and DR scores explained small proportions of the variance in GMIP scores (3 and 5%, respectively). IR and MC scores did not add to the explanation of GMIP score variance. Increases in FR, PA, and MC scores were associated with decreases in GMIP scores, indicating that higher performance on WMT memory subtests was associated with a lower possibility of scoring below WMT failure cutoffs due to significant memory impairment. Overall, since two of the three memory subtest scores (FR and PA) accounted for the majority of the variance in GMIP scores, preliminary support was provided for the sensitivity of the GMIP score to WMT memory subtest performance, and thus, to general memory ability. These results thus suggest that the GMIP score is likely a valid indicator of WMT scores

below failure cutoff due to significant memory impairment (as defined by neuropsychological test scores in the borderline range in this study).

Further support for, as well as a slight threat against, the validity of the GMIP score was obtained in a final regression that examined how much the WMT memory and PV composites explained GMIP score. As hypothesized, findings indicated that the WMT memory composite explained a greater proportion of the variance in GMIP scores than did the PV composite. The WMT memory composite accounted for 85% of the variance in GMIP scores, with the PV composite explaining an additional 5% of the variance. Increases in memory composite scores were associated with decreases in GMIP scores, demonstrating that better performance on the WMT memory subtests was associated with lower GMIP scores. This finding makes sense given that lower GMIP scores would be expected to be associated with higher memory scores, thereby suggesting that WMT performance was not impacted by significant memory impairment. Echoing conclusions drawn from the first regression model, these results provide additional support for the validity of the GMIP score as an indicator of WMT scores below failure cutoff due to poor performance on memory tests.

Finally, an unanticipated finding emerged when examining the correlation matrix of the final regression model that posed a slight threat to the validity of the GMIP. The correlation matrix indicated that the PV and memory composites were strongly correlated ($r = .73$, $p < .001$). Such a strong correlation suggested that they measured a similar construct and thus posed a slight threat to the validity of the GMIP score. The magnitude of this relationship was also not anticipated, as the PV composite is composed of subtests that are purported to be measures of PV and the memory composite is composed of

subtests that are supposed to be measures of memory (Green, 2005). However, this finding makes more sense when put into context; that is, some amount of memory functioning is required to score above failure cutoff on the PV subtests. As noted above, though, such an unexpected finding may have also reflected the model's inaccurate representation of the sample data and, consequently, is not generalizable to other samples.

As exemplified by results discussed in this section, preliminary support was found for the validity of the GMIP in the current epilepsy sample. Depending on the regression model, GMIP performance explained from 30-55% of the variance in LM scores. Overall, GMIP group patients had significantly lower LM scores than non-GMIP group patients. On average, the GMIP group demonstrated LM scores in the borderline range, compared with LM scores in the average range for non-GMIP group patients. Thus, on average, GMIP scores signified the presence of borderline memory impairment on testing, providing initial support for the validity of the GMIP in this population. Despite these encouraging results, neuropsychologists are still urged to interpret GMIP scores within the context of clinical history, behavioral observations, and overall neuropsychological performance, as, like with any test, there is always some amount of error associated with classification accuracy.

Limitations and Directions for Future Research

As with all research, methodology is an important factor to consider in the interpretation of current findings. The present study was the first of its kind to categorize patients into WMT performance groups, and to subsequently examine how such groups performed across a variety of neuropsychological measures. Although WMT classifications were based on normative cut scores (Green, 2005) and patients in each

group had significantly different scores across all WMT subtests, results of subsequent analyses may have differed had an alternate design been employed. For example, would more practically significant differences on certain neuropsychological outcomes (e.g., significantly different scores in different ranges on the DTBM in patients in the optimal and suboptimal groups) have emerged between groups if patients with IR, DR, and CNS scores in the caution range and patients with MC and PA scores in the warning range had been placed into the optimal instead of suboptimal group? Such a design would have left patients with WMT scores in the failure range (at least one IR, DR, or CNS score ≤ 82.5% and GMIP scores < 30) in the suboptimal group. Patients with GMIPs would have remained in the GMIP group. As the current study did not employ such a design, it remains unknown if such proposed changes to group categorization would have drastically altered results, yet the possibility remains. Future studies are encouraged to employ this suggested modification in design and then re-examine the relationship between WMT performance and neuropsychological test scores in patients with epilepsy.

An additional potential limitation regarding design was that the LM domain was comprised of verbal and visual memory tests, therefore potentially framing LM as a unitary construct. Theoretically, LM is not believed to be a unitary construct, but rather an umbrella construct under which verbal and visual memory fall as unique yet related types of memory. In practice, neuropsychologists conceptualize and interpret performance on verbal and visual memory tests differently, particularly in patients with TLE (52% of the sample) who may display impairments on verbal and/or visual memory tests depending on seizure type, focus, and lateralization. The present study included both types of memory tests in the LM domain in an attempt to represent general memory

functioning, similar to the approach taken by Green et al. (2001), Green et al. (2002), and Rohling and Demakis (2010). However, upon examination of the composition of the current LM domain, it became apparent that the majority of the domain (69%) was comprised of verbal memory tests. Such a composition – verbal memory heavy – was quite similar to that of the LM domains constructed in studies that utilized similar designs (Green et al., 2001; Green et al., 2002; Rohling & Demakis, 2010), thereby providing preliminary justification for the composition of the current LM domain.

The LM domain in the current study was not deconstructed into two separate domains – verbal memory and visual memory – for a variety of reasons. First, reliability analysis results revealed that the LM domain possessed a strong level of reliability (α = .93), indicating that all tests that comprised the domain represented the overarching construct of memory. Next, one of the main goals of this study was to examine the relationship between WMT scores and overall neuropsychological performance as measured by the DTBM. The DTBM was created by computing the average of cognitive domain *z*- scores; therefore, having two memory domains instead of one would have had little impact on overall DTBM scores. Further, as previously mentioned, other studies that utilized a similar design (constructing cognitive domains and an OTBM or DTBM) did not separate verbal and visual memory tests into their own domains. Instead, Green et al. (2001), Green et al. (2002), and Rohling and Demakis (2010) included both verbal and visual memory tests in their LM domains, which represented general memory functioning.

