2002

Qualitative scoring of the Rey 15-Item Memory Test in a forensic population

James A. Martin
Louisiana State University and Agricultural and Mechanical College, jamarti@lsu.edu

QUALITATIVE SCORING OF THE REY 15-ITEM MEMORY TEST IN A FORENSIC POPULATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
James A. Martin
B.S., University of Georgia, 1988
M.S., Augusta State University, 1991
M.A., Louisiana State University, 1999
August, 2002

Acknowledgments

The author would like to acknowledge the Eastern Louisiana Mental Health System for the use of their facilities, records, and patient volunteers. The contributions of both psychology and clerical staff were especially beneficial in making this project possible.

Table of Contents

List of Tables

Abstract

Several studies have examined the ability of the Rey 15-Item Memory Test (MFIT) to identify malingering of memory problems among a variety of psychiatric and neurologically impaired populations. The consensus has been that the quantitative scoring method is overly sensitive to genuine memory impairment and lacks sensitivity to simulated amnesia. However, a reexamination of these studies and available data indicates the MFIT is both valid and effective at identifying actual malingerers among civil litigants, and a number of these studies were limited through inappropriate inclusion of severely impaired patients and research designs of questionable validity. Also, the performance of a group for whom malingering of memory complaints is a relevant issue (criminal defendants) has been overlooked. The present study expands upon previous investigations by comparing the MFIT performance of a known group of forensic malingerers to a group of non-malingering pretrial criminal defendants and non-malingering post-trial forensic inpatients, and by examining the utility of a qualitative scoring approach hypothesized to enhance the MFIT's detection ability.

Using the quantitative method, a low sensitivity of 47.7% was obtained for malingerers. Minimal improvement was found when qualitative scoring was incorporated (56.8%), although confidence in correct identification was increased with very low total scores (<5) and failure to recall at least 3 of the first 6 items. While the quantitative method yielded high specificity for non-malingering post-trial patients (86.7%), this was not the case for the more clinically relevant non-malingering pretrial patients (56.2%). However, specificity was increased for both non-malingering groups through the addition of qualitative scoring. Although both the quantitative and

combined quantitative and qualitative scoring methods were found to be accurate at identifying criminal forensic malingerers, neither was found to be more accurate than base rate prediction alone. It is concluded that the lack of effectiveness can be attributed to 1) decreased sensitivity to less blatant forms of malingering, and 2) the adverse impact of lower intelligence and psychiatric symptoms affecting the ability to attend and organize cognitive processes on the MFIT recall for actual patients.

Introduction

Malingering is defined as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs (American Psychiatric Association [APA], 1994). It is recognized that a variety of somatic (e.g., chronic pain), psychiatric (e.g., psychosis), and neurocognitive (e.g., amnesia, low intelligence) symptoms are susceptible to being feigned (Main & Spanswick, 1995; Resnick, 1993; Schretlen, Van Gorp, Wilkins, & Bobholz, 1992). Given the potential losses to society for successful malingering, conservatively estimated to be a $5.36 billion annual cost in the U.S.A. (Gouvier, Lees-Haley, & Hammer, in press), it is not surprising that a great amount of effort has gone into the clinical identification of malingerers. One method has been in the use of existing psychological assessment instruments, for example using the Minnesota Multiphasic Personality Inventory (MMPI) to detect feigning of chronic pain (Dush, Simons, Platt, Nation, & Ayres, 1994). Another method has involved the development of new instruments specifically for the detection of malingering, with examples including the M-Test (Beaber, Marston, Michelli, & Mills, 1985) and Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) to detect feigned psychosis, and the Dot Counting Test (DCT; Rey, 1941) to detect feigned neurocognitive impairment.

Neuropsychologists have been increasingly called upon to evaluate for malingering within legal settings (Bernard, 1990). Among civil litigants and disability claimants, the prevalence for malingering of cognitive deficits has been noted to vary widely, with estimates ranging between 15 to 64% (Greiffenstein, Baker, & Gola,

1994; Trueblood & Schmidt, 1993). Within criminal forensic settings, prevalence estimates for the malingering of psychosis and associated cognitive deficits are within a narrower range. Rogers (1986) provided a prevalence estimate of 4.5% for definite malingering and 20% for suspected malingering among criminal defendants being evaluated for insanity. A survey of 320 highly experienced forensic experts (Rogers, Sewell, & Goldstein, 1994) found that 15.7% of forensic evaluatees were classified as malingering, and a more recent survey of 221 forensic experts yielded a similar estimate of 17.4% (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998).

Amnesia is one of the more frequently malingered neuropsychological symptoms among both civil and criminal litigants. However, the simulation of chronic organic amnesia (e.g., severe anterograde amnesia due to traumatic brain injury) has typically been associated with compensation litigation, whereas the simulation of limited forms of amnesia (i.e., amnesia specific to the crime itself) has been more frequently associated with criminal cases (Schacter, 1986a). For example, in a review of legal cases, Schacter (1986b) noted that between 30 to 65% of individuals convicted of homicide claimed limited amnesia for the crime. Moreover, he noted there was no case in which an individual with chronic organic amnesia had come before the court on a serious criminal matter.

This was corroborated by Rubinsky and Brandt (1986), who found no case in which amnesia, in and of itself, was determined to have rendered an individual incompetent to stand trial or to have negated criminal responsibility. Rather, simulated amnesia is typically seen as part of insanity pleas. As Rubinsky and Brandt noted, several state courts have held the position that "amnesia is sometimes an incident of insanity" (p. 30). This is consistent with current research, which has

shown that individuals with certain psychiatric conditions (e.g., schizophrenia, major depression) have associated cognitive deficits involving memory and attention (e.g., Malloy & Duffy, 1994). As such, pretrial criminal defendants have been noted to malinger psychosis and associated cognitive impairment in order to be found incompetent to stand trial, demonstrate diminished criminal responsibility, obtain an insanity verdict, or to mitigate sentencing (Gothard, Rogers, & Sewell, 1995; Resnick, 1997; Rubinsky & Brandt, 1986; Schacter, 1986a; 1986b).

The consensus in the legal community is that memory impairment is easily faked, yet almost impossible to disprove (Wiggins & Brandt, 1988). However, there is a paucity of empirical investigation substantiating this position with respect to criminal defendants. Nonetheless, several investigators offer encouragement that criminal defendants attempting to malinger memory impairment may be identifiable through existing neuropsychological and malingering measures. For example, Rogers and Cruise (1998) postulated these individuals are more likely to be extreme in their presentations, given that only grossly psychotic (and thus cognitively impaired) presentations have a good probability of achieving the proper goal. This is especially relevant for individuals accused of violent crimes (e.g., homicide), who have more at stake in terms of being convicted and thus more likely to overplay their role (Schacter, 1986a). Furthermore, these contentions should be considered within the context of common erroneous perceptions about amnesia. For example, Iverson (1995) found that individuals instructed to simulate amnesia would often fake total amnesia (i.e., personal information, mother's maiden name), and 10% stated they would completely forget presented material. Also, laypersons often do not distinguish between the etiologically distinct amnestic disorders, and thus are more likely to present with

inconsistent symptoms (e.g., severe retrograde amnesia in conjunction with mild head injury) (Gouvier, Prestholdt, & Warner, 1988; Rubinsky & Brandt, 1986; Wiggins & Brandt, 1988).

Despite the finding that criminal defendants are likely to simulate limited as opposed to chronic amnesia, Schacter (1986a) acknowledged these individuals may also simulate memory impairment on current assessments to further substantiate their claims. Still, other confounds must be considered. Individuals already having a psychiatric condition and cognitive deficits may simply exaggerate existing impairment, making the task of separating true deficits far more difficult (Pachana, Boone, & Ganzell, 1998). As pointed out by Hayes, Hilsabeck, and Gouvier (1999), it is not uncommon for attorneys to coach their clients with regard to psychological and neuropsychological assessments. Likewise, malingerers may receive factual information from a variety of sources (i.e., other patients, health care professionals), as well as having prior legal and assessment experiences, which could contribute to more sophisticated presentations.

Thus, while it is possible that criminal defendants who attempt to malinger problems with memory can be identified with existing assessments, the presence of potential confounds makes consistently accurate identification difficult. Rogers, Harrell, and Liff (1993) have identified six systematic strategies that may prove useful for improving the detection of feigning. These include (1) the "performance curve" strategy, which assumes that malingerers will not consider item difficulty in choosing which questions to fail (i.e., failing easy items while passing more difficult ones). A more commonly employed strategy known as (2) "symptom validity testing" involves presenting items with two forced-choice alternatives, with the assumption that

malingerers will fail the items at a below-chance level of performance (i.e., only getting 10% of the items right whereas one would get 50% right by randomly guessing). The (3) "magnitude of error" strategy simply assumes there will be qualitative differences between malingerers and bona fide patients in the types of wrong responses made. Two similar strategies which have received little empirical focus include (4) "atypical (symptom) presentation" and (5) "psychological sequelae" (i.e., the presence of psychological symptoms not typically associated with a given neurological insult).

The final method is referred to by Rogers et al. (1993) as (6) the "floor effect" strategy, which assumes the individual attempting to malinger will fail at a task which even grossly impaired individuals are likely to perform adequately. This has relevance for forensic malingering evaluations, given the presumption that individuals attempting to feign insanity are more likely to overplay their role in an effort to appear more disturbed (Resnick, 1997; Rogers & Cruise, 1998). One such measure, the Rey Fifteen-Item Memory Test (Rey, 1964), has been the focus of empirical investigations for more than a decade.

The Rey Fifteen-Item Memory Test

In 1964, Andre Rey described a screening measure designed to detect memory malingering. Although never formally named by Rey, it has been referred to as the "Rey Memory Test" (Goldberg & Miller, 1986), the "Rey 15-Item Memory Test" (Back et al., 1996), and "Rey's 15-Item Visual Memory Test" (Arnett, Hammeke, & Schwartz, 1995). For purposes of clarity, Rey's measure will be referred to throughout the remainder of this text as the "Memory for Fifteen Items Test" (MFIT), which is its most commonly used name (Greiffenstein et al., 1994; Hayes, Hale, & Gouvier, 1998). The MFIT is a measure of immediate span of apprehension (i.e., short-term memory) (Leng & Parkin, 1995). It consists of an 8.5" x 11" card on which are printed 15 items (letters, numbers, and shapes) arranged in 3 columns and 5 rows (Appendix A). The examinee is told there are 15 <u>different</u> (emphasized) items to remember, which are to be reproduced immediately on a blank sheet of paper following a 10-second exposure to the stimulus card. Although it is presented as a difficult task, it is actually quite simple because there is redundancy among items that reduces the amount of information to be remembered (i.e., three main ideas). Thus, the MFIT relies upon the floor effect strategy for detecting malingering (Rogers et al., 1993), which assumes the naive malingerer will be misled into overplaying their role and choose to perform poorly on this very simple task.

Immediate memory span is an aspect of memory typically preserved, even in individuals with severe organic amnesia (Leng & Parkin, 1995). Although the assumption underlying the MFIT has intuitive appeal, empirical investigation is necessary to demonstrate validity and effectiveness for this purpose. Rogers et al. (1993) described three study designs through which this can be accomplished.

"Simulation designs" are analogue studies which involve instructing normal subjects to feign (simulate) a particular disorder, the performance of which can then be compared to that of individuals with the disorder. In a "known-groups" design, the performance of actual malingerers (identified by independent clinicians) is compared to that of individuals with and without the given disorder. Finally, "differential prevalence" designs involve making comparisons with a group assumed to be malingering based solely upon the context of the evaluation (i.e., presumption that individuals seeking compensation for head injury are more likely to malinger neuropsychological sequela).

In the following sections, empirical investigations from each of these categories, as well as strictly normative studies of the MFIT, will be examined. Particular emphasis will be placed on the instrument's sensitivity (i.e., percentage of malingerers correctly identified) and specificity (i.e., percentage of non-malingering patients correctly identified). This will be followed by an examination of the validity and effectiveness of the MFIT in detecting malingered memory deficits among various populations, in addition to the improved prospects offered through an alternative qualitative scoring method. Finally, the present study will be presented and these issues discussed in relation to use of the MFIT among pre- and post-trial forensic inpatients.

Definition of Statistical Terms

Before discussing investigations of the MFIT, a definition of the statistical concepts to be used throughout the remainder of the text is in order. Cutoff scores are often used with psychological tests to make decisions concerning the presence or absence of a given condition. Using the MFIT and malingering status as an example, scores falling below the cutoff of nine items are assumed to indicate malingering, whereas scores at or above the cutoff are assumed to not be indicative of malingering. A "true positive" occurs when an individual who is malingering is correctly identified as malingering by their test performance (falling below the cutoff); the perentage of true positives is referred to as the sensitivity of the test. Conversely, a "true negative" occurs when an individual who is not malingering is correctly identified as such. The percentage of true negatives is known as the specificity of the test. When an individual who is not malingering is incorrectly classified as malingering by the test, a false positive (Type I error) is said to occur, whereas a "false negative" or Type II error occurs when a malingerer is misclassified as not malingering. An illustration of these concepts is provided in Table 1.

Table 1

Classification Table for the MFIT

	Actual Condition	
Predicted Condition (based on MFIT scores)	Malingering	Not Malingering
Malingering (less than 9 items)	True Positive (Sensitivity)	False Positive (Type I Error)
Not Malingering (9 or more items)	False Negative (Type II Error)	True Negative (Specificity)

Gouvier, Hayes, and Smiroldo (1998), citing Faust and Nurcombe (1989), recommended a useful strategy for determining if a test is a valid and effective indicator of malingering. A test is considered a valid (accurate) indicator if the sensitivity divided by the false negative error rate exceeds the false positive error rate divided by the specificity. If a measure is an effective indicator (i.e., more accurate than the base rate), the base rate for the condition (i.e., number of malingerers divided by the total population) will be greater than the combined error rate (false positive + false negative) for the instrument. However, in situations where the base rate exceeds 50%, the equation for determining effectiveness becomes (1 – base rate > false positive + false negative). An important point is that while a test may be valid, it may still not be an effective indicator.

Normative Studies

It was first suggested by Lezak (1976) that a cutoff of less than nine items (three rows) correctly recalled on the MFIT be used in suspecting malingering, as only "significantly deteriorated patients" would recall fewer items. However, this suggestion was apparently based upon clinical observations rather than empirical fact. It was not until 10 years later that the first empirical investigation of the MFIT was conducted. In an effort to establish normative data, Goldberg and Miller (1986) examined the performance of 50 acutely disturbed psychiatric inpatients and 16 mentally retarded individuals on the MFIT. While none of the psychiatric inpatients fell below the suggested cutoff of nine items, 37.5% of the mentally retarded sample did. Furthermore, the authors recommended using the total number of items correct, as a greater number of individuals in both groups failed to meet the 3-row criterion. These findings raised an important issue regarding the influence of intellect on MFIT performance, as none of the psychiatric inpatients had IQ estimates in the mentally retarded range (mean IQ of 101.1, range of 70 to 123).

While the study of Goldberg and Miller (1986) offered some utility in investigating Lezak's (1976) claim pertaining to "significantly deteriorated", it failed to examine the MFIT with respect to a population with demonstrated memory impairment. Assuming that individuals feigning memory impairment would most likely attempt to imitate individuals with brain damage, Bernard and Fowler (1990) investigated the MFIT with 18 predominately diffuse head injury patients with significant memory impairment (M time postinjury = 18.9 weeks) and 16 normal controls. Using the 9-item cutoff, it was found that 11.2% of the patients with head injuries would have been incorrectly suspected of malingering (i.e., specificity of

88.8%), as compared to none of the controls. However, the authors noted that lowering the cutoff to less than eight items correctly recalled resulted in none of the head-injured patients being incorrectly classified; they subsequently recommended using this more conservative score.