For the sake of thoroughness, analyses in the current study were re-conducted to explore whether splitting the LM domain into separate verbal and visual cognitive domains would impact results. Results of the re-analyses were very similar to results of

the original analyses and did not result in patients in any WMT performance group scoring differently than they did on the LM domain or DTBM in initial analyses. Thus, exploratory re-analyses results revealed that constructing two LM domains failed to meaningfully impact results, thereby providing preliminary support for the inclusion of both verbal and visual memory tests in the LM domain. It remains unknown if results would have been different in both the initial analyses and re-analyses had the sample demonstrated significant cognitive impairment. Future studies are therefore encouraged to implement the modified design (separate verbal and visual memory domains) with markedly impaired patients with epilepsy to explore potential impacts on results.

Another limitation emerged during data analysis: on average, the current sample did not demonstrate significant cognitive impairment on testing; that is, other than the GMIP group averaging scores in the borderline range on LM and PS domains. This limitation was unexpected, as the majority of the patients were pre-surgical. Typically, epilepsy surgical candidates have poor to no seizure control, and, accordingly, may demonstrate significant cognitive impairment in activities of daily living and on neuropsychological testing (Bortz, 2003; Tavakoli et al., 2011). However, since the current sample was, on average, young (M = 39.98, SD = 14.28) and, on average, had been experiencing seizures for less than 20 years (M = 16.96, SD = 1.75), it is likely that most patients were not yet exhibiting signs of marked impairment on testing (Jokeit & Ebner, 1999). As such, the lack of a largely severely cognitively impaired sample may have impacted the regression models exploring the relationship between WMT performance and neuropsychological test scores. More specifically, the presence of a more markedly cognitively impaired sample may have led to the emergence of practically

significant differences on other neuropsychological measures in addition to the LM and PS domains among optimal and GMIP groups. Future research is therefore encouraged to replicate the current design with a sample of epilepsy patients who display more global impairment on neuropsychological testing. Additional research might also wish to replicate this study using a sample of patients with longer average durations of seizure disorder than patients in the current study, thereby increasing the chances of obtaining a sample with significant cognitive impairment.

Next, pre-surgical and non-surgical patients were both included in the current study. The majority of extant literature exploring PVT within the epilepsy population has not reported surgical status (e.g., Cragar et al., 2006; Drane et al., 2006; Hill et al., 2003; Hoskins et al., 2010). Instead, patients have been described as undergoing video-EEG monitoring to determine candidacy for resection surgery, and decisions regarding surgical candidacy status have not typically been reported. Therefore, non-surgical patients may have been included in previous studies, but this remains unknown because of how samples have been described. The present study reported surgical status and included both pre- and non-surgical patients in analyses. Including both pre- and non-surgical patients may have impacted findings, as, for example, non-surgical patients may not be as motivated to undergo neuropsychological evaluation as pre-surgical patients. Non-surgical patients may be less motivated for testing because the results of their evaluation may be less critical to their medical care as potential surgical risks and outcomes are not being determined. However, current results indicated that a higher percentage of pre-surgical patients fell into the suboptimal performance group than did non-surgical patients (39 versus 28%, respectively), indicating that the majority of non-surgical

patients in the current sample appeared motivated for testing. As patients in WMT performance groups were not further stratified into groups based on surgical status, the potential effect of surgical status on neuropsychological functioning and possible interaction effect of WMT performance and surgical status on neuropsychological functioning was not explored. Future research is therefore encouraged to more accurately describe sample characteristics including surgical status, and to explore the possible impact of surgical status on WMT performance and neuropsychological functioning in patients with epilepsy.

Similarly, the current sample did not include post-surgical patients with epilepsy. Post-surgical patients were not included because pre-surgical and post-surgical WMT and neuropsychological data were not available for the majority of patients. Therefore, it would have been difficult to control for the potential impact of surgery on WMT performance and also on neuropsychological measures when examining differences among WMT performance groups. It remains unknown how including post-surgical patients might have impacted findings, including whether there would have been differences in WMT performance between pre- and post-surgical patients. Future research should compare the WMT performance of pre- and post-surgical patients with epilepsy, as well as examine possible differences in neuropsychological functioning between patient groups in light of WMT performance, in order to further expand the epilepsy PVT research base.

Moving on, another potential limitation of the current study was that seizure data (e.g., seizure duration, age of onset, seizure type, seizure frequency, date of last seizure) were not included in analyses. These data were not included in regression models because

the main focus of the study was to explore the relationship between WMT performance and neuropsychological test scores in patients with epilepsy, and including seizure data would have changed the nature and broadened the scope of the study. However, incorporating seizure data might have impacted current findings. For example, research has found that patients who experienced seizure activity within 24 hours of, or during, neuropsychological testing may perform below their typical level of cognitive functioning on neuropsychological measures (Rennick, Perez-Boria, & Rodin, 1969; Aldenkamp & Arends, 2004a; Aldenkamp & Arends, 2004b). Therefore, including seizure data such as date of last seizure and seizure frequency might have explained unaccounted for variance in cognitive domain and DTBM scores. Further research investigating the relationship between PVT and neuropsychological scores within the epilepsy population should include seizure data as predictor or control variables in data analysis models to clarify and expand upon current findings.

Another limitation was the study's small sample size ($N = 81$). Although power analyses indicated that the sample was large enough to detect large effects in all but one analysis, results may have differed and been more generalizable if the sample had been larger. For example, if a substantially greater amount of patients with epilepsy were included, a higher proportion of the sample may have demonstrated more significant levels of cognitive impairment. As such, more practically significant differences on neuropsychological measures among WMT performance groups might have emerged, or the GMIP might have been found to be, on average, associated with severe as opposed to borderline memory impairment. Additional studies are therefore encouraged to replicate

the current design with larger sample sizes in hopes of further validating and building upon current findings.