Morgan (1991) conducted a normative study to establish the effect of true memory impairment on MFIT performance. The MFIT was administered to 60 neurology patients, none of whom were involved in litigation or disability proceedings, along with measures of new learning ability and memory. The sample consisted primarily of individuals with alcoholic encephalopathy and probable dementia of the Alzheimer's type, and all subjects were tested (where relevant) at a minimum of 1-month postinsult (e.g., after encephalitis or termination of alcohol consumption). Morgan found that 12 patients (20%) failed to achieve the criterion of at least nine figures correctly recalled ($\underline{M}$=5.4). A closer inspection of the data revealed that among those failing the MFIT, 92% obtained average memory impairment ratings in the severe to profound range, and the MFIT failure participants were significantly older. Morgan additionally noted there was no greater sensitivity of MFIT performance to verbal or nonverbal memory impairment, as the majority of subjects demonstrated global impairment. Morgan concluded that Lezak's (1976) claim regarding significantly deteriorated was supported, as only individuals with the most severe memory impairment failed to achieve the suggested cutoff.

Based upon the findings of Goldberg and Miller (1986) concerning the influence of intellect on MFIT performance, Hays, Emmons, and Lawson (1993) sought to provide ability-based norms for this measure. Among a sample of 300 adult psychiatric inpatients, it was found that IQ and age correlated highly with the total

number of items recalled on the MFIT (r=.60 and -.29, respectively), which taken together accounted for 43% of the variance in MFIT scores. Although the investigators failed to mention the diagnostic makeup of their sample, it was noted that 72 % of individuals with an IQ lower than 65 fell below the 9-item cutoff. This is almost double the number of subjects falling below the cutoff in the Goldberg and Miller (1986) study, who had an average IQ of 63.4 (range between 40 and 69).

A more recent normative investigation of the MFIT was conducted by Back et al. (1996). Under the premise that individuals feigning psychosis often feign cognitive impairment as well, Back et al. examined the performance of 30 individuals with schizophrenia on the MFIT. Using the 9-item cutoff, they found that 13% of the sample would have been incorrectly suspected of malingering. A regression analysis revealed MFIT performance to be unrelated to patients' mental and psychiatric status (as determined through performance on the Mini-Mental State Examination and ratings on the Brief Psychiatric Rating Scale), although education accounted for 37% of the variance in scores.

With the exception of the study by Morgan (1991), a major shortcoming common to normative studies has been the assumption that participants had no incentive(s) for feigning/exaggerating their complaints (e.g., gaining staff attention, acquiring better services, etc.). Nonetheless, they provide evidence that the specificity of the MFIT is compromised by one's level of intellect, particularly when intellectual deficiency is superimposed upon a psychiatric condition. Furthermore, individuals with severe memory impairment do perform below the suggested cutoff at a rather high rate. The MFIT therefore appears to be of limited utility in detecting

feigning/exaggeration of memory complaints among individuals with an established history of psychiatric impairment or intellectual deficiency.

Simulation Design Studies

The first study of the MFIT using a simulation design was conducted by Bernard (1990). Using a sample of college students, the performance of a control group was compared to that of a group instructed to "fake believable memory impairments" without financial incentive, and a group instructed to fake memory impairment with a financial incentive, across multiple measures of memory. No differences were found between groups in the total number of items recalled on the MFIT, and it is noted that the average scores for the two simulating groups (13.2 and 13.3, respectively) were well above the suggested cutoff. Bernard contended that this finding might have resulted from the placement of the MFIT at the end of a battery that contained more difficult tests, thus making the intent of the measure more obvious. He suggested the MFIT be placed at the beginning of test batteries to avoid compromising the test's detection ability. This contention was confirmed in a subsequent study by Bernard, Houston, and Natoli (1993). By placing the MFIT at the beginning of the battery, it was found that a group of college students simulating memory impairment recalled significantly fewer items than a group of controls. Although the authors noted the average number of items recalled by the simulated malingerers ($\underline{M}$=10.0) remained above the suggested cutoff, an inspection of the tables reveals this group did recall fewer correct rows ($\underline{M}$=2.2) than the recommended 3-row cutoff.

Schretlen, Brandt, Kraft, and Van Gorp (1991) conducted a more extensive study using two groups of simulators. In the first group, community volunteers and college students were instructed to feign either psychogenic amnesia, amnesia due to head trauma, or amnesia of unspecified etiology. The second simulating group was

composed of inpatients with substance abuse problems given instructions to feign "insanity." The performance of simulators was compared to that of 80 normal controls and a variety of patients including 10 amnesics with "severe" memory problems, 55 patients with a history of "moderately severe" traumatic brain injury (TBI), 9 patients with dementia (primarily Huntington's disease), 34 patients with severe mental illness (i.e., schizophrenia), 40 patients with a variety of neuropsychiatric diagnoses, and 7 civil litigants under suspicion of malingering memory complaints. It was found that the genuine amnesics, severe psychiatric patients, and individuals suspected of malingering recalled significantly fewer items than individuals simulating amnesia. No differences were found between normal controls, patients with TBI, and normals simulating amnesia, although the patients simulating insanity were found to recall significantly fewer items than the normal controls. None of the simulators fell below the 9-item cutoff.

Using the 9-item cutoff, a sensitivity of 0% was found for the normals simulating various forms of amnesia and 36% for patients simulating insanity. Comparatively, the sensitivity for the group suspected of malingering was 43%. While it is possible the performance of the two simulating groups was affected by the failure of the experimenters to provide detailed scenarios, this finding does bring into question the effects of simulating different conditions as well as the validity of using college students in simulation studies.

In terms of specificity, Schretlen et al. (1991) found that 73% of all patients combined fell above the 9-item cutoff. From this and the comparatively lower performance of the amnesics, the authors concluded that the MFIT lacked adequate specificity and was overly sensitive to genuine memory impairment. However, the

use of groups of widely disparate sizes (some with very low Ns) and failure to make statistical corrections for experiment wide error (i.e., comparisons between 9 groups using three dependent measures) violates statistical assumptions and brings into question the validity of these findings.

A simulation design study by Guilmette, Hart, Giuliano, and Leininger (1994) examined the efficacy of the MFIT in relation to an abbreviated version of the Hiscock Forced-Choice Procedure (A-HFCP). Participants included 20 inpatients in a brain injury rehabilitation hospital, 20 psychiatric inpatients with a predominant diagnosis of Major Depression, and 20 college students instructed to "fake believable memory deficits." Guilmette et al. found that the brain-injured subjects recalled significantly fewer items on the MFIT than both the simulated malingerers and psychiatric patients. Furthermore, 45% of the brain-injured subjects fell below the 9-item cutoff, as compared to only 15% of the simulated malingerers. In contrast, none of the brain-injured subjects and only 5% of the psychiatric patients were misclassified on the A-HFCP, and 85% of the simulated malingerers were correctly detected. While these results appear dismal for the MFIT, it is noted the brain-injured subjects consisted primarily of individuals who suffered from cerebrovascular accidents or head injuries, and who demonstrated moderate to severe cognitive deficits on a variety of neuropsychological tests (including memory). This is an important point to consider, given the claims of Lezak (1976).

Using an expanded scoring system incorporating both quantitative and qualitative methods, Arnett et al. (1995) conducted a two-part study comparing the performance of simulators (college students) with that of neurological patients on the MFIT. In the first phase, the performance of undergraduates was compared to a mixed

neurological sample consisting primarily of closed head injury with intracerebral hemorrhage, cerebrovascular accident, and brain tumor cases (M time since injury = 1.67 years). It was found that undergraduate simulators recalled significantly fewer correct rows in proper sequence, correct rows in any sequence, and total items in the correct location than neurological patients. However, there was no difference between simulators and patients in term of the number of correct items regardless of location (which is the common scoring method), and it is noted the average number of items recalled by the simulators was 12.1 (SD=2.6). Using the 9-item cutoff, sensitivity was found to be 63% and specificity 74%; these numbers changed minimally when using 8- and 7-item cutoffs. Arnett et al. concluded from this phase of the study that less than two correct rows in proper location be used as the cutoff, as this yielded an improved specificity of 97% (sensitivity = 47%).

The second phase of the Arnett et al. (1995) study was an attempt to replicate the findings of the first phase using a similar group of neurological patients and "more sophisticated" simulators (i.e., medical students). The same pattern of results emerged, and using the 9-item cutoff a similar rate of sensitivity and specificity was found (76% and 80%, respectively). The authors concluded the number of correct rows in proper location was the best discriminator, as only two neurological patients in the entire sample were misidentified.

Although the Arnett et al. (1995) study was beneficial in providing evidence that individuals with traumatic brain injury can perform adequately on the MFIT, certain limitations are worth noting. First, simulators were only instructed to exaggerate problems with brain damage, rather than specifically feign memory impairment. This may account for the findings of low sensitivity when using the

recommended cutoff. Moreover, simulators were tested in a group format and using only the MFIT, which the authors acknowledged may have affected the results. Second, the authors noted that four of the neurological subjects were not fully oriented to person, place, and time. While this may be indicative of significant cognitive impairment, no specific mention is made to their performance on the MFIT, thus their potential contribution to lowered specificity. Furthermore, the authors mentioned that all patients received a full neuropsychological battery, yet failed to report on the level of performance of this group. As such, it is unknown what the actual level of functioning was for this group (except that they all had IQ's above 70), which limits the conclusions to be drawn from this study regarding neurological impairment and MFIT performance.

Two major findings have emerged from simulation design studies of the MFIT. First, nearly all of these studies used samples of college students to simulate the performance of malingerers. While an advantage in this procedure is increased experimental control, it comes at the expense of external validity. Rogers et al. (1993) noted that it is unknown as to what extent the performance of these groups generalizes to that of malingerers in real-world settings. As Rogers and Cruise (1998) discovered, factors such as appreciation for negative incentives (i.e., consequences for unsuccessful feigning), ability to identify with the scenario, and relevance to participants can affect response styles. Given that these factors are often not incorporated into simulation studies, the validity of findings from these studies is questionable. This may explain the widely discrepant findings of sensitivity among these studies, i.e., 0% in Schretlen et al. (1991) to 76% in Arnett et al. (1995). Other noted confounds include variations in the amount of detail provided in scenarios,

disorders participants were instructed to feign (i.e., memory deficits vs. insanity), administration format (i.e., group vs. individual, beginning vs. end of test battery), and educational differences between simulators and patient controls. As such, estimates of sensitivity derived from simulation studies should be considered cautiously.

The second major finding from simulation studies of the MFIT comes in the form of additional normative data. Specifically, these studies provided further evidence that individuals with established neurological and psychiatric conditions resulting in severe cognitive impairment have greater difficulty in "passing" the MFIT than those individuals with less severe impairments. In keeping with Lezak's (1976) claim regarding "significantly deteriorated" individuals, the findings of low specificity among such populations is not unexpected. It could therefore be argued that it would be more appropriate to examine the specificity for such populations separate from that of individuals with less severe deficits, as these are the individuals malingerers are most likely to imitate (e.g., Greiffenstein et al., 1996; Iverson, 1995).

Differential Prevalence Designs

There have only been two studies examining performance on the MFIT using a differential prevalence design. The first of these was conducted by Lee, Loring, and Martin (1992), who compared the MFIT performance of 100 inpatients with temporal lobe epilepsy (TLE), 40 outpatients with neurological disorders not involved in litigation, and 16 individuals with neurological disorders involved in litigation. TLE patients had demonstrated memory impairment (less than 5th percentile) on at least one of four standardized memory tests, and were not involved in litigation. The non-litigating outpatient sample was comprised of a variety of neurological disorders (e.g., tumor, closed head injury), whereas the outpatients involved in litigation were predominately cases of mild closed head injury. Analyses revealed no differences in the total number of MFIT items recalled between TLE patients and non-litigating outpatients. However, both groups were found to recall significantly more items than the litigating outpatient group. Lee et al. recommended using a cutoff of seven or fewer items, as this yielded the best overall sensitivity and specificity.

A closer inspection of the Lee et al. (1992) study reveals a problem with the reported sensitivity and specificity. Instead of reporting values for a 9-item cutoff, the authors report values for an 8-item cutoff. In examining the data provided in their tables, use of the 9-item cutoff would produce a specificity of 93% for both the TLE and non-litigating outpatient groups, compared to a sensitivity of 32.5% for the litigating outpatients. Specificity changed minimally when using the recommended 7-item cutoff (i.e., 96% for TLE, 95% for non-litigating outpatients), and there was no change in sensitivity for the litigating outpatients. This small amount of change hardly warrants a recommendation for changing cutoff scores. Moreover, a cutoff of six

items would have been even more accurate, as no patient in their study obtained a score below 7 items. Nonetheless, the finding of high levels of specificity are encouraging, as Lee et al. noted the TLE patients with impaired memory were selected on the basis that their pathology almost always affects structures underlying new learning. Thus, there is evidence from a large sample that individuals with memory impairment can perform quite well on the MFIT.

In the other study using a differential prevalence design (Griffin, Normington, & Glassmire, 1996), the authors compared two groups with "no incentive" to malinger (i.e., permanently psychiatrically disabled individuals and "normals" from community programs) with a group of "possible malingerers" (i.e., individuals filing psychological disability claims). It was found that 19.8% of the possible malingerers fell below the 9-item cutoff, as compared to only 9.4% of the normals. Problematic was the finding that a higher percentage of the permanently disabled group (28.3%) fell below the recommended cutoff.

Rogers et al. (1993) noted an intrinsic problem with differential prevalence designs is the assumption that the context of the evaluation results in dissimilar rates of feigning (i.e., individuals in litigation are more likely to be malingering). It is virtually impossible to identify true prevalence rates for malingering from these designs, given no other objective criteria for malingering are used. Moreover, these studies often overlook the possibility that individuals with "no incentive" for malingering based on a given criterion (e.g., non-litigating status) do not have other incentives for malingering (e.g., avoidance of responsibility) (Greiffenstein et al., 1994). This is best exemplified when the results of the Lee et al. (1992) study are compared with those of Arnett et al. (1995). Specifically, Arnett et al. examined the

MFIT performance of their neurological patient group with respect to the patients' litigation status. No significant differences were found on any of the MFIT indices under investigation between patients in litigation and those not in litigation; moreover, the authors noted there was a trend for the patients in litigation to actually perform better. This stands in contrast to the inferior performance of the litigating outpatients in the Lee et al. study, from which it can be concluded that estimates of sensitivity from differential prevalence studies are quite variable and potentially unreliable.

Known-Groups Designs

To date, five studies have examined the MFIT using a known-groups design. In an effort to improve upon previous designs, Greiffenstein et al. (1994) devised objective criteria (i.e., improbable outcomes for mild head injury) for identifying a group of individuals with persistent postconcussive syndrome (PPCS) with high likelihood of malingering. Criteria included improbable poor performance ($\geq$ -3 SD) on two or more neuropsychological measures, contradiction between collateral sources and symptom history, total disability in a major social role after one year, and claims of remote memory loss. Mildly head injured PPCS patients who met two or more of these criteria were classified as probable malingerers (N=43). Their performance was compared to that of 33 neuropsychological referrals with severe TBI (i.e., comatose admission status with medical instability), and 30 referrals with mild head injury and PPCS but no objective malingering markers. Comparisons were made across several measures of memory, including the MFIT. It was found that the probable malingerers recalled significantly fewer correct items on the MFIT than both the severe TBI and mild head injury with PPCS groups.

With regard to sensitivity and specificity, 62% of probable malingerers were correctly classified using a cutoff of nine items, whereas a specificity of 88% was found for TBI patients and 93% for PPCS patients. Although the sensitivity of 62% was not significantly greater than Greiffenstein et al.'s (1994) estimated base rate of 59% for probable malingering, they noted that this estimate may have been inflated due to referral source bias (i.e., majority of the mild head injury clients were referrals from insurance companies and attorneys). As such, the authors stated a more reasonable estimate of base rates for malingering among PPCS clients lies in the range

of 33 to 60%. Whereas the findings of specificity are encouraging, all of the cases studied were "a number of years" postinjury, which limits the generalizations one can draw to more acute cases.

Based upon the findings that criminal defendants often claim amnesia to either prove incompetence or escape criminal responsibility, Simon (1994) conducted an examination of the MFIT among non-retarded forensic inpatients. The experimental group was comprised of 14 male pretrial defendants who had been given a diagnosis of malingering by the staffing team, using a wide range of information sources (excluding the MFIT) and in accordance with DSM-III-R criteria (APA, 1987). The performance of this group was compared to a control group of 14 forensic inpatients with serious mental illness. All of the control subjects had been legally declared "Not Guilty by Reason of Insanity" (NGBRI), and thus were assumed to have no incentive to malinger. Using the recommended 9-item cutoff, it was found that 85.7% of both groups were correctly classified. Using the recommended 3-row cutoff, sensitivity increased to 100%, although this came at the expense of reducing specificity to an unacceptable 43%.

Aside from the relatively small number of subjects studied, Simon (1994) noted some factors that may have affected his findings. First, it was noted that all of the suspected malingerers were rather extreme in their presentations. Thus, lower rates of sensitivity would be expected among more sophisticated malingerers and those attempting to feign more subtle deficits. Second, the control group consisted of patients with chronic psychosis and/or dementia. Although the average IQ for controls was 90, it was still significantly correlated with the total number of items recalled ($\underline{r}$=.72). Thus, lower specificity is not unexpected when one takes into consideration

the demonstrated combined effects of psychiatric illness and intellectual functioning on MFIT performance (Hays et al., 1993). This was further demonstrated in a known-groups investigation by Hayes, Hale, and Gouvier (1997), who found that non-malingering mentally retarded forensic inpatients performed worse than mentally retarded pretrial defendants diagnosed with malingering across several standard malingering measures (including the MFIT).

Millis and Kler (1995) used below chance performance on a forced-choice symptom validity test (Recognition Word Memory Test [RMT]; Warrington, 1984) to identify a group of "clinical malingerers" from among a sample of individuals reporting closed head injuries and pursuing compensatory litigation. The MFIT performance of this group (N=7) was compared to that of seven inpatients with TBI in the acute phase of recovery (mean time post-injury of 1.1 months). It was found that the malingering group recalled significantly fewer total items (M=8.0) than the TBI group (M=12.1). Millis and Kler further reported a sensitivity of 57% and specificity of 100% when using the 7-item cutoff recommended by Lee et al. (1992). They advised against using the 9-item cutoff, as specificity dropped to 71%.

A re-examination of the data supplied by Millis and Kler (1995) reveals these estimates of sensitivity and specificity to be in error. Apparently, they considered the 9-item cutoff to mean nine or fewer items recalled, when in fact it was intended to mean less than nine items. As such, use of the 9-item cutoff continued to yield a specificity of 100%, as all TBI patients recalled nine or more items. This is an encouraging finding, as this group was reported to have sustained moderate to severe injuries (i.e., Glasgow Coma Scales at time of injury between 13 and 3). Even more encouraging is the fact that only two of the clinical malingerers recalled nine or more

items, which places sensitivity at 71%. However, the fact that all of the clinical malingerers were selected on the basis of below chance performance on another measure makes it highly probable they were rather blatant in their deception strategies. Also, the restricted size of the sample makes it difficult to generalize these findings.

In a follow up to their 1994 study, Greiffenstein, Baker, and Gola (1996) examined alternative scoring methods for the MFIT using an expanded sample of TBI patients (N=60) and probable malingerers (N=90). Using a more conservative 10-item cutoff (fewer than 10 items recalled) derived through discriminant function analysis, sensitivity for the possible malingering group was found to be 64%, comparable to the rate of 62% found using the 9-item cutoff in the 1994 study. Unfortunately, specificity dropped from 88% to 72% for the TBI group when using the raised cutoff. The authors surmised that inclusion of TBI patients with dense amnesia (N=5) was inappropriate, as there was no practical clinical basis for giving this measure to individuals with objective severe neurological disease, and that malingering is not typically an issue with these patients. Removing these cases from the analyses, specificity with the 10-item cutoff increased to 78%. Given that sensitivity did not change between these studies whereas specificity was adversely affected, it stands to be argued that the 9-item cutoff be maintained, with malingering suspected whenever 8 or fewer items are recalled.

In summary, the known-groups investigations of the MFIT offer more promising results regarding the instrument's discriminatory power than both simulation and differential prevalence studies. This is relevant, given the limitations and questionable validity of the latter designs mentioned by Rogers et al. (1993). While the results from these investigations are believed to be more generalizable to

real-world samples, a limitation remains the lack of reliable and uncontaminated criteria for identifying malingerers, although the Greiffenstein et al. studies were designed to address this issue. Another problem concerns the fact that the findings reported here are primarily applicable to individuals seeking civil litigation (i.e., disability and/or compensation claims secondary to mild head injury). Only two studies have investigated the MFIT among pretrial criminal defendants (Hayes et al., 1997; Simon, 1994), and three of these studies (Hayes et al., 1997; Millis & Kler, 1995; Simon, 1994) were restricted in their sample sizes. The latter two studies also had an increased likelihood the diagnosed malingerers were rather blatant in their deception strategies. As a consequence, generalizations based upon these findings are necessarily limited. There remains a need for replication with larger samples, in particular criminal defendants, and including those who are less blatant in their efforts to feign mental disorders or cognitive impairment.

Summary of Findings

This review provides evidence the MFIT has adequate specificity among a variety of populations. Five studies provided data from which estimates of specificity could be established for psychiatric patients, from which the following can be concluded. Specificity tends to be high for non-retarded acutely ill psychiatric inpatients (e.g., Goldberg & Miller, 1986, 100%; Guilmette et al., 1994, 90%) and non-retarded individuals with chronic forms of psychosis (combined data from Back et al., 1996, and Simon, 1994, 87%). Combining these data with the data from Hays et al. (1993), the overall specificity for non-retarded psychiatric patients is 83.8% (268 of 320 patients correctly identified). Although Back et al. found MFIT performance to be unaffected by symptom severity and mental status, Schretlen et al. (1991) reported very poor performance for their "severe psychiatric" group. It should be noted that "severe" was never defined in the later study, although scores on the Mini-Mental State Examination were highly correlated ($\underline{r}$=.81) with MFIT items recalled for the combined groups of neurological and psychiatric patients.

A much lower specificity of 62.5% was found for individuals with a diagnosis of mild to moderate mental retardation (Goldberg & Miller, 1986), and even lower rates (32.4%) among psychiatric inpatients with intellectual functioning in the moderate to mild mental retardation range (Hays et al., 1993). While the majority of patients studied suffered from some form of psychosis or depression, a setback in these studies was failure to adequately specify the diagnostic makeup of the sample and provide symptom severity ratings (e.g., Goldberg & Miller, 1986; Hays et al., 1993; Schretlen et al., 1991). This limits the conclusions that can be reached regarding particular diagnostic groups, although the evidence suggests that individuals

with a psychiatric diagnosis other than mental retardation can perform adequately on the MFIT.

For neurological conditions, individuals having suffered intracerebral hemorrhage or injuries resulting in severe global cognitive impairment or severe organic amnesia had the poorest recall on the MFIT (Greiffenstein et al., 1996; Guilmette et al., 1994; Morgan, 1991; Schretlen et al., 1991). Otherwise, specificity was adequate independent of whether the nature of the insult was diffuse or specific (Bernard & Fowler, 1988; Lee et al., 1992), or whether the individual was in the acute or late phase of recovery (Greiffenstein et al., 1994; Millis & Kler, 1995). Specificity estimates for these populations ranged from 74 to 100%, and of the studies reviewed, six provided data from which an overall specificity could be computed (Arnett et al., 1995; Bernard & Fowler, 1988; Greiffenstein et al., 1994; Lee et al., 1992; Millis & Kler, 1995; Morgan, 1991). From these studies, a cutoff of nine items correctly recalled resulted in a specificity of 85.8% (266 of 310 patients) for individuals with neurological conditions, most of whom were experiencing significant problems with memory.

Less encouraging are the findings of sensitivity for the MFIT. When examining data from simulation studies, the MFIT was found to have inadequate sensitivity. Three studies provided sufficient information from which an overall estimate of sensitivity could be calculated (Arnett et al., 1995; Guilmette et al., 1994; Schretlen et al., 1991). From these studies, 53 out of 141 total simulators failed to recall nine or more items, which yields an overall sensitivity of 37.5%. Even lower figures were obtained when the results of the two differential prevalence designs were combined (Griffin et al., 1996; Lee et al., 1992). Of the 107 individuals suspected of

malingering in these two studies, 24 failed using the 9-item cutoff, resulting in a dismal sensitivity of 22.4%. As noted, the validity of these designs and the generalization of the findings is questionable. This is evident in the much higher rates of sensitivity found in the studies using diagnosed malingerers. Examining the combined data from these studies (Greiffenstein et al., 1994; Millis & Kler, 1995; Simon, 1994), 44 out of 64 individuals diagnosed as malingering through independent means were correctly identified through their performance on the MFIT, producing a more encouraging sensitivity of 68.8%. For civil litigants seeking disability and/or financial compensation for their injuries, the overall sensitivity was 64%; for forensic evaluatees, 85.7%.

Using data from the known-groups studies, the 9-item cutoff for the MFIT appears to be a valid indicator of malingering among civil litigants/disability claimants (.64/.36 > .14/.86). It is also an effective indicator, whether one uses more liberal (i.e., 60 to 66%) or conservative (15 to 33%) estimates of malingering base rates (e.g., 1 - .66 > .122 + .05; .33 > .122 + .05). The small combined error rate (17.2%) results in a rather broad base rate effectiveness range (i.e., effective when the base rate falls between 18 to 82%). Based on this data, the MFIT is both valid and effective over a broad range in distinguishing between individuals malingering memory problems from those with memory problems due to neurological causes.

Only one study has provided information from which validity and effectiveness can be determined for criminal defendants (Simon, 1994). Using the data from this study and the findings from both non-retarded and retarded psychiatric patients (as these are the groups criminal defendants are most likely to attempt simulating; Rubinsky & Brandt, 1986), the MFIT appears to be a valid indicator for

malingering memory problems secondary to psychiatric illness (.857/.143 > .162/.838) and psychiatric illness plus mental retardation (.857/.143 > .676/.324). Given an estimated 16 to 20% malingering base rate among criminal defendants, the MFIT appears to be effective at identifying malingerers among non-retarded psychiatric inpatients (combined error rate of 16.2%), but not among psychiatric patients with mental retardation (combined error rate of 59%). However, definite conclusions cannot be made in this regard, as false negative estimates were derived from a single study with only 14 malingerers, all of which were noted to be extreme in their presentations. Further research with this population is clearly indicated to determine the effectiveness of the MFIT.

Qualitative Scoring

Using the recommended quantitative scoring system, a minimum 10% of both neurological and psychiatric patients would be incorrectly suspected of malingering memory complaints. In cases involving a dual diagnosis of psychosis and mental retardation, the rates may be greater than 64%. Although it would be inappropriate to establish a diagnosis of malingering on the basis of a single screening measure and in the absence of other objective data, important ramifications can still result. For example, one may be subjected to more extensive testing, stigmatized with a label and suffer loss of reputation, lose deserved financial compensation, or even face an undeserved prison sentence. Thus, modifications that can increase specificity and improve effectiveness warrant investigation.

One method has been lowering the cutoff for the MFIT. On the basis of their respective findings, several investigations (e.g., Guilmette et al., 1994; Lee et al., 1992) have recommended using lower cutoffs either for the total number of items or the total number of rows recalled. However, it was noted these alternate cutoffs were found to either result in minimal changes in specificity, or were erroneously based upon miscalculations of the data. More importantly, they resulted in decreasing the sensitivity of the MFIT (i.e., Greiffenstein et al., 1996). Thus, what is needed is a method that increases specificity while either maintaining or even improving sensitivity. One potential method advocated by Rogers et al. (1993) is the "magnitude of errors" strategy. This method simply involves examining the differences in patterns of incorrect responses between either simulators or known malingerers and relevant clinical comparison groups. This can be accomplished through examining the qualitative rather than quantitative errors on the MFIT.

The examination of qualitative errors is not new to neuropsychology. In 1961, Benton and Spreen examined the differences in qualitative errors between normals simulating brain damage and actual head injured patients on the Benton Visual Retention Test. It was found that simulators made more errors of distortion of recalled designs (i.e., bizarre figures) while making fewer perseverative and omission errors than actual brain damaged subjects. Interestingly, it was observed in the first normative study of the MFIT (Goldberg & Miller, 1986) that mentally retarded patients were more prone to making errors involving perseveration and reversal, whereas psychiatric patients typically omitted a single row. Although the authors failed to provide scoring criteria for qualitative errors, they suggested these error types be scrutinized to avoid false positives.

Several of the studies mentioned have examined qualitative errors in some fashion. Only one type of qualitative error (i.e., repeated items) was examined by Schretlen et al. (1991). While no significant differences were found between any groups on this type of error, it was significantly correlated with IQ ($\underline{r}$=-.29), a finding consistent with that of Goldberg and Miller (1986). Morgan (1991) found the majority (40%) of his neurological sample made the error of misordering geometric shapes (actually, 65% of those able to recall the shapes). Errors of perseveration (repetition) were also common, while less common errors involved repeating rows or continuing rows (e.g., extending letters from A to E). It was further noted that 93% of the sample recalled all of the capital letters and 83% recalled all Arabic numerals, thus providing evidence for a primacy effect in recall. Morgan therefore suggested that failure to recall these two rows should raise the suspicion of malingering, as these were cases of true memory impairment and 75% of those who failed the MFIT still recalled the

capital letters. Interestingly, one would expect better recall of items from the last two rows, given the findings of a recency effect in the short-term recall of amnesiacs (Wiggins & Brandt, 1988). This unexpected pattern of performance may be attributable to a combination of short exposure time (10 seconds) and impaired processing speed.

In the simulation study of Guilmette et al. (1994) it was briefly mentioned that some of the simulators added additional items in their recall, which led the authors to suggest that a qualitative analysis may improve sensitivity for the MFIT. Conversely, no differences were found between simulators and neurological patients on any of the 14 qualitative indices examined in the two simulation studies of Arnett et al. (1995). Hypothesizing that qualitative errors would be less related to those factors shown to affect specificity (i.e., age and intelligence), Griffin et al. (1996) administered the MFIT to 90 undergraduate simulators to develop a new qualitative scoring system. They identified 10 qualitative "malingering" errors, which were examined with their groups of analog malingerers, psychiatrically disabled non-malingerers, and "normal" (less severely ill) controls. The disabled group made significantly more "wrong item" errors (i.e., production of recognizable figure not on stimulus card) than the control group, and the control group made significantly more "Roman numeral" errors (i.e., tally marks drawn as Roman numerals) than the disabled group, which the authors judged to be artifacts of illness. The analog malingerers were found to make more dyslexic (i.e., character reversal), embellishment (i.e., elaboration or adornment of recognizable character), gestalt (i.e., failure to make a 3 x5 arrangement when recalling 15 items), and row sequence errors than the combined groups of disabled and

controls. Moreover, it was found that analog malingerers produced significantly more total qualitative errors than the non-malingering groups.