Lack of generalizability is the final limitation of the present study. As described in the Results chapter, data largely violated the assumption of normality. Transformations made to the data were not helpful, so data remained untransformed and, therefore, non-normal. Although ANOVA and multiple regression are considered fairly robust tests, it is possible that the non-normal data had a negative impact on the ability of such tests to estimate reliable statistics and produce models representative of the sample data. Various assumptions of regression were also violated for some of the multiple regression models, indicating that results must be interpreted with caution and cannot be generalized to other samples of patients with epilepsy or to other patient populations. Replications and expansions of the current study with largely normal datasets that do not violate many of the assumptions of the statistics being employed are encouraged to provide further validation of current findings.

Conclusion

Results of the present study indicated that previously reported base rates of suboptimal performance in the epilepsy population might vary considerably (e.g., from 4-28%) because such rates likely included patients with significant memory impairment in their suboptimal PV groups. Seeking to clarify that variance in base rate, results of the current study revealed that 21% of the sample scored below WMT failure cutoff because of significant memory impairment (as defined by LM scores in the borderline range). This finding suggested that similar patients might have been misclassified as performing suboptimally in previous studies that did not employ WMT GMIP analysis.

Interestingly, current results also revealed that a substantial percentage of patients (36%) performed suboptimally on the WMT. This finding likely indicates that a significant number of patients simply underperformed, for reasons unknown, on the WMT and various neuropsychological measures administered during testing. This finding may also reflect how patients were sorted into performance groups. Namely, patients with IR, DR, and CNS scores in the caution range were sorted into the suboptimal group, instead of the optimal group as has typically been done in prior studies, along with patients who scored below failure cutoffs and did not have a GMIP (i.e., "clear fail" performance). As discussed earlier in the discussion section, sorting patients with caution range scores into the suboptimal group likely increased the base rate of suboptimal performance attained in the current study.

Regarding the relationship between PVT scores and neuropsychological test scores, current results were largely consistent with extant research and indicated that WMT performance accounted for variance in overall neuropsychological performance, though not in as high a rate as previously reported values. In the current study, WMT performance accounted for 29% of the variance in overall neuropsychological performance, and from 17 to 40% of the variance in cognitive domain scores. Of note, PV did not impact all cognitive domains equally, a finding consistent with existing research that suggests that certain domains (e.g., learning and memory) may be more sensitive to the impact of suboptimal performance than others. Additionally, WMT performance groups scored significantly different across most, but not all, neuropsychological measures, with patients in the suboptimal and GMIP groups obtaining significantly lower scores on most measures than patients in the optimal group.

Notably, the sample was not as cognitively impaired as would be expected given a majority (70%) pre-surgical sample. The lack of a significantly cognitively impaired sample may have accounted for the lack of practically significant difference on some variables among performance groups.

Finally, preliminary support was found for the validity of the GMIP in identifying WMT scores below failure cutoff due to significant memory impairment. More specifically, patients in the GMIP performance group averaged LM scores in the borderline range – indicating the presence of significant impairment – compared to non-GMIP group patients who averaged LM scores in the average range. Although a slight threat was identified regarding the validity of the GMIP in this sample, this threat should be interpreted with caution as the finding may have reflected the regression model's inaccurate representation of the sample data.

Overall, results of this study encourage the use of the WMT and GMIP analysis in patients with epilepsy. It should be noted, though, that similar to extant PV research with various patient populations, the majority of variance in cognitive domains and overall neuropsychological profile remained unaccounted for. Other possible explanatory variables (e.g., seizure and medical data, AEDs, neuropathology, psychopathology, disability status) were not included in current regression models because they were considered to fall outside the focus area of this study. Thus, the unexplained variance in neuropsychological performance should be investigated in future studies that include a wide range of possible explanatory variables. Further, neuropsychologists should be aware that WMT scores do not likely explain the majority of variance in cognitive domain scores or overall neuropsychological performance in patients with epilepsy.

Therefore, neuropsychologists should continue to consider the impact of clinical history and medical status, current level of functioning (e.g., ability to live independently, perform basic activities of daily living, etc.), psychological status, behavioral observations, and demographic factors, along with PVT scores, on neuropsychological performance in patients with epilepsy.

References

Aron, A., Aron, E. N., & Coups, E. J. (2006). *Statistics for psychology* (4th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

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

Aldenkamp, A., & Arends, J. (2004a). Effects of epileptiform EEG discharges on cognitive function: Is the concept of "transient cognitive impairment" still valid? *Epilepsy and Behavior, 5*, S25-S34.

Aldenkamp, A., & Arends, J. (2004b). The relative influence of epileptic EEG discharges, short nonconvulsive seizures, and type of epilepsy on cognitive function. *Epilepsia, 45*, 54-63.

American Academy of Clinical Neuropsychology (AACN). (2007). AACN practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist, 21*, 209-231.

Ardila, A., Galeano, L. M., & Rosselli, M. (1998). Toward a model of neuropsychological activity. *Neuropsychology Review, 8*, 171-190.

Army Individual Test Battery. (1944). Manual of Directions and Scoring. Washington, DC: War Department, Adjutant General's Office.

Axelrod, B. N., Goldman, B. S., & Woodard, J. L. (1992). Interrater reliability in scoring the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist, 6*, 143-155.

Axelrod, B. N., Ricker, J. H., & Cherry, S. A. (1994). Concurrent validity of the MAE visual naming test. *Archives of Clinical Neuropsychology, 9*, 317-321.

Belsey, D. A., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, NY: Wiley.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test—Revised.* Odessa, FL: Psychological Assessment Resources, Inc.

Benedict, R. H. B., & Brandt, J. (2007). *Hopkins Verbal Learning Test-Revised/Brief Visuospatial Memory Test-Revised professional manual supplement.* Lutz, FL: Psychological Assessment Resources, Inc.

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist, 12*, 43-55.

Benedict, R. H. B., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8*, 145-153.

Benton, A. L., Hamsher, K. deS, & Sivan, A. B. (1994a). *Multilingual Aphasia Examination* (3rd ed.). Iowa City, IA: AJA.

Benton, A. L., Hannay, H. J., & Varney, N. R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology, 25*, 907-910.

Benton, A. L., Sivan, A. B., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1994b). *Contributions to neuropsychological assessment. A clinical manual* (2nd ed.). New York, NY: Oxford University Press.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Cortex, 4*, 344-358.

Benton, A. L., Varney, N. R., & Hamsher, K. deS. (1978). Visuospatial judgment: A clinical test. *Archives of Neurology, 35*, 364-367.

Berg, E. A. (1948). A simple objective test for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15-22.

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

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*, 19-45.

Bickart, W. T., Meyer, R. G., & Connell, D. K. (1991). The symptom validity technique as a measure of feigned short-term memory deficit. *American Journal of Forensic Psychology, 9*, 3-11.

Binder, L. M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4*, 25-36.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15*, 170-182.

Binder, L. M. (2002). The Portland Digit Recognition Test: A review of validation data and clinical use. *Journal of Forensic Neuropsychology, 2*, 27-41.

Binder, L. M., & Kelly, M. P. (1996). Portland Digit Recognition Test performance by brain dysfunction patients without financial incentives. *Assessment, 3*, 403-409.

Binder, L. M., Kindermann, S. S., Heaton, R. K., & Salinsky, M. C. (1998). Neuropsychological impairment in patients with nonepileptic seizures. *Archives of Clinical Neuropsychology, 13*, 513.

Binder, L. M., Salinsky, M. C., & Smith, S. P. (1994). Psychological correlates of psychogenic seizures. *Journal of Clinical and Experimental Neuropsychology, 16*, 524.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*, 141-147.

Boone, K. B. (Ed.) (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective.* New York, NY: Guilford Press.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23*, 729-741.

Boone, K. B., & Lu, P. H. (2007). Non-forced-choice effort measures. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 27-43). New York, NY: Oxford University Press.

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., … Warner-Chacon, K. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology, 17*, 625-642.

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

Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., … Berman, N. G. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The *b* Test. *Archives of Clinical Neuropsychology, 15*, 227-241.

Bortz, J. J. (2003). Neuropsychiatric and memory issues in epilepsy. *Mayo Clinic Proceedings, 78*, 781-787.

Bowden, S. C., Shores, E. A., & Mathias, J. L. (2006). Does effort suppress cognition after brain injury? A re-examination of the evidence for the Word Memory Test. *The Clinical Neuropsychologist, 20*, 858–872.

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury.

Brady, J. P., & Lind, D. L. (1961). Experimental analysis of hysterical blindness: Operant conditioning techniques. *Archives of General Psychiatry, 4*, 331-339.

Brandt, J., & Benedict, R. H. B. (2001). Hopkins Verbal Learning Test-Revised. Lutz, FL: Psychological Assessment Resources.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106-148.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, J., & Cohen. P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Conder, R., Allen, L., & Cox, (1992). *Manual for the computerized assessment of response bias.* Durham, NC: CogShell.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20*, 191-198.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman & Hall.

Cragar, D. E., Berry, D. T. R., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2002). A review of diagnostic techniques in the differential diagnosis of epilepsy and non-epileptic seizures. *Neuropsychology Review, 12*, 31-64.

Cragar, D. E., Berry, D. T. R., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2006). Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *The Clinical Neuropsychologist, 20*, 552-566.

Daigneault, S., Braun, C., Gilbert, B., & Proulx, R. (1988). Canonical and factorial structures of a battery of "frontal" neuropsychological measures. *Journal of Clinical and Experimental Neuropsychology, 10,* 58.

De Renzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect disturbances in aphasics. *Brain, 85,* 665-678.

des Rosiers, G., & Kavanagh, D. (1987). Cognitive assessment in closed head injury: Stability, validity and parallel forms for two neuropsychological measures of recovery. *International Journal of Clinical Neuropsychology, 9,* 162-173.

Devinsky, O. (2004, February 1). Absence seizures. Retrieved from http://www.epilepsy.com/epilepsy/seizure_absence

Dodrill, C. B. (1978). The hand dynanmometer as a neuropsychological measure. *Journal of Consulting and Clinical Psychology, 46,* 1432-1435.

Dodrill, C. B. (2008). Do patients with psychogenic nonepileptic seizures produce trustworthy findings on neuropsychological tests? *Epilepsia, 49,* 691-695.

Drane, D. L., Williamson, D. J., Stroup, E S., Holmes, M. D., Jung, M., Koerner, E., … Miller, J. W. (2006). Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia, 47,* 1879-1886.

Dricker, J., Butters, N., Berman, G., Samuels, I., & Carey, S. (1978). The recognition and encoding of faces by alcoholic Korsakoff and right hemisphere patients. *Neuropsychologia, 16,* 683-695.

Egelko, S., Gordon, W. A., Hibbard, M. R., Diller, L., Lieberman, A., Holliday, R., … Orazem, J. (1988). Relationship among CT scans, neurologic exam, and neuropsychological test performance in right-brain-damaged stroke patients. *Journal of Clinical and Experimental Neuropsychology, 10,* 539-564.

Ehrenstein, W. H., Heister, G., & Cohen, R. (1982). Trail Making test and visual search. *Archiv für Psychiatrie and Nervenkrankheiten, 231,* 333-338.

Epilepsy Foundation of America (EFA). (2010). What is epilepsy. Retrieved from http://www.epilepsyfoundation.org/aboutepilepsy/whatisepilepsy/index.cfm

Ettinger, A. B., & Kanner, A. M. (2007). *Psychiatric issues in epilepsy: A practical guide to diagnosis and treatment* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Fals-Stewart, W. (1991). An interrater reliability study of the Trail Making Test (Part A and B). Unpublished manuscript.

Faust, D., Hart, K., & Guilmette, T. J. (1998.) Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56,* 578-582.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19,* 508-515.

Field, A. (2008). Analysis of covariance (ANCOVA). Retrieved from http://statisticshell.com/docs/ancova.pdf

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publications Ltd.

Fossum, B., Holmberg, H., & Reinvang, I. (1992). Spatial and symbolic factors in performance on the Trail Making Test. *Neuropsychology, 6,* 71-75.