The hypothesis of Griffin et al. (1996) regarding the influence of age and education was weakly supported, as age accounted for 4.1% of the variance in total qualitative error scores for possible malingerers and IQ accounted for 4.1% of the variance in qualitative errors for the non-malingering sample. However, there are major limitations to these findings. The authors varied the MFIT instructions (i.e., added instructions to write down items "just as they appeared on the card") in order to produce more scorable errors. This goes against standard administration procedures, and implies that under standard conditions few qualitative errors will be produced. This is especially bothersome, given that Griffin et al. noted a low base rate for qualitative errors (i.e., average analog malingerer made only one type of qualitative error).

Finally, only one study using a known-groups design has investigated a qualitative scoring method for the MFIT. In the Greiffenstein et al. (1996) study, alternative scoring strategies were examined for the MFIT. While these were not reported as qualitative methods per se, an inspection of their scoring criteria reveals that one particular method (i.e., "spatial" scoring) would constitute a qualitative method. Specifically, this score reflects a correct ordering of elements within rows, similar to the "misordering" error of Morgan (1991). Interestingly, the use of spatial scoring was found to result in improved rates of sensitivity (69%) and specificity (82%). While this study did not investigate other types of qualitative errors, it does provide preliminary evidence that qualitative methods may produce better results when using a valid research design.

In summary, there are indications that individuals with known organic dysfunction and/or impaired cognitive functioning make specific types of qualitative errors. These mainly include errors of perseveration and misordering of elements, which is not unexpected given the nature of the underlying pathology (e.g., Malloy & Richardson, 1994). Furthermore, one study (Morgan, 1991) found that individuals with organically based memory problems exhibited a primacy effect in their MFIT recall. Qualitative scoring may therefore prove useful in improving the detection ability of the MFIT. However, several limitations preclude definite conclusions in this matter. This includes a lack of uniformity in the types of errors examined across studies, and the fact that no studies using a known-groups design have systematically investigated the many identified types of qualitative errors. Thus, there remains a need for more research in this area.

The Present Study

There is evidence the MFIT has adequate sensitivity and specificity with a variety of populations. However, a number of shortcomings have been identified, including inadequate sample sizes, inappropriate inclusion of severely brain injured patients, failure to make statistical corrections for multiple comparisons, failure to include diagnostic information and symptom ratings for psychiatric patients, failure to consider alternative incentives for and assessment of feigning/exaggerating complaints among "non-malingering" subjects, miscalculations of data, and reliance upon study designs of questionable ecological validity. Despite these shortcomings, the MFIT appears to be a valid and effective screening measure for malingered memory complaints within neuropsychological/civil forensic settings. There are also indications that a qualitative scoring method may enhance the MFIT's efficacy in this regard, although no systematic investigation of this method using a sound research design has been conducted.

One group for whom the MFIT may prove useful is pretrial criminal defendants. These individuals often present with memory deficits relating either to the crime itself or as part of a clinical picture in which cognitive deficits are an associated feature (e.g., psychosis) to procure an insanity verdict or be seen as incompetent to stand trial (Rubinsky & Brandt, 1986; Schacter, 1986a; 1986b). However, there has only been one investigation to date examining the performance of this group with the MFIT (Simon, 1994). While this study possessed ecological validity, conclusions were limited due to the small sample size and use of subjects who were extreme in their presentations. Interestingly, Rogers and Cruise (1998) have contended that criminal defendants would be more likely to give extreme presentations, considering

the negative consequences of a conviction and that extreme symptoms may be thought to be necessary for acquiring an insanity verdict. While this has intuitive appeal and offers encouragement that such individuals would be easily identified through their performance on malingering measures, it does pose one major confound. The clinical populations that criminal defendants may be attempting to mimic (individuals found not guilty by reason of insanity) are often characterized by severe and chronic psychosis and significant intellectual/cognitive deficits. The observation that these populations have been noted to perform quite poorly on the MFIT (e.g., Hays et al., 1993) poses a serious threat to the ability of the quantitative scoring system in differentiating these groups. Thus, there remains a need for a known-groups investigation of the MFIT using both a larger sample of criminal defendants and relevant comparison group, and investigating alternative scoring strategies.

The present study extended upon previous investigations of the MFIT through the following methods. First, this study examined both quantitative and qualitative scoring methods for the MFIT using a known-groups design. The experimental group consisted of an archival group of pretrial criminal defendants (forensic inpatients judicially ruled as incompetent to stand trial) who were independently diagnosed as malingering and given the MFIT under natural clinical conditions (i.e., as part of a comprehensive psychological evaluation for malingering). In keeping with the recommendations of Rogers and Cruise (1998), the performance of this group was compared with that of two relevant clinical comparison groups (e.g., individuals with similar criminal backgrounds and individuals with similar psychiatric symptoms). These consisted of a second archival sample of pretrial (PT) forensic inpatients who had not been suspected of malingering and also given the MFIT under normal

conditions, and a current control sample of forensic inpatients found NGBRI by the local state courts.

Second, this study extended upon previous research by verifying the presumed "non-malingering" status of the NGBRI comparison group. This was done through the administration of additional malingering screening measures (e.g., M-Test, Dot Counting Test). Last, information pertaining to the MFIT's sensitivity, specificity, validity, and effectiveness was computed for both the quantitative and qualitative scoring methods, both separately and combined.

Hypotheses

The following hypotheses were offered: 1) Patients diagnosed as malingering would recall fewer total items on the MFIT than non-malingering patients (PT and NGBRI); 2) Since psychiatric patients often exhibit varying degrees of intellectual and cognitive impairment, it was hypothesized the non-malingering patients would make more qualitative errors involving perseveration, misordering of elements, and character reversals and rotations than diagnosed malingerers; 3) Diagnosed malingerers, in an attempt to portray themselves as insane, would make more qualitative errors involving bizarre and unusual reproductions and addition of extraneous elements than non-malingering patients; 4) Diagnosed malingerers were expected to recall fewer items from the beginning of the MFIT (i.e., failure to show a primacy effect) than the non-malingering patients (i.e., Morgan, 1991); and 5) Linear discriminant analysis incorporating both quantitative and qualitative variables would yield substantially better discriminating power than the total number of items recalled alone.

Methods

Participants

Three groups of patients were examined in the present study. The first group, diagnosed malingerers (DM), consisted of an archival sample of 44 adult male inpatients (pretrial criminal defendants) evaluated at a state forensic hospital between the years 1991 and 2000. Individuals in this group had been judicially declared incompetent to stand trial by the local courts and remanded for further evaluation and treatment. These patients were identified by the interdisciplinary treatment team (i.e., psychiatrist, psychologist, social worker, nurse, and security staff) as either feigning or exaggerating problems with psychosis, cognitive dysfunction, or both. This was accomplished on the basis of behavioral observations of inconsistent and/or atypical symptom presentations (e.g., unable to do the simplest arithmetic yet observed playing spades on the unit) and/or improbably poor performance on the initial mental status and competency examinations. Patients in this group were subsequently referred for a psychological evaluation and given the MFIT, along with other measures of malingering, as part of a comprehensive battery. Based upon behavioral observations and the combined results of assessment, these individuals were diagnosed as malingering using the indices outlined in either the DSM-III-R or DSM-IV (APA, 1987; 1994). Although the results of psychological assessment were utilized in reaching a final diagnostic decision in most cases, in none of these cases did the MFIT contribute to the determination of the diagnosis of malingering.

For comparison, data was collected on a second archival patient group consisting of 32 adult male pretrial criminal defendants (PT) also found incompetent to stand trial and remanded for further evaluation and treatment during the same time

period. Patients in this group were selected on the basis they had not been suspected of or diagnosed as malingering and had also been administered the MFIT, although as part of a more comprehensive psychological (i.e., diagnostic) or neuropsychological assessment to check for level of effort. As with the DM group, the MFIT did not contribute to the determination of PT patients' non-malingering status.

According to Rogers and Cruise (1998), comparison groups in malingering research should be relevant to forensic evaluations in terms of having a representative range of mental disorders and criminal backgrounds. Although the PT group would appear to meet this criteria, that they were not suspected of or diagnosed as malingering does not rule out the possibility that some patients in this group were in fact malingering. Data was therefore collected on a third and current group of 30 adult male forensic inpatients who had been judicially declared NGBRI by the courts and remanded to the same state forensic hospital for continued care and treatment. This group was selected for purposes of comparison because: 1) They have a representative range of mental disorders and criminal backgrounds; 2) They have a higher incentive for "faking good" for purpose of release, and thus are considered less likely to be "faking bad;" 3) They are a highly relevant comparison group in that the NGBRI status is the typical goal of pretrial malingerers.

Materials

Rey 15-Item Visual Memory Test. The Rey 15-Item Visual Memory Test (MFIT; Lezak, 1976; Rey, 1964) was developed as a screening measure for malingered memory impairment. It consists of 15 items, arranged in three rows and five columns, which are printed on an 8.5" x 11" sheet of paper. The respondent is told they will be shown a card that has 15 different (emphasized) symbols on it. They

are further informed they will have 10 seconds to view the card, after which time the card will be taken away and they will be asked to draw everything they can remember. The traditional quantitative scoring method involves simply counting the total number of items correctly recalled, with scores of less than nine items recalled used in raising the suspicion of malingering.

For the present study, MFIT protocols for all participants were scored in terms of the number of correct items recalled and along the various qualitative dimensions identified through previous investigations. These qualitative errors were defined to the fullest extent possible based upon the descriptions provided in the respective studies.

Negative Impression Management Scale. The Negative Impression Management (NIM) scale of the Personality Assessment Inventory (Morey, 1991) was used to screen for the exaggeration of psychopathology among the NGBRI participants (Appendix B). It consists of nine items, selected on the basis of very low endorsement rates, which represent an exaggerated unfavorable self-impression or extremely bizarre and unlikely symptoms (e.g., "I have visions in which I see myself forced to commit crimes") (Morey, 1996). Items are rated by the respondent on a 4-point scale, ranging from 0 'false, not at all true' to 3 'very true' which are then summed to obtain a total raw score (possible range of 0 to 27). The PAI was standardized with large clinical, normal census-matched, and college student samples (N>1,000 each). Overall, good psychometric properties were reported by Morey (1991). The median Cronbach's alpha across the 22 full scales was above .80 for the entire normative sample, with an average 3-week test-retest reliability of .78 for a combined sample of college students and community volunteers.

For the NIM scale, an alpha of .74 was reported for the normative clinical sample and a 3-week test-retest reliability of .75 for the combined normal sample. Morey (1991) suggested that a raw score of 13 or more be used as a cutoff, as scores in this range were more than 2 standard deviations above the mean for the clinical sample. He reported on a simulation study which revealed that use of this cutoff resulted in correct classification of 86.5% of college student simulators, while only 5.9% of a clinical sample and 1.5% of a normal sample were misclassified. Also, a known-groups study by Rogers, Sewell, Cruise, Wang, and Ustad (1998) revealed that NIM scores above 13 were a useful screen for identifying criminal forensic malingerers. Morey (1996) provides further clinical interpretations for T-score ranges, as well as a summary of additional simulation design investigations of the scale. In the present study, a raw score of 13 or more was used as the cutoff for possible malingering.

M-Test. The M-Test (Beaber et al., 1985) was also used to assess the possibility of feigning or exaggeration of psychiatric symptoms among the NGBRI participants. The M-Test was developed as a screening measure to specifically detect the malingering of schizophrenic symptoms. It consists of 33 true-false items, divided into three scales. The C (Confusion) scale contains eight "attitude-belief" items, none of which involve symptom endorsement, and which subjects are expected to endorse in a fixed direction (e.g., "I believe that cancer is a horrible disease" should be answered true). These items are used to assess comprehension, and are placed at the beginning of the test. The body of the test contains randomly ordered items from the M (Malingering) and S (Schizophrenia) scales. The M scale items reflect atypical and extremely severe symptoms not characteristic of mental illness (e.g., "I believe that

God has appointed me to teach the Zolan beliefs to all people that I meet"), whereas S scale items are genuine symptoms associated with schizophrenia (e.g., "Periodically I am bothered by hearing voices that no one else hears"). Using a sample of schizophrenia patients and undergraduate simulators, a score of 4 or more positively endorsed M scale items was identified as the cutoff for suspected malingering in the pilot study of Beaber et al. (1985).

In a review of studies using the M-Test, Smith (1997) noted that simulation studies provided encouraging estimates of sensitivity and specificity, whereas studies with suspected malingerers yielded more variable results. In an effort to improve the utility of this instrument, Rogers, Bagby, and Gillis (1992) developed "Rule-Out" and "Rule-In" criteria for the M-Test using those items with the highest positive and negative predictive power as determined through a known-groups investigation. Two options were provided, one that was more conservative and maximized specificity (Option A) and one which maximized sensitivity (Option B). Both the original and revised scoring criteria were applied to the M-Test data for the NGBRI group, as well as those archival patients who received the measure.

Dot Counting Test. The Dot Counting Test (DCT) is another screening measure for malingering developed by Andre Rey (Lezak, 1995; Rey, 1941). It consists of 12 cards on which are printed a series of dots. The first 6 cards contain sets of ungrouped dots, whereas the last 6 cards contain sets of grouped dots. Respondents are presented with the cards in a fixed non-sequential order, and instructed to count the number of dots as quickly as possible. It is expected the non-malingering patient will take longer to count the ungrouped dots than the grouped dots, with increases in counting time proportional to increases in the number of dots. Binks, Gouvier, and

Waters (1997) found the total number of incorrect responses for both grouped and ungrouped dots provided better discrimination than five other indices derived from DCT protocols between college student simulators and a heterogeneous sample of patients referred for neuropsychological evaluation. For the present study, the cutoff for suspected malingering was established as either the total grouped counting time exceeding the total ungrouped counting time, or the total number of errors (grouped plus ungrouped) greater than one standard deviation above the mean of the patient sample of Binks et al. (1997) (i.e., more than 4 errors).

Expanded Brief Psychiatric Rating Scale. The Expanded Brief Psychiatric Rating Scale, Version 4.0 (BPRS-E; Ventura et al., 1993) was used to obtain current (2-week) ratings of psychiatric symptoms for the NGBRI participants (see Appendix C). The BPRS-E is used to assess 24 separate dimensions of psychiatric symptoms (e.g., hallucinatory behavior, flat affect). Each dimension is rated on a behaviorally anchored 7-point Likert scale, with ratings ranging from 1 'not present' to 7 'extremely severe.' NGBRI participants were assessed by doctoral students trained to a minimum intraclass correlation coefficient (ICC) of .80 according to University of Chicago Center for Psychiatric Rehabilitation criterion ratings. Ratings were based upon a combination of behavioral observations and participants' answers to standard questions posed in an interview format.

Shipley Institute of Living Scale. The Shipley Institute of Living Scale (SILS; Zachary, 1991) was used to provide an estimate of the general intellectual functioning of NGBRI participants. The SILS has been widely used in a variety of settings with both adolescents and adults to provide a brief measure of verbal intellectual ability. It consists of a Vocabulary subtest containing 40 items assessing verbal skills, and an

Abstraction subtest containing 20 items which assess abstract thinking skills. Age-based norms are provided from which the estimated WAIS-R Full Scale IQ can be computed from the total score. Since the SILS' inception in the 1930's, additional norms have been collected for a large (N=290) sample of mixed psychiatric patients. Furthermore, Zachary reported on two studies which found a correlation of .85 between SILS estimated IQ and actual WAIS-R IQ's for adult psychiatric patients.