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

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indictor Profile. *Law and Human Behavior, 24,* 59-82.

Gervais, R. O., Green, P., Allen, L. M., & Iverson, G. L. (2001a). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology, 2,* 1-20.

Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001b). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology, 28,* 1892-1899.

Gierok, S. D., Dickson, A. L., & Cole, J. A. (2005). Performances of forensic and non-forensic adult psychiatric inpatients on the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 20,* 755-760.

Goldberg, T. O., & Miller, H. R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42,* 792-795.

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word Memory Test performance in amnestic patients with hippocampal damage. *Neuropsychology, 23,* 529-534.

Gorissen, M., Sanz de la Torre, J. C., & Schmand, B. (2003) In Green, P. (2005). *Green's Word Memory Test for Microsoft Windows. User's Manual Revised.* Edmonton, Canada: Green's Publishing.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 34,* 404-411.

Green, P. (2001). Why clinicians often disagree about the validity of test results. *NeuroRehabilitation, 16,* 231-236.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows. User's manual*. Edmonton, Canada: Green's Publishing.

Green, P. (2005). *Green's Word Memory Test for Microsoft Windows. User's manual revised*. Edmonton, Canada: Green's Publishing.

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America, 18,* 43-68.

Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT) user's manual for Microsoft Windows.* Edmonton, Canada: Green's Publishing Inc.

Green, P., & Allen, L. M. (1999). Performance of neurological patients on the Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB). *Supplement to the Word Memory Test and CARB '97 manuals.* Durham, NC: CogniSyst.

Green, P., Allen, L. M. & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US version 1.1*. Durham, NC: Cognisyst.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23,* 741-750.

Green, P., & Iverson, G. L. (2001). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15*, 492-497.

Green, P., Iverson, G. L., & Allen, L. M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*, 813-819.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2,* 97-124.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe memory impairment. *Applied Neuropsychology, 18,* 86-94.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15,* 1045-1060.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia methods with a large clinical sample. *Psychological Assessment, 6,* 218-224.

Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., & Ciota, M. (2006a). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *NeuroToxicology, 27,* 940-950.

Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., & Ciota, M. (2006b). Classification accuracy of the Test of Memory Malingering in persons reporting exposure to environmental and industrial toxins: Results of a known-group analysis. *Archives of Clinical Neuropsychology, 21,* 439-448.

Grosz, H. J., & Zimmerman, J. (1965). Experimental analysis of hysterical blindness: A follow-up report and new experimental data. *Archives of General Psychiatry, 15,* 255-260.

Grote, C. L., & Hook, J. N. (2007). Forced-choice recognition tests of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 44-79). New York, NY: Oxford University Press.

Grote, C., Kooker, E., Garron, D., Nyenhuis, D., Smith, C., & Mattingly, M. (2000). Performance of compensation-seeking and non-compensation-seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology, 22,* 709-719.

Grote, C. L., Smith, C. A., & Ruth, A. (2001). Neuropsychological evaluation of the patient with seizures. In A. B. Ettinger, & A. M. Kanner (Eds.). *Psychiatric issues in epilepsy* (pp. 31-43). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Guilmette, T., Hart, K., & Guiliano, A. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. *The Clinical Neuropsychologist, 7,* 59-69.

Guilmette, T. J., Hart, K. J., Guiliano, A. J., & Leninger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock forced-choice procedure. *The Clinical Neuropsychologist, 8,* 283-294.

Haaland, K. Y., & Delaney, H. D. (1981). Motor deficits after left or right hemisphere damage due to stroke or tumor. *Neuropsychologia, 19,* 17-27.

Hamsher, K. deS., Levin, H. S., & Benton, A. L. (1979). Facial recognition with focal brain lesions. *Archives of Neurology, 36*, 837-839.

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Hamer, R. M. (1988). Productive naming and memory in depression and Alzheimer's type dementia. *Archives of Clinical Neuropsychology, 3*, 313-322.

Hawkins, K. A., Sledge, W. H., Orleans, J. F., Quinland, D. M., Rakfeldt, J., & Hoffman, R. E. (1993). Normative implications of the relationship between reading vocabulary and Boston Naming Test performance. *Archives of Clinical Neuropsychology, 8*, 525-537.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual revised and expanded*. Odessa, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an extended Halstead-Reitan battery*. Odessa, FL: Psychological Assessment Resources, Inc.

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

Heilbronner, R. L., Henry, G. K., Buck, P., Adams, R. L., & Fogle, T. (1991). Lateralized brain damage and performance on Trail Making A and B, Digit Span Forward and Backward, and TPT memory and location. *Archives of Clinical Neuropsychology, 6*, 251-258.

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

Hennekens, C. H., & Buring, J. E. (1987). *Epidemiology in medicine*. Boston: Little, Brown and Company.

Henry, M., Merten, T., Wolf, S. A., & Harth, A. (2009). Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients. *Journal of Clinical and Experimental Neuropsychology, 32*, 19-27.

Hermann, B. P., Seidenberg, M., Dow, C., Jones, J., Rutecki, P., Bhattacharya, A., & Bell, B. (2006). Cognitive prognosis in chronic temporal epilepsy. *Annals of Neurology*, 60, 80-87.

Hermann, B. P., Seidenberg, M., Haltiner, A., & Wyler, A. R. (1992). Adequacy of language function and verbal memory performance in unilateral temporal lobe epilepsy. *Cortex, 28*, 423-433.

Hermann, B. P., Seidenberg, Lee, E. J., Chan, F., & Rutecki, P. (2007). Cognitive phenotypes in temporal lobe epilepsy. *Journal of the International Neuropsychological Society, 13*, 12-20.

Hermann, B. P., & Wyler, A. R. (1988). Effects of anterior temporal lobectomy on language function. *Annals of Neurology, 23*, 585-588.

Hessen, E., Lossius, M. I., & Gjerstad, L. (2008). Behavioural adjustment in seizure-free epilepsy patients on monotherapy. *Seizure, 17*, 422-430.