Procedures

The present study was reviewed and approval granted by the Institutional Review Boards of Louisiana State University, the Louisiana Department of Health and Hospitals, and the Eastern Louisiana Mental Health System Forensic Division. The hospital charts of all patients (both archival and current) were then thoroughly reviewed for the following sources of information and entered onto a demographics form (Appendix D). First, demographic information of theoretical relevance to MFIT performance (age at the time of testing, years of education, and any available IQ scores) were collected. Second, information for determining the groups was collected. This included the number of prior arrests, commitment charge(s), and maximum possible prison sentence as outlined in West's (1998) Louisiana State statutes. Last, information useful for verifying patients' malingering status was gathered (i.e., psychiatric diagnoses, prior suspicion or diagnosis of malingering, documented behavioral observations of staff, reasons for testing referral, results of any competency and malingering assessments), as was other potentially relevant data (i.e., medication status, documented history of neurological injury, number of previous hospitalizations for psychosis, documented history of mental retardation).

Selection Procedures for the DM and PT Groups. The selection criteria for inclusion in the archival DM group included 1) pretrial legal status at the time of evaluation 2) a diagnosis of malingering by the interdisciplinary treatment team, and 3) having been administered the MFIT, with a copy of the patient's reproduction available for re-scoring. The same criteria were used in selecting archival PT subjects, with the noted exception that malingering had been definitely ruled out by the treatment team.

Selection Procedures for NGBRI Participants. Potential participants in the NGBRI group were screened by a clinician independent of the study for the presence of factors that could impair ability to provide informed consent (e.g., severe disorganization or intellectual impairment). Patients who were determined to be suitable for the study and capable of giving consent, and for whom there were no current or past indicators or diagnoses of malingering (as determined through chart review) were approached for participation and had read to them the study consent form (Appendix E). Participants' level of understanding was then assessed through having them paraphrase major points and answering questions quizzing their understanding of the consent form. As outlined by Simon (1994), to reduce the likelihood of confusion and possible malingering among NGBRI participants they were further informed that they were no longer facing legal charges and their results would not be shared with the courts or hospital staff.

NGBRI participants were subsequently administered the MFIT and SILS and were rated on the BPRS-E. As previously mentioned, patients of NGBRI status are considered to have less incentive to malinger impairment and more incentive to look good for purposes of less restrictive placement (Hayes, Hale, & Gouvier, 1998).

While this is a reasonable assumption, it fails to take into account other incentives for malingering (e.g., acquiring a single room, gaining staff attention). Moreover, a shortcoming in previous studies has been a failure to verify (through standardized assessment) the malingering status of those patients assumed to be non-malingering. Thus, NGBRI patients were given additional malingering screening measures (NIM scale, M-Test, and DCT) to rule out the possibility of malingering. All NGBRI subjects were paid two dollars for their participation.

Scoring Procedures. Following the collection of data, the MFIT protocols for all subjects (excluding identifying information other than the study number) were scored along the specified quantitative and qualitative dimensions (see Table 2) by two independent raters (doctoral candidates in clinical psychology with experience in using the MFIT) blind to subjects' diagnoses and group membership. Raters were given the opportunity to become familiarized with the scoring criteria and have questions regarding these criteria clarified prior to rating the protocols. It was noted there were some areas of confusion regarding the scoring criteria for Primacy Effect, Capitalization, Row Perseveration, and Wrong Item errors. This was clarified through joint discussion and the scoring criteria were revised (see italicized print in Table 2). After the ratings were completed, the agreement between raters was assessed through intraclass correlation coefficients (ICC's) and kappa agreement, where appropriate.

As a check for accuracy, interrater reliability (ICC and kappa) was also calculated for the data derived from hospital records. This was accomplished through selecting 20 subjects at random and having an independent rater (clinical psychology graduate student not affiliated with the study) review the subjects' hospital records and complete the demographics form.

Table 2

Scoring Procedures for the MFIT

Quantitative: Total number of items correctly recalled (regardless of location).

Qualitative: (scored for the total number of occurrences for each of these types, with the exception of Primacy Effect, Capitalization, Roman Numeral, and Gestalt).

Primacy Effect -	the total number of items recalled from the first two rows of the stimulus card, *regardless of location*.
Item Perseveration -	repetition of an individual item (e.g., A A).
Row Perseveration -	repetition of an entire row (*differentiate from Capitalization error*).
Row Extension -	continuing the elements of a row (e.g., A B C D E).
Reversal -	reversing a symbol (e.g., "b" drawn as "d").
Rotation -	rotating a correctly drawn figure by more than 30 degrees.
Capitalization -	drawing the row of small cap letters in large caps (*must be in the correct location, i.e. third row*). Not counted in total score.
Roman Numeral -	drawing the tally marks as Roman numerals. Counted in total score.
Within Row Error -	misordering the elements within a row (e.g., B A C).
Row Sequence Error-	misordering of the rows (e.g., 1 2 3 before A B C).
Between Row Error -	rearranging the items from two or more rows (e.g., A 2 B).
Wrong Item Error -	number of produced recognizable figures not on stimulus card (*differentiate from row extension*).
Extraneous Element -	number of indistinguishable or bizarre figures not on the stimulus card (differentiate from row extension and wrong item error).
Gestalt Error -	failure to reproduce a 3 x 5 configuration when 15 items are present
Embellishment -	adornment or elaboration of recognizable characters (e.g., smiley face drawn in the circle); also include elaborate drawings in place of the stimulus material

Results

Interrater Reliability

MFIT Scoring Indices. The reliability between raters' scoring of all subjects' MFIT protocols was computed using primarily ICC. Kappa coefficients (percent agreement corrected for chance) were used for the three qualitative variables scored for their presence or absence (i.e., Capitalization, Roman Numeral, and Gestalt) (Table 3).

Table 3

Interrater Reliabilities for the Quantitative and Qualitative MFIT Scoring Indices

Scoring Index	ICC	Kappa
Total Correct (quantitative)	.97	---
Primacy	.95	---
Item Perseveration	.86	---
Row Perseveration	.94	---
Row Extension	.90	---
Reversal	1.00	---
Rotation	1.00	---
Within Row Error	.91	---
Between Row Error	.84	---
Row Sequence Error	.81	---
Wrong Item	.85	---
Extraneous Element	.98	---
Embellishment	.86	---
Roman Numeral	---	1.00
Capitalization	---	.95
Gestalt	---	*

Note. * Unable to compute kappa coefficient due to the low frequency of occurrence.

Overall, the interrater reliability for the quantitative and qualitative scoring indices was found to be at an acceptably high level (>.80), with more than half of the indices (56%) over .90. Since such a high level of reliability was obtained, one rater was randomly selected and her scores used for statistical analyses. Further, it was noted the "Gestalt" error previously identified by Griffin et al. (1996) was found to have occurred only once out of the 106 protocols. This qualitative score was therefore eliminated from further analyses.

Hospital Records Review. To ensure the information from hospital records was obtained accurately and reliably, 20 subjects were randomly selected to have their records reviewed by an independent rater. This information was entered onto a demographics form and subjected to analysis for interrater reliability through a combination of ICC and kappa. The average reliability for all items was .895 (range of .80 to 1.00), which is at an acceptable level of reliability.

Verification of Group Membership

The following is a summary of the information that went into the treatment teams' clinical decisions regarding patients' malingering status. This was done to better determine the accuracy of diagnostic decisions and estimate the probability of malingering among the non-malingering samples.

Diagnosed Malingerers. Inclusion in the DM group was based upon a diagnosis of malingering by the hospital interdisciplinary treatment teams. Diagnoses were established primarily on the basis of behavioral observations by multiple staff and in various situations (e.g., during initial psychiatric and psychological mental status and competency evaluations, interactions on and off the unit). The most frequently observed behavior was inconsistency in either the patient's presentation or

report of symptoms, as noted in 70.5% of cases. This was followed by an unusually impaired or improbably poor performance on initial examinations (65.9%) and the presentation of atypical or rare symptoms (50.0%). Less common were presentations involving unusually or extremely severe symptoms (29.6%), vague descriptions of symptoms (22.7%), highly inconsistent demonstration of abilities (18.2%) and improbable or absurd symptoms (15.9%). Furthermore, most diagnosed malingerers were observed to engage in a combination of at least three separate behaviors believed to be indicative of malingering (Rogers, Gillis, Dickens, & Bagby, 1991). Based on these observations, staff had determined that 6 patients were malingering cognitive impairment, 4 were malingering psychosis, and 34 were malingering elements of both.

In the majority of cases the results of a formal psychological (malingering) evaluation were used to confirm observations. Of the 44 patients in the DM group, 41 had been administered at least one malingering measure (primarily the SIRS, M-Test, or DCT) in addition to the MFIT. Among the three patients who did not receive additional assessment, one was highly uncooperative with efforts and had been diagnosed as malingering on a previous admission. The other two had been diagnosed on the basis of their symptom presentations, which were noted to be of a rather extreme nature.

Of the 41 DM patients formally assessed, 37 (90.2%) met criteria for malingering on at least one measure other than the MFIT. Specifically, 22 of the 30 patients given the SIRS were determined to be malingering elements of psychosis. Among the 8 patients who "passed" the SIRS, 6 scored in the malingering range on the M-Test, DCT, or both, as did 9 of the 11 patients who did not receive the SIRS. Of interest is the finding that, among the four DM patients who did not score in the

malingering range on any additional measure, three obtained extremely low MFIT total scores (2, 3, and 4) and one scored at the cutoff. Looking at historical factors, nine patients had a documented preexisting psychiatric condition involving psychosis (i.e., through previous hospitalizations, psychiatric diagnosis, and/or past treatment with neuroleptics). Another three patients had a highly probable history of psychosis (through independent reports but no available confirming records), and one had a documented history of mental retardation. None of the DM had a documented or confirmed neurological condition (e.g., through electroencephalogram or CAT scan) and only two had a history of traumatic brain injury (both noted to present with impairments inconsistent with either the extent or location of injury). Thus, a high percentage (70.5%) of patients diagnosed as malingering were presenting with problems involving psychosis, cognitive impairment, or both but without any prior history of these conditions. However, almost one third of the diagnosed malingerers had experience with these conditions and therefore may have been more sophisticated (i.e., less blatant) in their presentations.

Finally, presence of an antisocial personality disorder (in conjunction with other factors) is one of the criteria used for making a diagnosis of malingering (APA, 1994). Among the DM group, 14 (31.8%) met diagnostic criteria for antisocial personality disorder. A chi-square analysis revealed this number to be significantly different (p=.025) than the number of PT patients diagnosed with antisocial personality disorder (6.3%).

Pretrial (non-malingerers). Although patients in this group were selected on the basis they had not been suspected of or diagnosed as malingering, this did not rule out the possibility that individuals in this group were malingering. It was therefore

decided to look at other available information to better assess this possibility. In terms of diagnoses, the primary Axis I diagnosis for this group was schizophrenia (most in conjunction with a substance abuse diagnosis), and the primary Axis II diagnosis was mental retardation (see Table 4). Among the 32 PT patients, only 10 (31.3%) had no prior documented history of either psychosis or mental retardation. A chi-square revealed this number to be significantly different from the number of DM patients with no prior history, $\underline{X}^2$ (4, $\underline{N}$=76) = 11.092, $\underline{p}$=.001.

Table 4

Primary Axis I and Axis II Diagnoses for the Two Non-Malingering Groups

Diagnosis	PT ($\underline{N}$=32)	NGBRI ($\underline{N}$=30)
AXIS I		
Schizophrenia	14 (43.8%)	21 (70.0%)
Substance Abuse	6 (18.8%)	2 (6.7%)
Major Depression	5 (15.6%)	3 (10.0%)
Substance-Induced Dementia	3 (9.4%)	0 (0.0%)
Organic Mental Disorder	2 (6.3%)	1 (3.3%)
None	1 (3.1%)	0 (0.0%)
Other	1 (3.1%)	3 (10.0%)
AXIS II		
Mental Retardation (Mild)	15 (46.9%)	3 (10.0%)
Borderline Intellect	7 (21.9%)	4 (13.3%)

Among the reasons for administration of the MFIT, 12 patients were given the MFIT as part of a comprehensive neuropsychological evaluation involving potential memory problems. Another 13 were given the MFIT as part of an overall diagnostic assessment, one in which the patient was presenting with acute and severe symptoms (typically disorganization). Staff had determined that four of these patients were experiencing a first episode of psychosis. The MFIT was given as a check for level of

effort on competency evaluation and personality/intellectual assessment in six cases. Only one patient received the MFIT as part of a comprehensive malingering evaluation, although he had never been suspected of actual malingering by the hospital staff (i.e., one sanity commission member was uncertain and suggested the possibility be ruled out through assessment).

Despite the absence of true suspicion of malingering, 20 PT patients had been given the SIRS, M-Test, or DCT, either alone or in combination. This is common practice within forensic settings, given the legal implications and potential serious nature of some patients' charges (e.g., death penalty in First Degree Murder cases). None of the 13 patients given the SIRS were found to be malingering on this measure, neither were the 5 patients who received only the M-Test or the 2 patients who only received the DCT. Of the PT patients who received the M-Test (N=14), only one scored in the malingering range when using the original criteria of Beaber et al. (1985), whereas two would have fallen in this range using the revised criteria of Rogers, Bagby, and Gillis (1992). Using Rey's original scoring criteria (Lezak, 1983), three patients scored in the malingering range on the DCT, whereas two would have done so using the criteria of Binks et al. (1997). In none of these cases did an individual who scored in the malingering range on one test do so on a second test.

Examining hospital staffs' behavioral observations of PT patients, only three behavioral indicators suggestive of malingering (Rogers et al., 1991) were observed with any degree of regularity. The most frequently observed behavior was unusually impaired or improbably poor performance on initial examinations by seven patients (21.9%). This was followed by inconsistency in the presentation or report of symptoms (12.5%) and unusually severe or extreme symptoms (9.4%). However,

staff indicated these behaviors were most likely due to either acute psychosis or impaired intellectual functioning, which combined with the results of standardized assessment would indicate the probability of malingering to be rather low in this sample.

NGBRI Controls. To confirm the non-malingering status of the NGBRI group, participants were given additional malingering screening measures including the M-Test, DCT, and Negative Impression Management (NIM) scale of the PAI. It was found that no one scored above the suggested cutoff for malingering on the NIM (M=4.43, SD=3.58, range = 0 to12). Only one NGBRI participant scored in the malingering range using the original DCT criteria, whereas two fell in the malingering range when using the criteria of Binks et al. (1997). On the M-Test, six patients scored in the malingering range when using the original criteria, a number that doubled when using the revised and improved criteria of Rogers, Bagby, and Gillis (1992). The poor performance on this measure was not entirely surprising, given the aforementioned problems with variable rates of sensitivity and specificity for this measure (Smith, 1997). Additionally, Hankins, Barnard, and Robbins (1993) noted that one problem with the M-Test was that it is essentially a cognitive task, and found that the cognitive status (level of cognitive impairment) of the individual could significantly impact their ability to engage in the task. This finding was confirmed in the current sample through significant correlations ($p<.05$) between BPRS-E items reflecting cognitive status (i.e., disorientation and conceptual disorganization) and the total M (malingering) scale items ($r=.315$ and .338, respectively).

These results appear to verify the non-malingering status of patients in the NGBRI group. Only one NGBRI patient scored in the malingering range on more

than one measure (M-Test or DCT). However, this was the case only when the revised M-Test scoring criteria were used, which in turn appear to have less specificity than the original scoring criteria, at least for this sample. Moreover, this patient had a longstanding and documented history of mental health problems involving both mental retardation and schizophrenia (i.e., five previous hospitalizations), and had recently obtained a WAIS-3 IQ of 55. It was therefore decided his performance was most likely attributable to genuine psychiatric problems and low IQ and he was retained for the study. A breakdown of patients' malingering classification on each of the three primary malingering measures is provided in Tables 5 to 7.