Hill, S. K., Ryan, L. M., Kennedy, C. H., & Malamut, B. L. (2003). The relationship between measures of declarative memory and the Test of Memory Malingering in patients with and without temporal lobe dysfunction. *Journal of Forensic Neuropsychology, 3*, 1-18.

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

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*, 967-974.

Hoaglin, D., & Welsch, R. (1978). The hat matrix in regression and ANOVA. *American Statistician, 32*, 17-22.

Hoptman, M. J., & Davidson, R. J. (1993). Benton's Facial Recognition Task: A psychometric evaluation. Paper presented at the meeting of the International Neuropsychological Society, Galveston, TX.

Hoskins, L. L., Binder, L .M., Chaytor, N. S., Williamson, D. J., & Drane, D. L. (2010). Comparison of oral and computerized versions of the Word Memory Test. *Archives of Clinical Neuropsychology, 25*, 591-600.

Howe, L. L. S., Anderson, A. M., Kaufman, D. A. S., Sachs, B. C., & Loring, D. W. (2007). Characterization of the Medical Symptom Validity Test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology, 22*, 753-761.

Howe, L. L. S., & Loring, D. W. (2009). Classification accuracy and predictive ability of the Medical Symptom Validity Test's dementia profile and general memory impairment profile. *The Clinical Neuropsychologist, 23,* 329-342.

Huff, F. J., Collins, C., Corkin, S., & Rosen, T. J. (1986). Equivalent forms of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 8,* 556-562.

Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology, 17,* 1-23.

Inman, T. H., Vickery, C. D., Berry, D. T. R., Lamb, D. G., Edwards, C. L., & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The Letter Memory Test. *Psychological Assessment, 10,* 128-139.

Iverson, G., Green, P., & Gervais, R. (1999). Using the Word Memory Test to detect biased responding in head injury litigation. *The Journal of Cognitive Rehabilitation,* March/April, 2-6.

Jokeit, H., & Ebner, A. (1999). Long term effects of refractory temporal lobe epilepsy on cognitive abilities: A cross sectional study. *Journal of Neurology, Neurosurgery, and Psychiatry, 67,* 44-50.

Jones-Gotman, M., Smith, M. L., Risse, G. L., Westerveld, M., Swanson, S. J., Giovagnoli, A. R., … Piazzini, A. (2010). The contribution of neuropsychology to diagnostic assessment in epilepsy. *Epilepsy and Behavior, 18,* 3-12.

Kaplan, E., Goodglass, H., & Weintraub, S. (1978). *The Boston Naming Test* (Experimental ed.). Boston, MA: Kaplan & Goodglass.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Keary, T. A., Frazier, T. W., Belzile, C. J., Chapin, J. S., Naugle, R. I., Najm, I. M., & Busch, R. M. (2013). Working memory and intelligence are associated with Victoria Symptom Validity Test hard item performance in patients with intractable epilepsy. *Journal of the International Neuropsychological Society, 19,* 314-323.

Keary, T. A., Frazier, T. W., Busch, R. M., Kubu, C. S., & Iampietro, M. (2007). Multivariate neuropsychological prediction of seizure lateralization in temporal epilepsy surgical cases. *Epilepsia, 48,* 1438-1446.

Kelland, D. Z., Lewis, R., & Gurevitch, D. (1992). Evaluation of the Repeatable Cognitive-Perceptual-Motor Battery: Reliability, validity and sensitivity to Diazepam. *Journal of Clinical and Experimental Neuropsychology, 14,* 65.

Kløve, H. (1963). Clinical neuropsychology. In F. M. Forster (Ed.), *The medical clinics of North America*. New York, NY: Saunders.

Lafayette Instrument. (1989). *Instruction/Owner's Manual for the Grooved Pegboard Test.* Lafayette, IN: Author.

Lally, S. (2003). What tests are acceptable for use in forensic evaluation? A survey of experts. *Professional Psychology: Research and Practice, 34,* 491-498.

Larrabee, G. J. (1992). On modifying recognition memory tests for detection of malingering. *Neuropsychology, 6,* 23-27.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17,* 410-425.

Larrabee, G. J. (2005). Assessment of malingering. In G. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (pp. 115-158). New York, NY: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 625-631.

Larrabee, G. J., & Curtis, G. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Neuropsychology, 17,* 536-547.

Lee, G. P., & Hamsher, K. (1988). Neuropsychological findings in toxicometabolic confusional states. *Journal of Clinical and Experimental Neuropsychology, 10,* 769-778.

Lee, G., Loring, D., & Martin, R. (1992). Rey's 15-Item visual memory test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment, 4,* 43-46.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1995). Forensic neuropsychological test usage: An empirical study. *Archives of Clinical Neuropsychology, 11,* 45-51.

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. R. (1992). Factor structure of the Wechsler Memory Scale—Revised within a comprehensive neuropsychological battery. *Neuropsychology, 6,* 239-249.

Levick, W. (1982). *The Judgment of Line Orientation Test: A clinical evaluation.* Master's psychology thesis, University of Newcastle, Australia.

Levin, H., Grossman, R. G., & Kelly, P. J. (1976). Aphasic disorders in patients with closed-head injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 39*, 1062-1070.

Levin, H., Grossman, R. G., Sarwar, M., & Meyers, C. A. (1981). Linguistic recovery after closed head injury. *Brain and Language, 12*, 360-374.

Levin, H. S., Hamsher, K. deS., & Benton, A. L. (1975). A short form of the Test of Facial Recognition for clinical use. *Journal of Psychology, 91*, 223-228.

Levin, B. E., Llabre, M. M., & Reisman, S. (1991). Visuospatial impairment in Parkinson's disease. *Neurology, 41*, 365-369.

Lewis, R. F., & Rennick, P. M. (1979). *Manual for the Repeatable Cognitive-Perceptual-Motor Battery*. Clinton Township, MI: Ronald F. Lewis.

Lezak, M. D. (1982). The test-retest stability and reliability of some tests commonly used in neuropsychological assessment. Paper presented at the 5th European Conference of the International Neuropsychological Society, Deauville, France.

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York, NY: Oxford University Press.

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

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley & Sons, Inc.

Locke, D. E. C, Berry, T. R., Fakhoury, T. A., & Schmitt, F. A. (2006). Relationship of indicators of neuropathology, psychopathology, and effort to neuropsychological results in patients with epilepsy or psychogenic non-epileptic seizures. *Journal of Clinical and Experimental Neuropsychology, 28*, 325-340.

Loring, D. W., Lee, G. P., & Meador, K. J. (2005). Victoria Symptom Validity test performance in non-litigating epilepsy surgery candidates. *Journal of Clinical and Experimental Neuropsychology, 27*, 610-617.

Lynch, W. J. (2004). Determination of effort level, exaggeration, and malingering in neurocognitive assessment. *Journal of Head Trauma Rehabilitation, 19*, 277-283.

Marcangelo, M. J., & Ovsiew, F. (2007). Psychiatric aspects of epilepsy. *Psychiatry Clinics of North America, 30*, 781-802.

Martin, R., Bolter, J., Todd, M., Gouvier, W., & Niccolls, R. (1993). Effects of sophistication and motivation on the detection of malingered memory

performance using a computerized forced-choice task. *Journal of Clinical and Experimental Neuropsychology, 15*, 867-880.

Martin, R., Hayes, J., & Gouvier, W. (1996). Differential vulnerability between post-concussion self-report and objective malingering tests in identifying simulated mild head-injury. *Journal of Clinical and Experimental Neuropsychology, 18*, 265-275.

Massman, P. J., & Doody, R. S. (1996). Hemispheric asymmetry in Alzheimer's disease is apparent in motor functioning. *Journal of Clinical and Experimental Neuropsychology, 18,* 110-121.

Matthews, C. G., & Haaland, K. Y. (1979). The effect of symptom duration on cognitive and motor performance in Parkinsonism. *Neurology, 29*, 951-956.

Matthews, C. G., & Kløve, H. (1964). *Instruction manual for the Adult Neuropsychology Test Battery*. Madison, WI: University of Wisconsin Medical School.

Mattis, S. (1988). *Dementia Rating Scale*. Odessa, FL: Psychological Assessment Resources.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 11*, 386-401.

Menard, S. (1995). *Applied logistic regression analysis.* Sage University paper series on quantitative applications in the social sciences, 07-106. Thousand Oaks, CA: Sage.

Miller, L. S., & Rohling, M. L. (2001). A statistical interpretative method for neuropsychological test data. *Neuropsychology Review, 11*, 143-169.

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*, 1094-1102.

Mittenberg, W., Theroux-Fichera, S., Zielinski, R. E., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology: Research and Practice, 26*, 491-498.

Monsch, A. U., Bondi, M. W., Butters, N., Paulsen, J. S., Salmon, D. P., Brugger, P., & Swenson, M. R. (1994). A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. *Neuropsychologia, 8*, 25-30.

Monsch, A. U., Bondi, M. W., Butters, N., & Salmon, D. P. (1992). Comparison of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Archives of Neurology, 49*, 1253-1258.

Moore, P. M., & Baker, G. A. (2002). The neuropsychological and emotional consequences of living with intractable temporal lobe epilepsy: Implications for clinical management. *Seizure, 11*, 224-230.

Morrison, M. W., Gregory, R. J., & Paul, J. J. (1979). Reliability of the Finger Tapping Test and a note on sex differences. *Perceptual and Motor Skills, 48*, 139-142.

Myers, R. (1990). *Classical and modern regression with applications* (2nd ed.). Boston, MA: Duxbury.

Nitch, S. R., & Glassmire, D. M. (2007). Non-forced-choice measures to detect noncredible cognitive performance. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 78-102). New York, NY: The Guilford Press.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw–Hill.

Pallant, J. (2005). *SPSS survival manual.* Crows Nest, Australia: Allen & Unwin.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57*, 367-372.

Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-192). New York, NY: Guilford.

Pankratz, L., Fausti, S. A., & Peed, S. (1975). Case study: A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43*, 421-422.

Partington, J. E., & Leiter, R. G. (1949). Partington's Pathway Test. *The Psychological Service Center Bulletin, 1*, 9-20.

Philpott, L. M. (1992). *The effects of severity of cognitive impairment and age on two malingering tests: An investigation of the Rey Memory Test and Rey Dot Counting Test in Alzheimer's patients and normal middle aged/older adults.* Unpublished doctoral dissertation, California School of Professional Psychology, Los Angeles, CA.

Pontius, A. A., & Yudowitz, B. S. (1980). Frontal lobe system dysfunction in some criminal actions in a Narratives Test. *Journal of Nervous and Mental Disease, 168*, 111-117.

Powell, D. H., Kaplan, E. F., Whitla, D., Weintraub, S., Catlin, R., & Funkenstein, H. H. (1993). *MicroCog: Assessment of Cognitive Functioning.* San Antonio, TX: The Psychological Corporation.

Prigatano, G. P., Smason, I., Lamb, D. G., & Bortz, J. J. (1997). Suspected malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology, 12*, 609-619.

Provins, K. A., & Cunliffe, P. (1972). The reliability of some motor performance tests of handedness. *Neuropsychologia, 10*, 199-206.

Psychological Assessment Resources Staff. (1992). *Finger Tapper user's guide*. Lutz, FL: Psychological Assessment Resources, Inc.

Rabin, L., Barr, W., & Burton, L. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33-65.

Raven, J. C. (1976). *Standard Progressive Matrices.* Oxford, England: Oxford Psychological Press.

Reddon, J. R., Stefanyk, W. O., Gill, D. M., & Renney, C. (1985). Hand dynamometer: Effects of trails and sessions. *Perceptual and Motor Skills, 61*, 1195-1198.

Reitan, R. M. (1969). *Manual for the Administration of Neuropsychological Test Batteries for Adults and Children.* Indianapolis, IN: Neuropsychology Laboratory, Indiana University Medical Center.

Reitan, R. M., & Davison, L. A. (1974). *Clinical Neuropsychology: Current Status and Applications*. Washington, DC: V. H. Winston.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson, AR: Neuropsychology Press.