Table 5

Malingering Classification of DM and PT Patients Using the SIRS

Classification	DM (N=30)	PT (N=13)
Malingering	22 (73.3%)	0 (0%)
Not Malingering	8 (26.7%)	13 (100%)

Table 6

Malingering Classification of DM, PT, and NGBRI Patients Using the DCT

Classification	DM (N=23)	PT (N=12)	NGBRI (N=30)
(Grouped Time > Ungrouped Time)			
Malingering	13 (56.5%)	3 (25.0%)	1 (3.3%)
Not Malingering	10 (43.5%)	9 (75.0%)	29 (96.7%)
(Total Number of Errors)			
Malingering	2 (8.7%)	2 (16.7%)	2 (6.7%)
Not Malingering	21 (91.3%)	10 (83.3%)	28 (93.3%)

Table 7

Malingering Classification of DM, PT, and NGBRI Patients Using the M-Test

Classification	DM (N=36)	PT (N=14)	NGBRI (N=30)
(Original Criteria)[a]			
Malingering	26 (72.2%)	1 (7.1%)	6 (20.0%)
Not Malingering	10 (27.8%)	13 (92.9%)	24 (80.0%)
(Revised Criteria)[b]			
Malingering	29 (80.6%)	2 (14.3%)	12 (40.0%)
Not Malingering	7 (19.4%)	12 (85.7%)	18 (60.0%)

Note: [a] Original criteria of Beaber et al. (1985), with malingering indicated by a score of $\geq$ 4 positively endorsed M-scale items. [b] Revised criteria of Rogers, Bagby, & Gillis (1992), with malingering indicated by a score of < 4 on selected Rule Out items and > 2 on selected Rule In items.

Demographic Data Analysis

A one-way analysis of variance (ANOVA) was used to examine group differences on the demographic variables of age and education. There was a significant difference between groups for age ($F[2,103] = 4.486$, $p = .014$), and a post hoc test for Least Significant Differences (LSD) revealed the NGBRI control group to be significantly older than both the PT and DM patients (both p's <.05). There was also a significant group difference for years of education ($p = .021$), which a post hoc LSD test revealed to be attributable to patients in the DM group having fewer years of education than the PT group ($p = .007$). A chi-square analysis did not reveal any

differences between the three groups in terms of racial composition, $\underline{X}^2$ (2, $\underline{N}$=106) = 2.795, $\underline{p}$ = .247 (Table 8).

Table 8

Demographic Variables by Group

Demographic Variable	DM (N=44)	PT (N=32)	NGBRI (N=30)
Age	31.55[a] (10.55)	32.94[b] (11.02)	38.63[a b] (8.91)
Education	8.14[a] (2.61)	9.78[a] (2.55)	9.20 (2.55)
Ethnicity			
Caucasian	5	8	7
African-American	39	24	23

Note: [a b] Denotes groups which significantly differed at the .05 level.

That NGBRI controls were significantly older was not unexpected, as this legal status is not typically obtained until after competency restoration efforts have been completed and the individual has been through the process of trial. However, previous findings of a significant correlation between age and the total number of MFIT items recalled (Griffin et al., 1996; Hays et al., 1993) indicates the need to co-vary for the effects of age in subsequent analyses. With regard to education, of primary importance is the relative equivalence of the two non-malingering groups (PT and NGBRI) on this variable. More precisely, it would be difficult to ascertain the true effects of educational level on the MFIT performance of the DM group since their scores were of dubious validity. Also, the educational level of this group was of

questionable validity as it was often obtained through self-report pursuant to difficulty in procuring school records.

In examining the relevance of comparison groups (Rogers & Cruise, 1998), there were no significant differences between groups on those variables reflecting criminal background. Specifically, one-way ANOVAs revealed no significant differences between DM, PT, and NGBRI patients in the total number of arrests prior to hospitalization ($F[2, 103] = .581$, $p = .561$), total number of felony convictions ($F[2, 103] = .928$, $p = .399$), or total number of misdemeanor convictions ($F[2, 103] = 1.249$, $p = 2.91$) (Table 9).

Table 9

Mean Number of Prior Arrests, Felony, and Misdemeanor Convictions by Group

Historical Variable	DM (N=44)	PT (N=32)	NGBRI (N=30)
Total Prior Arrests	9.43 (10.66)	7.56 (9.82)	7.27 (7.16)
Total Felony Convictions	1.02 (1.37)	0.53 (1.14)	0.93 (2.23)
Total Misdemeanor Convictions	1.07 (1.77)	1.78 (3.70)	0.83 (1.68)

The three comparison groups were found to be relatively homogenous in terms of the primary (i.e., more serious) commitment charge, as indicated by a non-significant chi-square, X^2 (12, N=106) = 13.566, p=.329 (Table 10). Of interest is the finding that approximately half of all patients had been involuntarily committed for violent offenses against person (e.g., murder, aggravated rape, aggravated assault and battery). These types of charges could potentially result in severe punishment (e.g., life imprisonment and death penalty), which presumably would provide stronger incentives to malinger (Schacter, 1986a).

Table 10

Percentage of Patients in Each Group by Primary Commitment Charge

Charge	DM (N=44)	PT (N=32)	NGBRI (N=30)
Murder	20.5%	21.9%	16.7%
Attempted Murder	13.6%	18.8%	6.7%
Other Violent Crimes Against Person	18.2%	6.3%	33.3%
Non-Violent Crimes Against Person	6.8%	3.1%	6.7%
Aggravated Theft/Property Crime	20.5%	18.8%	26.7%
Minor Theft-Related Crimes	9.1%	9.4%	6.7%
Drug-Related Offenses	11.4%	21.9%	3.3%

On the basis of these results, it would appear the three comparison groups met the criteria of Rogers and Cruise (1998) for relevance in that they possessed similar criminal backgrounds.

Quantitative Scoring

The original (quantitative) scoring method for the MFIT involved totaling the number of items correctly recalled, with fewer than nine items recalled being used as the cutoff for raising the suspicion of malingering (Lezak, 1976). Applying this criteria to patients' MFIT protocols, approximately 21 of the 44 DM patients would have been correctly suspected of malingering, providing a sensitivity of 47.7% and a false negative rate of 52.3%. For the two non-malingering groups, a high false positive rate of 43.8% was obtained for PT patients, although a more encouraging false positive rate of 13.3% was found for NGBRI patients. It was also found that the average number of items recalled by DM patients fell below the 9-item cutoff (M=8.27, SD=4.05), while that for the non-malingering PT (M=9.78, SD=3.02) and NGBRI patients (M=11.90, SD=3.18) fell above the cutoff. Using a one-way

ANOVA, a significant between-groups difference was found in mean total scores ($p<.001$), which a post hoc LSD test revealed to be attributable to NGBRI patients scoring higher than both PT patients ($p=.020$) and DM patients ($p<.001$). These results were virtually unchanged when using an analysis of covariance (ANCOVA) to control for the effects of age ($p<.001$). A breakdown of the number of patients in each group by quantitative scores is provided in Table 11.

Table 11

Number of Patients in Each Group by Quantitative MFIT Scores

Total Items Recalled	DM (N=44)	PT (N=32)	NGBRI (N=30)
2	3 (6.8)	0 (0)	0 (0)
3	4 (15.9)	0 (0)	0 (0)
4	2 (20.5)	0 (0)	1 (3.3)
5	3 (27.3)	1 (3.1)	1 (6.7)
6	5 (38.6)	3 (12.5)	0 (6.7)
7	2 (43.2)	6 (31.3)	2 (13.3)
8	2 (47.7)	4 (43.8)	0 (13.3)
9	9 (68.2)	2 (50.0)	3 (23.3)
10	3 (75.0)	2 (56.3)	1 (26.7)
11	2 (79.6)	4 (68.8)	2 (33.3)
12	1 (81.8)	5 (84.4)	7 (56.7)
13	0 (81.8)	0 (84.4)	0 (56.7)
14	2 (86.4)	1 (87.5)	4 (70.0)
15	6 (100)	4 (100)	9 (100)

Note. Cumulative percentage in parentheses.

Hypothesis 1. In the first hypothesis it was predicted that non-malingering patients would obtain significantly higher total MFIT scores than malingering patients. While the finding of a significant difference in quantitative scores between NGBRI

and DM patients lends partial support to this hypothesis, that PT and DM patients did not significantly differ was contrary to expectations. This is of greater concern as the PT and DM patients are most likely to receive the MFIT in clinical settings.

In terms of validity (sensitivity/false negatives > false positives/specificity) and effectiveness (base rate > false positives + false negatives or 1- base rate > false positives + false negatives), the traditional quantitative scoring method was found to be valid for differentiating both PT (.477/.523 > .438/.562) and NGBRI patients (.477/.523 > .133/.867) from DM. However, this method did not appear to be an effective indicator for malingering with this sample. Specifically, the combined error rates for PT and DM (.303+.184) and NGBRI and DM (.311+.054) were in excess of the estimated malingering base rate among criminal defendants (16%; Rogers et al., 1998). These results attest to the need to examine alternative scoring methods that might improve the efficacy of the MFIT.

Qualitative Scoring

Contrary to the findings of Griffin et al. (1996), a relatively high base rate for qualitative errors was found among all three patient groups (Table 12). Among the DM group, only 25% made but one type of qualitative error, and over half (52.3%) made two or more distinct types of errors. Among the non-malingering PT and NGBRI groups these rates were 46.9% and 36.7% for one qualitative error type, and 40.6% and 46.7% for two or more qualitative error types, respectively. A one-way ANOVA revealed no significant differences between these groups in terms of the average number of distinct error types ($F[2, 103] = .599$, $p = .551$) or the total number of all errors made ($F[2, 103] = .242$, $p = .786$).

Table 12

Number of Patients in Each Group by the Number of Distinct Qualitative Error Types

Number of Distinct Error Types	Group DM (N=44)	PT (N=32)	NGBRI (N=30)
0	10 (22.7%)	4 (12.5%)	5 (16.7%)
1	11 (25.0%)	15 (46.9%)	11 (36.7%)
2	16 (36.4%)	4 (12.5%)	6 (20.0%)
3	5 (11.4%)	3 (9.4%)	7 (23.3%)
4	1 (2.3%)	4 (12.5%)	1 (3.3%)
5	1 (2.3%)	1 (3.1%)	0
6	0	1 (3.1%)	0

This would indicate that malingering forensic patients did not make more qualitative errors overall than their non-malingering counterparts. As such, the focus is now shifted to examining group differences in the specific types of qualitative errors made. A breakdown of the number of patients within each group making each of the identified types of qualitative errors is provided in Table 13.

Table 13

Number of Patients in Comparison Groups Making Each Qualitative Error

Qualitative Error	Group DM (N=44)	PT (N=32)	NGBRI (N=30)	p
Item Perseveration	8 (18.2%)	6 (18.8%)	3 (10.0%)	.566
Row Perseveration	6 (13.6%)	1 (3.1%)	4 (13.3%)	.273
Roman Numeral	3 (6.8%)[a]	1 (3.1%)[b]	7 (23.3%)[a b]	.020
Capitalization	2 (4.6%)[a]	7 (21.9%)[a]	4 (13.3%)	.074
Within Row	14 (31.8%)	11 (34.4%)	9 (30.0%)	.933
Between Row	9 (20.5%)	2 (6.3%)	2 (6.7%)	.096
Row Sequence	7 (15.9%)	7 (21.9%)	10 (33.3%)	.211
Row Extension	2 (4.6%)	1 (3.1%)	2 (6.7%)	.804
Wrong Item	9 (20.5%)	12 (37.5%)[a]	4 (13.3%)[a]	.066
Extraneous Element	3 (6.8%)	6 (18.8%)	5 (16.7%)	.254
Embellishment	4 (9.1%)	3 (9.4%)	4 (13.3%)	.821
Reversal	1 (2.3%)	1 (3.1%)	0 (0%)	.645
Rotation	3 (6.8%)	2 (6.3%)	0 (0%)	.353

Note. [a] and [b] denote groups differing from each other at the .05 level of significance using Fisher's Exact Test of Significance. p represents the overall significance level for chi-square.

Hypothesis 2. The second study hypothesis was that non-malingering patients would make more qualitative errors involving perseveration, misordering of elements, character reversals and rotations than malingering patients. This expectation was based upon previous findings of these particular error patterns among the mentally retarded and psychiatric patients (Goldberg & Miller, 1986; Greiffenstein et al., 1996; Morgan, 1991). This hypothesis was not supported by the present results, as there

were no significant group differences (using chi-square and Fisher's Exact Test) for errors involving perseveration (Item and Row Perseveration), misordering of elements (Within Row, Between Row, and Row Sequence) or rotations and reversals of items (Rotation and Reversal). In fact, it was found that very few non-malingering patients made errors involving rotations and reversals, and there was a non-significant trend for malingerers to make more errors involving misordering of elements between rows (Between Row error) than non-malingering patients. The most frequently occurring error among all groups involved a misordering of items within a row (Within Row error).

Hypothesis 3. In the third hypothesis it was expected that DM patients would make more errors involving bizarre and unusual reproductions and addition of extraneous elements than PT and NGBRI patients (i.e., Griffin et al., 1996). Rather, a chi-square analysis revealed these groups to be relatively equal in the number of Embellishment (p=.821) and Extraneous Element errors (p=.254) (refer back to Table 13). In conjunction with the results from the second hypothesis, it would appear that malingerers were just as likely to make those errors typically associated with genuine psychiatric and intellectual impairment, while non-malingering psychiatric patients were just as likely to make those errors believed to be indicative of malingering.

Hypothesis 4. The fourth hypothesis was that malingerers would recall fewer items from the first 2 rows of the MFIT, i.e., fail to exhibit a primacy effect in their recall. This was based upon the findings of Morgan (1991), that even individuals with severe organic memory problems could recall these items and hence failure to do so should suggest malingering. It was found that DM patients recalled an average of 4.30 (SD=2.12) items from the first 2 rows, as compared to 5.28 (SD=1.25) for PT patients

and 5.27 (SD=1.26) for NGBRI patients. Using a one-way ANOVA, the overall difference between groups was found to be significant (F[2, 103]=4.417, p=.014). A post hoc LSD test revealed that both PT and NGBRI recalled more items than DM (p=.013 and .016, respectively), and that PT and NGBRI did not significantly differ in their recall of these items (p>.90). A breakdown of the number of patients in each group by the number of items recalled from the first 2 rows is provided below (Table 14).

Table 14

Primacy Effect (Items Recalled from First 2 Rows) by Group Membership

Number of Items Recalled	DM (N=44)	PT (N=32)	NGBRI (N=30)
0	5 (11.4)	0 (0)	0 (0)
1	1 (13.6)	0 (0)	0 (0)
2	1 (15.9)	1 (3.1)	0 (0)
3	10 (38.6)	5 (18.8)	6 (20.0)
4	2 (43.2)	0 (18.8)	2 (26.7)
5	2 (47.7)	4 (31.3)	0 (26.7)
6	23 (100)	22 (100)	22 (100)

Note. Cumulative percentage in parentheses.

The majority of all patients recalled more than three items from the first 2 rows. Only one non-malingering patient (PT) recalled fewer than three items, compared to 7 (15.9%) of the malingerers. In fact, a closer inspection of the data revealed that the PT patient who recalled only two primacy items still had a quantitative score above the cutoff, whereas all of the DM patients exhibiting a primacy score of two or less fell below the 9-item cutoff. Furthermore, 60% of the PT

and 66.7% of the NGBRI patients who had primacy scores of three produced quantitative scores above the 9-item cutoff, as compared to only 20% of the malingerers who had primacy scores of three. These findings suggest that failure to recall three or more items from the first 2 rows is highly indicative of malingering, particularly when combined with failure to recall 9 or more total items. Furthermore, it is noted that failure to recall six items from the first 2 rows resulted in improved classification over the total quantitative score for the PT group (specificity of 68.7%), whereas the classification (sensitivity) for DM remained unchanged. This would indicate that one would do better by simply calculating primacy scores.