Rennick, M., Perez-Boria, C., & Rodin, E. A. (1969). Transient mental deficits associated with recurrent prolonged epileptic clouded state. *Epilepsia, 10*, 397-405.

Rey, A. (1941). L'examen psychologie dans las cas d'encephalopathie traumatique. *Archives de Psychologie, 28*, 286-340.

Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaires de France.

Reynolds, C. R. (1998). *Detection of malingering during head injury litigation.* New York, NY: Plenum Press.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury, 19*, 539-543.

Rohling, M. L., Allen, L. M., & Green, P. (2002). Who is exaggerating cognitive impairment and who is not? *CNS Spectrums, 7*, 387-395.

Rohling, M. L., & Demakis, G. J. (2010). Bowden, Shores, & Mathias (2006): Failure to replicate or just failure to notice. Does effort still account for more variance in neuropsychological test scores than TBI severity? *The Clinical Neuropsychologist, 24*, 119-136.

Rosen, W. G. (1980). Verbal fluency in aging and dementia. *Journal of Clinical Neuropsychology, 2*, 135-146.

Sadock, B. J., & Sadock, V. A. (2007). *Kaplan & Sadock's synopsis of psychiatry: Behavioral sciences/clinical psychiatry* (10th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Sattler, J. M. & Ryan, J. J. (2009). *Assessment with the WAIS-IV.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Sawrie, S. M., Chelune, G. J., Naugle, R. I., & Luders. H. O. (1996). Empirical methods for assessing meaningful change following epilepsy surgery. *Journal of the International Neuropsychological Association, 2*, 556-564.

Schachter, S. C. (2009). Seizure disorders. *Medical Clinics of North America, 93*, 343-351.

Schretlen, D., Brandt, J., Krafft, L., & van Gorp, W. (1991). Some caveats in using the Rey 15-Item Memory Test to detect malingered amnesia. *Psychological Assessment, 3*, 667-672.

Shapiro, A. M., Benedict, R. H. B., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test-Revised. *The Clinical Neuropsychologist, 13*, 348-358.

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

Shute, G. E., & Huertas, V. (1990). Developmental variability in frontal lobe function. *Developmental Neuropsychology, 6*, 1-11.

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology, 24*, 721-728.

Sirven, J. I. (2002). Demystifying seizures and epilepsy: Introduction to the symposium on seizures. *Mayo Clinic Proceedings, 77*, 977-979.

Slick, D. J., Hopp, G., Strauss, E. (1995). *The Victoria Symptom Validity Test*. Odessa, FL: Psychological Assessment Resources.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. (1996). Victoria Symptom Validity Test: Efficiency for detection of feigned memory impairment and relationship to neurological test and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18*, 911-922.

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test*: *Professional Manual.* 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.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19*, 465-473.

Slick, D. J., Tan, J. E., Strauss, E., Mateer, C. A., Harnadek, M., & Sherman, E. M. S. (2003). Victoria Symptom Validity Test scores of patients with profound memory impairment: Nonlitigant case studies. *The Clinical Neuropsychologist, 17*, 390-394.

Spiegel, E. (2006). *The Rey 15-Item Memorization Test: Performance in under-represented populations in the absence of obvious motivation to feign neurocognitive symptoms*. Unpublished doctoral dissertation, Fuller Theological Seminary, Pasadena, CA.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests* (2nd ed.). New York, NY: Oxford University Press.

Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research, 157*, 191-200.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Stevens, J. (2003). Partial and semipartial correlations. Retrieved from http://pages.uoregon.edu/stevensj/MRA/partial.pdf

Strauss, E., & Wada, J. (1988). Hand preference and proficiency and cerebral speech dominance determined by the carotid amytal test. *Journal of Clinical and Experimental Neuropsychology, 10*, 169-174.

Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 225-286). Lisse, Netherlands: Swets and Zeitlinger.

Sweet, J. (2011). *Professional practice guidelines for the use and interpretation of SVTs: The time has come* [PowerPoint slides]. Retrieved from http://wordmemorytest.com/DISCUSSIONS/EFFORT_EFFECTS/Sweet_SVT_lecture1.pdf

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16,* 495-505.

Tavakoli, M., Barekatain, M., Doust, H. T. N., Molavi, H., Nouri, R. K., Moradi, A., … Zare, M. (2011). Cognitive impairments in patients with intractable temporal lobe epilepsy. *Journal of Research in Medical Sciences, 16*, 1466-1472.

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

The Psychological Corporation. (2002). *WAIS – III, WMS – III technical manual.* USA: Author.

Theodor, L H., & Mandelcorn, M. S. (1973). Hysterical blindness: A case report and study using a modern psychophysical technique. *Journal of Abnormal Psychology, 82*, 552-553.

Thompson, L. L., & Heaton, R. K. (1989). Comparison of different versions of the Boston Naming Test. *The Clinical Neuropsychologist, 3*, 184-192.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM).* North Tonawanda, NY: Multi-Health Systems, Inc.

Torta, R., & Keller, R. (1999). Behavioral, psychotic, and anxiety disorders in epilepsy: Etiology, clinical features, and therapeutic implications. *Epilepsia*, *40*(Suppl. 10), S2-S20.

Trahan, D. E. (1991). Judgment of line orientation in patients with unilateral vascular lesions or severe head trauma. *Journal of Clinical and Experimental Neuropsychology, 13*, 23-24.

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.

Warrington, E. K., & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. *Cortex, 3,* 317-326.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children – Revised.* New York, NY: The Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale–Revised manual.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale–Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale–Third Edition.* San Antonio, TX: The Psychological Corporation.

Wilkinson, G. S. (1993). *Wide Range Achievement Test–3*. Lutz, FL: Psychological Assessment Resources.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test–4 professional manual*. Lutz, FL: Psychological Assessment Resources.

Williamson, D. J., Drane, D. L., Stroup, E. S., Miller, J. W., & Holmes, M. D. (2003). *Most patients with psychogenic seizures do not exert valid effort on neurocognitive testing.* Poster presented at the Annual Meeting of the American Epilepsy Society, Boston, MA.