Before proceeding to the final study hypothesis, it is noted significant between-groups differences occurred on those qualitative variables (Roman Numeral, Capitalization, and Wrong Item) for which there was no a priori hypotheses. No hypotheses were offered for these variables for two reasons. First, these particular errors had only been identified and examined through one previous investigation (Griffin et al., 1996), with the resulting conclusion that both reflected "artifacts of illness." This was because one error (Roman Numeral) was found significantly more among normal controls and the other (Wrong Item) significantly more among the severely ill. Consistent with their conclusions, it was found that significantly more NGBRI patients made Roman Numeral errors than both PT and DM (Fisher's Exact Test, $\underline{p} < .05$), and significantly more PT patients made Wrong Item errors than NGBRI ($\underline{p} < .05$) (refer to Table 13).

Second, there appears to be a lack of consensus that the Capitalization error should be counted as an actual error. This was evident in discussions with the two raters for this study as well as consultations with four clinical psychologists from

forensic settings. It was found that significantly more PT patients made this error than DM patients (Fisher's Exact Test, $p < .05$). Based upon these factors, it was decided to recalculate the quantitative scores for all patients to include those items previously counted wrong due to a Capitalization error (refer back to Table 2). Inclusion of these items increased the total (quantitative) scores to above the 9-item cutoff for one NGBRI, four PT, and no DM patients. This resulted in improved false positives and specificity for both PT (43.8/56.2 to 31.3/68.7) and NGBRI (13.3/86.7 to 10.0/90.0), with no change in the sensitivity and false negatives for DM (47.7/52.3). However, the combined false positive and false negative error rates for PT and DM (43.4%) and NGBRI and DM (35.1%) still exceeded the established base rate for malingering (16%).

Hypothesis 5. In the final hypothesis it was predicted that a combination of both quantitative and qualitative variables would provide better discrimination between malingerers and non-malingerers than the quantitative score alone. This was demonstrated to some extent in the last section, where considering the Primacy Effect and total score together strengthened suspicions about malingering and where correcting the total score for Capitalization errors resulted in improved classification for non-malingerers. Another method for exploring this hypothesis is through the use of linear discriminant analysis (LDA), a statistical procedure that allows one to determine the best combination of variables for differentiating between two or more groups.

The independent (predictor) variables for LDA were derived as follows. First, since the total score yielded better discrimination among groups after being corrected for Capitalization errors, the corrected total score was entered as the quantitative

variable. Since Item and Row Perseveration errors are essentially measures of the same phenomenon (i.e., repetition), they were combined into a single variable (Perseverative errors). This was also done for Reversal and Rotation errors, as these are both errors involving the spatial rotation of items (letters and objects, respectively). Age was entered as an independent variable to control for its effects on scores. The Roman Numeral and Wrong Item errors were omitted from this analysis since there were no a priori hypotheses for these variables and they appeared to contribute more to differentiation between the two non-malingering groups (i.e., "artifacts of illness"). This resulted in 11 independent variables (Table 15), which fits within the recommended criteria for LDA of having 10 subjects per independent variable (Pedhazur, 1982). PT and NGBRI patients were grouped together to form a single "non-malingering" (NM) group (N=62), with group status (DM vs. NM) as the dependent variable.

Table 15

Independent Variables Used in the Linear Discriminant Analysis

Variable	Wilks' Lambda	Level of Significance	Canonical Correlation
Corrected Total Score	.868	.000	.636
Primacy Effect	.952	.024	.366
Age	.962	.047	.322
Between Rows	.962	.046	-.323
Perseveration	.977	.123	-.248
Row Sequence	.981	.163	.224
Extraneous Element	.981	.157	.228
Rotations	.992	.368	-.144
Embellishment	.999	.728	.056
Within Row	1.000	.834	.034
Row Extension	1.000	.835	.033

In the initial step of LDA, it was found that the Wilks' lambda (ratio of the within-groups sum of squares to the total sum of squares) for the Corrected Total Score, Primacy Effect, Age, and Between Rows error were all significant using univariate F-tests (all p's<.05). This indicates that the proportion of variance in these scores attributable to differences between malingerers and non-malingerers was significant. However, an inspection of the canonical correlations (correlation between predictor variable and the resulting discriminant function) revealed that only four of the nine predictor variables for which there was an a prior hypothesis went in the expected direction. Specifically, higher scores on those variables with negative canonical correlations were more associated with malingering, whereas higher scores on variables with positive correlations were associated with non-malingering. Thus, contrary to hypotheses the presence of Between Rows, Perseveration, and Rotation errors were associated more with malingering and Extraneous Element and Embellishment errors were associated more with non-malingering.

Next, the discriminative ability of all independent variables (that were not linear combinations of the other independent variables) was determined through a forced-entry LDA with a minimum entry tolerance criteria of .001. This method was chosen as it allows for the maximum number of variables to be examined, given the information provided by a variable about group discrimination has not already been supplied by the other variables in the equation (Norusis, 1988). It was found that all of the variables in Table 15 met minimum tolerance criteria and thus were used in the analysis. The resulting Wilks' lambda for the discriminant function was .726, indicating the function accounted for approximately 27.4% of the variance between

groups. This was found to be statistically significant, X^2 (11, N=106) = 31.490, p=.001, and resulted in 79.2% of all patients being correctly classified (Table 16).

Table 16

Classification Results from the Linear Discriminant Analysis Using Forced-Entry

	Actual Group	
Predicted Group	DM (N=44)	NM (N=62)
DM	30 (68.2%)	8 (12.9%)
NM	14 (31.8%)	54 (87.1%)

The overall rate for correct classification was substantially improved over that obtained through both the quantitative (61.3%) and the (Capitalization) corrected quantitative method (66.0%). A substantial increase was found in sensitivity (47.7 to 68.2%), and closer inspection of the data revealed that only 5 of the 32 PT patients (15.6%) were incorrectly classified.

Although this is encouraging, replication is necessary to demonstrate the validity and reliability of this method. Without the benefit of a second sample for cross-validation, this was accomplished through the "leave one out" method. In this method, each case in the sample is classified (malingering or non-malingering) using the discriminant function that is derived from the entire sample minus that particular case. It was found that the function using all of the variables did not cross-validate well, as only 62.3% of the cases were correctly classified (Table 17). This falls back within the range of the quantitative and corrected quantitative methods.

Table 17

Classification Resulting from Cross-Validation of the Forced-Entry LDA

Predicted Group	Actual Group DM (N=44)	NM (N=62)
DM	23 (52.3%)	19 (30.6%)
NM	21 (47.7%)	43 (69.4%)

An alternative approach was to use only those independent variables making the largest (and statistically significant) contribution to differentiating the two groups. This was done through an LDA using a stepwise entry method, with minimization of Wilks' lambda as the entry criterion. This resulted in a Wilks' lambda of .748 for the overall function, which was significant, X^2 (4, N=106) = 29.637, p=.000, and accounted for 25.2% of the variance between groups. Only four variables were retained for the function, which in order of the size of their pooled within-groups correlation with the discriminant function were the total corrected score, Between Rows error, age, and Extraneous Element. The only variable with a negative canonical discrimination function coefficient (thus higher scores associated with malingering) was the Between Rows error.

Use of this method resulted in 72.6% of the original cases being classified correctly (56.8% of malingerers and 83.9% of non-malingerers) (Table 18). More importantly, these results cross-validated when using the leave one out procedure, with 71.7% of the cases still being correctly classified. It was noted that only one patient (NM) was re-classified incorrectly.

Table 18

Classification Table for the Stepwise LDA and Cross-Validation

Predicted Group	Actual Group DM (N=44)		NM (N=62)	
DM	25 (56.8%)	*25 (56.8%)*	10 (16.1%)	*11 (17.7%)*
NM	19 (43.2%)	*19 (43.2%)*	52 (83.9%)	*51 (82.3%)*

Note. Numbers in italicized print represent cross-validated classifications.

The combined use of quantitative and qualitative scores resulted in improved classification, particularly for the two groups of greater clinical relevance (PT and DM). Although this method met statistical criteria for validity (.568/.432 > .177/.823), the combined error rate (27.4 %) remained above the estimated base rate for malingering (16%). Thus, the combined method did not make the MFIT better than base rate prediction alone.

Supplemental Analyses

Psychiatric Symptoms. The BPRS-E (Ventura et al., 1993) was given to all NGBRI participants to obtain ratings of psychiatric symptoms. Since none of the PT patients had been given this measure during hospitalization, symptom analyses were restricted to those patients in the NGBRI group (N=30). A series of one-tailed Pearson correlations were calculated between BPRS-E items and the MFIT scoring indices. It was found that only three symptoms had a moderate and significant correlation with total MFIT (quantitative) scores, and all three were psychotic symptoms that can affect the ability to attend, concentrate, and organize cognitive

processes. These symptoms were hallucinations (r = -.422, p = .01), disorientation (r= -.428, p=.009), and conceptual disorganization (r = -.468, p=.005).

The only psychiatric symptoms significantly correlated with the total number of qualitative errors were elevated mood (r=.429, p=.009) and flattened affect (r=.375, p=.021). Only one qualitative error (Wrong Item) was significantly associated with psychiatric symptoms, those being hallucinations (r=.552), disorientation (r=.587), and conceptual disorganization (r=.413) (all p's < .01), which appears to be consistent with the speculations of Griffin et al. (1996) regarding "artifact of illness." However, little can be made of this or any other of these findings (or lack thereof) between the individual qualitative errors and psychiatric symptoms, given the low number of patients making these errors (i.e., less than one-third; refer back to Table 13). Nonetheless, these results seem to indicate that psychotic symptoms which affect one's ability to attend, concentrate, and organize cognitive processes contribute to lower MFIT recall.

Intelligence. Low intelligence has been significantly associated with decreased MFIT recall (e.g., Hays et al., 1993). Fortunately, IQ estimates were available for 31 of the PT patients in this study. Looking back to Table 4, a larger proportion of PT patients had been diagnosed with either mental retardation or borderline intellectual functioning as compared to the NGBRI sample. A closer inspection reveals that it was predominately these PT patients (10 out of 14, 71.4%) who had been incorrectly classified based upon their quantitative scores. These diagnoses were confirmed through the finding that the average IQ for PT patients was 70.06 (SD=7.97, range = 54 to 84), falling just on the border of mild mental retardation. A one-way ANOVA revealed the average IQ for these patients to be significantly lower than that for

NGBRI patients (p=.016), who had an average IQ in the borderline range (M=75.17, SD=8.03, range = 55 to 91). Not surprisingly, the average IQ for DM patients (N=35) was well within the mental retardation range (M=63.26, SD=8.33, range = 36 to 80), a finding which supports the notion these patients were malingering problems with intelligence as well as psychosis.

Consistent with the findings from previous investigations, there was a significant positive one-tailed correlation between IQ and total MFIT scores for both PT (r=.390, p=.015) and NGBRI patients (r=.356, p=.027). A significant positive correlation was also found between IQ and the Primacy Effect for both PT (r=.362, p=.023) and NGBRI patients (r=.313, p=.046). Contrary to the findings of Griffin et al. (1996), there were no significant correlations for either patient group between IQ and the total number of qualitative error types or the total number of qualitative errors (all p's>.10). Although a closer inspection of the relationships between IQ and the individual qualitative errors was again hampered by the low number of patients making individual errors, it was noticed there were no significant correlations between any of these variables for the combined patient sample.

Age. Previous investigations (e.g., Griffin et al., 1996; Hays et al., 1993) reported an inverse correlation between age and total MFIT scores as well as qualitative errors. In the previous section it was found that age provided important information for group discrimination (i.e., increasing age associated with non-malingering status). Consistent with previous findings, age was found to be significantly and inversely correlated with total MFIT scores for both PT (r= -.350, p=.025) and DM patients (r= -.269, p=.038), though not for NGBRI (p>.150). There were no significant correlations for any group between age and any of the qualitative

indices. These findings seem to indicate that the primary effect of age is on recall per se, an effect that may be magnified by the level of psychiatric impairment.

Incentive to Malinger. Finally, it was mentioned that individuals having more at stake (e.g., facing severe punishment) would be more likely to engage in extreme presentations and "overplay the role" (Rogers & Cruise, 1998; Schacter, 1986a). It would therefore seem plausible that individuals facing larger prison sentences (or capitol punishment) would be apt to perform more poorly on the MFIT, in turn providing a contextual variable for interpreting results. Rather, there was not found a significant correlation between maximum possible sentence and total MFIT items recalled for patients who had been diagnosed as malingering (r= -.107, p=.244). This would indicate the need to exercise caution when interpreting results within a given context without considering other sources of information (e.g., behavioral observations).

Discussion

Since 1986, several investigations have established normative data for the MFIT among various patient populations displaying memory problems (e.g., neurological disorders, patients recovering from head injury, psychiatric patients, patients with mental retardation) (Back et al., 1996; Bernard & Fowler, 1990; Goldberg & Miller, 1986; Hays et al., 1993; Morgan, 1991). There have also been a series of investigations using simulation (e.g., Arnett et al., 1995; Bernard, 1990; Bernard et al., 1993; Guilmette et al., 1994; Schretlen et al., 1991), differential prevalence (Griffin et al., 1996; Lee et al., 1992), and known-groups designs (Greiffenstein et al., 1994; Millis & Kler, 1995; Simon, 1994). The general consensus has been that the MFIT is overly sensitive to genuine memory impairment (e.g., Schretlen et al., 1991), and the results of simulation and differential prevalence design studies have indicated low sensitivity to simulated or suspected malingering of memory problems.

A closer inspection of available data revealed this conclusion to be premature. Lezak's (1976) original contention was that only "significantly deteriorated" patients would have difficulty in recalling nine or more items. This was supported by the finding that patients with intracerebral hemorrhage, injuries resulting in severe global cognitive impairment, or severe organic amnesia had the poorest recall on the MFIT. Specificity among patients with neurological conditions was estimated to be approximately 85.8% (combined data from Arnett et al., 1995; Bernard & Fowler, 1988; Greiffenstein et al., 1994; Lee et al., 1992; Millis & Kler, 1995; Morgan, 1991). When a more sound study design was used (known-groups), the overall sensitivity

among civil litigants seeking disability or financial compensation for injuries was 64% (Greiffenstein et al., 1994; Millis & Kler, 1995).

These discoveries offer better support for use of the MFIT as a floor effect screening measure for malingered memory problems. However, there has been a paucity of research with the MFIT among pretrial criminal defendants, a population noted to malinger problems with memory and intellect in addition to psychosis (e.g., Gothard et al., 1995). Only two studies have been conducted so far; both were noted to have restricted sample sizes and only one (Simon, 1994) provided estimates of sensitivity and specificity. Furthermore, several MFIT investigations (Arnett et al., 1995; Goldberg & Miller, 1986; Griffin et al., 1996; Guilmette et al., 1994; Morgan, 1991; Schretlen et al., 1991) indicated that a qualitative scoring approach (magnitude of errors strategy; Rogers et al., 1993) might increase the detection ability of the measure. However, only one study has systematically investigated this approach (Griffin et al., 1996), and was limited through use of a design of questionable ecological validity (differential prevalence).

The present known-groups study examined the validity and effectiveness of both the traditional (quantitative) and qualitative scoring methods for the MFIT among a larger sample of criminal forensic inpatients. This was done through examining the MFIT protocols (administered under normal clinical conditions) from an archival sample of pretrial criminal defendants independently diagnosed as malingering, and an archival sample of pretrial criminal defendants for whom malingering had not been suspected and subsequently ruled out. Since NGBRI status is the goal of the pretrial malingerer, MFIT protocols were also collected from a sample of NGBRI patients for

purposes of comparison. The results of the scoring methods were examined within the context of five study hypotheses, summarized as follows.

Hypothesis 1

The first hypothesis dealt with the traditional quantitative scoring method (9-item cutoff), and it was predicted that malingering patients would recall fewer total items than non-malingering patients. As expected, the average number of items recalled by DM patients fell below the 9-item cutoff whereas the average number for the non-malingering PT and NGBRI patients fell above the cutoff. It was also found that NGBRI patients recalled significantly more items than DM patients. However, only partial support was gained as no significant differences were found between DM and PT in the total scores, and the average total score of the NGBRI group was significantly greater than that for PT. This raises concern, as the two non-malingering groups exhibited significantly different scores and the two groups most likely to receive the MFIT in criminal forensic settings (PT and DM) did not.

Several explanations can be offered for the significantly different quantitative scores of the PT and NGBRI groups. The first is that the archival nature of the PT group's data and their pretrial status (i.e., incentive to malinger) would make it difficult to rule out the presence of malingerers. However, close inspection of hospital staffs' behavioral observations of these patients in conjunction with the results of other psychological (malingering) assessment would indicate this is not so.

A second explanation would be that differences in the severity of psychiatric symptoms between the two groups contributed to differential performance. PT patients were typically admitted (and tested) during an acute phase of illness, whereas NGBRI patients had been hospitalized and treated for a longer time period. In fact, an

inspection of available records indicated that 40.6% of the PT patients had not yet received pharmacological treatment when they were assessed. By comparison, only two NGBRI patients were on no medications, and all had been hospitalized for 10 months or more. Furthermore, it had been noted that several of the PT patients had been determined by staff to be experiencing a first episode of psychosis, one typically characterized by disorganization. Together these findings indicate that PT patients had a more acute and severe symptom status. This is relevant, given that higher levels of conceptual disorganization, disorientation, and hallucinations were significantly associated with lower total scores among NGBRI participants. However, no firm conclusions can be reached at this time since no psychiatric symptom ratings were available for PT patients

A third and readily identified explanation is the contribution of lower intelligence in the performance of the PT group. Several investigations reported significant correlations between IQ and MFIT total scores (Goldberg & Miller, 1986; Hays et al., 1993; Simon, 1994), with the lowest reported specificity among individuals with both low IQ and (unspecified) psychiatric disturbance (Hays et al., 1993). Not surprisingly, the average IQ for PT patients (70.6) was significantly lower than that for NGBRI patients (75.2), and IQ was significantly correlated with the total score for both groups. Furthermore, closer inspection of the PT patients falling below the 9-item cutoff revealed that 57.1% had been diagnosed with mental retardation and 14.3% with borderline intellectual functioning. This occurred in conjunction with an Axis I diagnoses of schizophrenia, major depression, or some form of substance-induced cognitive disorder (dementia). It follows that the specificity for the PT group (56.2%) fell between that for individuals with mental retardation (62.5%; Goldberg &

Miller, 1986) and patients with various psychiatric problems and moderate to mild mental retardation (32.4%; Hays et al., 1993).

Although there was a trend for PT patients to obtain higher total scores than DM patients, this was not significant by conventional standards. This is important, as these are the two groups to be differentiated under clinical conditions. The most likely explanation for this finding can be found by way of comparison with the criminal forensic MFIT study of Simon (1994). In Simon's study, 85.7% of the diagnosed malingerers fell below the 9-item cutoff, compared to only 47.7% of the diagnosed malingerers in this study. Considering this with the assumption that malingering most likely lies on a continuum (e.g., Hayes et al., 1999; Resnick, 1993), it comes as no surprise that Simon reported the malingerers in his study to be rather extreme in their presentations. These discrepant rates of sensitivity raise the possibility that the severity of malingering in the current sample was more varied, an assumption supported by two findings. First, the behavioral observations of staff and the results of malingering assessment for the DM group (Tables 5 to 7) indicate that several individuals were less blatant in their efforts (i.e., exaggerating versus feigning). Second, 13 DM patients had an established preexisting condition of psychosis, an experience base that could potentially be used to develop a more sophisticated presentation style (e.g., Pachana et al., 1998). Consistent with this notion, only one DM patient with a preexisting psychosis fell below the cutoff, and even then by only a single point.

As noted by Rogers and Cruise (1998), the assumption underlying a floor effect measure is that the malingerer will "overplay" the role. Such an assumption would necessarily limit the utility of the measure among individuals engaging in less

extreme presentations, which seems to be supported by the present findings. However, it would also indicate better utility in identifying those being more blatant in their efforts. A closer inspection of the data in Table 11 reveals that only one non-malingering (NGBRI) patient had a quantitative score of less than 5, this was an individual with an established history of both psychosis and moderate mental retardation (tested IQ of 55). Only two non-malingering patients obtained a total score of 5, and both had IQ's in the lower end of the mild mental retardation range (less than 65). In comparison, 12 (27.3%) of the DM patients had total scores of 5 or less, seven of these had scores of 2 and 3.

In conclusion, one can be fairly confident in correctly identifying malingering when very low MFIT scores (< 5) are obtained. Otherwise, there is sufficient evidence that the traditional scoring method for the MFIT lacks specificity to the combined influences of lower intelligence and psychiatric symptoms affecting the ability to attend, concentrate, and organize cognitive processes. This method also lacks sensitivity to less extreme or more sophisticated forms of malingering. Together, these findings provide argument to examine alternative scoring methods, preferably those less affected by IQ and symptom status. This was examined in the remaining hypotheses.

Hypothesis 2

Several investigators (Goldberg & Miller, 1986; Greiffenstein et al., 1996; Morgan, 1991; Schretlen et al., 1991) observed qualitative errors involving perseveration, rotation, reversal, and misordering of items in the MFIT recall of patients with low IQ, primarily mental retardation. Given the presence of individuals experiencing problems with both psychosis and low IQ in criminal forensic settings

(e.g., Hayes et al., 1997), it was hypothesized these types of errors would be observed to a greater extent among the non-malingering patients. This hypothesis was not supported, as no significant between-groups differences were found for Item and Row Perseveration, Rotation, Reversal, Within Row, Between Row, and Row Sequence errors. In fact, only three non-malingering patients made a reversal or rotation error, and no significant correlations were found between IQ and these qualitative errors. Given the average IQ for non-malingerers fell in the lower end of the borderline range, this lack of significant findings is both unexpected and inconsistent with previous investigations. While the current study design did not permit examination of this issue, these results might suggest that the diminished cognitive functioning attributable to (or in conjunction with) psychosis is qualitatively different from that due to other organic factors.

There was observed an increased rate of errors involving perseveration and misordering of elements among malingerers. Although there was a non-significant trend for malingerers to make more Between Rows errors, it was later determined that the proportion of variance in this score attributable to differences between malingerers and non-malingerers was significant. While this is opposite of expectations, in retrospect it is not surprising. The MFIT was designed to facilitate recall through the grouping of related items (e.g., geometric shapes, first three letters of alphabet). The misordering of elements within a row (Within Row error) was the most frequently observed of all errors (one third of all groups), and it was observed this typically involved misordering of the geometric shapes. This is expected, as there is no "natural" order for these items (in contrast to "A B C"). However, placing together unlike items (e.g., A 3 O) would not be expected, except perhaps in cases involving

pronounced disorganization. The presence of this particular misordering error therefore appears to be more indicative of malingering.

Last, the presence of "non-malingering" errors among the DM might be explained by the presence of individuals with preexisting psychosis and low IQ. However, a closer inspection of the data revealed that only 3 or less of the 13 patients with an established preexisting condition made each of these specific types of errors. Rather, the results indicate these qualitative errors (with the exception of Between Rows) lack the ability to differentiate between malingering and non-malingering criminal forensic inpatients, at least when considered in isolation.

Hypothesis 3

The third hypothesis concerned the types of errors expected of malingerers. On the basis of previous findings (Benton & Spreen, 1961; Griffin et al., 1996; Guilmette et al., 1994), it was expected that DM would make more errors involving bizarre and unusual reproductions (Embellishment) and addition of extraneous elements than non-malingering patients. This would seem logical from a malingering standpoint, as someone desiring to appear "insane" might accomplish this through making "bizarre" errors. To the contrary, it was found that non-malingerers were just as likely to make these types of errors as malingerers, and the canonical correlation for these variables indicated a greater (although non-significant) association with non-malingering status.

One possible explanation is that the bizarre reproductions among non-malingerers reflects the bizarre or unusual thinking often seen in schizophrenia (APA, 1994). However, the small proportion of non-malingerers making these errors prohibited further statistical analysis of their association with specific symptoms. A

second and more likely explanation is that the hypothesis was predicated on findings derived from studies of questionable ecological validity (i.e., simulation and differential prevalence designs) (Rogers & Cruise, 1998). Nonetheless, these errors did not significantly differentiate malingerers from non-malingerers.

Hypothesis 4

Morgan (1991) observed that a high percentage of patient with severe to profound memory problems could recall the first 2 rows of the MFIT, i.e., exhibit a primacy effect in their recall. Although this was not examined in a malingering context, Morgan speculated that failure to exhibit a primacy effect might be indicative of malingering. This was the fourth hypothesis, that non-malingerers would exhibit more of a primacy effect in their recall compared to malingerers. Consistent with expectations, it was found that both PT and NGBRI patients recalled significantly more items from the first 2 rows than DM patients. Furthermore, no significant differences were found between the two non-malingering groups.

Additional important findings were made. First, failure to recall at least 3 of the 6 primacy items was highly indicative of malingering (only one non-malingering patient fell in this range). Scores of 3 were also more common among malingerers, although confidence in correct identification was increased when the total quantitative score fell below the 9-item cutoff. Second, using a cutoff of less than 6 primacy items resulted in improved specificity over the traditional scoring for PT patients (56.2 to 68.7%) while leaving sensitivity unchanged. Last, although lower primacy scores were associated with lower IQ, there were no significant correlations between this score and any of the 24 psychiatric symptoms assessed. That this score was

unaffected by symptom status makes it a more desirable alternative to the quantitative score.

An interesting aspect of the finding of a primacy effect is one would expect to see a recency effect in the short-term recall of individuals with memory problems (i.e., Wiggins & Brandt, 1988). A possible explanation lies in the finding that individuals with severe psychiatric disturbance have diminished processing speed (e.g., Malloy & Duffy, 1994; Malloy & Richardson, 1994). The allotted MFIT exposure time (10 seconds) may therefore be inadequate for these patients, in essence they only recall what they have had time to scan (i.e., first 2 rows). This would explain why the quantitative score was affected by those symptoms reflecting or contributing to impaired cognitive processing, whereas the primacy score was not. It is therefore suggested that future investigations consider varying the length of exposure time for these patients, in particular those who are acutely ill, as this may result in improved specificity. It is also suggested that greater emphasis be placed on primacy scores.

Hypothesis 5

Last, it was hypothesized that a combination of quantitative and qualitative scoring methods would provide better discrimination between malingerers and non-malingerers. This was supported through three findings. First, as discussed in the fourth hypothesis, quantitative scores falling below the cutoff in conjunction with lower primacy scores were found to be more indicative of malingering. Second, given the uncertainty that the Capitalization error should be counted as an error, quantitative scores were recalculated correcting for the presence of this error. This resulted in improved specificity for non-malingerers, in particular the PT patients (56.2 to 68.7%). Specificity for NGBRI patients increased marginally (86.7 to 90.0%), while

sensitivity was left unchanged (i.e., none of the DM falling below the cutoff made this type of error).

The third method through which this was examined was a linear discriminant analysis involving both quantitative and qualitative variables. Only those variables accounting for the largest proportion of variance between malingerers and non-malingerers (total Capitalization-corrected score, between rows error, and extraneous element) were used, controlling for the effects of age. Overall correct classification rose from 61.3% for the quantitative method to 72.6%, and was maintained upon cross-validation within the same sample (71.7%). Although specificity was at an acceptable level (83.9%), there was only a small increase in sensitivity (47.7 to 56.8%). These results indicated that higher total scores (Capitalization corrected) and extraneous elements were more indicative of non-malingering status, whereas lower total scores and between rows errors were more associated with malingering.

A problem with incorporating the qualitative scoring method was that, despite a high base rate for errors per se, less than one third of any group made any one particular type of error. Such a low base rate for errors necessarily limits their usefulness in discriminating between groups. This is similar to Griffin et al.'s (1996) finding of a low base rate for qualitative errors, from which it had been concluded that the presence of any "malingering" error be taken as evidence of feigning. The current results do not support this recommendation, as malingerers were just as likely overall to make those errors typically associated with genuine impairment while actual patients were just as likely to make those errors Griffin et al. believed to be indicative of malingering. Rather, these results suggest that the presence of between rows errors

and extraneous elements need to be taken in consideration when interpreting total MFIT scores in attempts to identify malingerers in criminal forensic inpatient settings.

Validity and Effectiveness

To summarize, 47.7% of diagnosed malingerers were correctly identified when using the traditional scoring method. This figure slightly increased to 56.8% when qualitative errors were incorporated. This is substantially less than the 85.7% sensitivity among blatant pretrial malingerers (Simon, 1994) and just below the estimated sensitivity of 64% for civil litigants. However, this was well above the overall sensitivity computed from studies using simulation (37.5%) and differential prevalence designs (22.4%).

In terms of specificity, 86.7% of NGBRI patients were correctly identified through their MFIT scores. This was almost identical to that for Simon's (1994) NGBRI control group (85.7%) and a sample of patients with schizophrenia (87%; Back et al., 1996). In comparison, only 56.2% of the non-malingering pretrial patients were correctly identified, although this number increased substantially (to approximately 80%) through incorporation of qualitative scoring. Incorporating qualitative scoring produced only a marginal increase for the NGBRI participants (90%).

The practical implication of these findings can be determined through computations of validity and effectiveness for the various approaches. The traditional scoring method was found to be valid for differentiating both groups of non-malingerers from malingerers. This was also found to be the case when using the Capitalization-corrected total score and the combined quantitative and qualitative procedure derived through LDA. However, the traditional scoring procedure was not

more accurate than utilizing base rate prediction alone in this regard, given the high combined error rates for PT and DM patients (48.7%). This leaves a base rate effectiveness range of approximately 49 to 51%, a number far in excess of the estimated 16% base rate for malingering among criminal defendants (Rogers et al., 1998). In comparison, combining the quantitative and qualitative methods produced a lower combined error rate of 27.4%, yielding a base rate effectiveness range of 28 to 72%. Although this is still above the estimated base rate, it is only slightly above the highest reported base rate of 25% for all forms of malingering (exaggeration to blatant fabrication) within criminal forensic settings (Rogers, 1986).

Before making final conclusions about the effectiveness of the combined approach, certain limitations need to be noted. First, the malingering base rate estimate that was used (16%) was derived from a survey of over 500 experienced forensic clinicians (Rogers et al., 1998). While such a large number of professional opinions would seem valid, the lack of objective diagnostic criteria for malingering combined with a failure to use objective measures undermines the validity of this estimate. This restricts the conclusions that can be reached on the effectiveness of the MFIT, or any other malingering measure, among criminal forensic populations. In the least, it provides argument for objectively determining local base rates.

Second, the effectiveness equation does not take into account possible measurement error in the determination of the combined error rate. Specifically, the interval of confidence surrounding the estimated false positive and false negative rates needs to be weighed when deciding if a measure is effective. While this is less of a concern when the base rate and error rates are widely disparate, it becomes problematic when the differences are minimal (as in the combined scoring approach).

Such situations provide argument for cross-validating results before reaching any firm conclusions about effectiveness.

Last, this study largely relied upon data from archival records. The time span covered by this study (one decade) was considerable, and a number of professionals were involved in making observations and assessments during this time. There were also variations in backgrounds and training that could not be accounted for. These factors make it difficult to determine to what extent, if any, that observations, diagnoses, and administration of assessments were done in a completely reliable fashion. Although this loss of experimental control has been a noted problem with known-groups designs in malingering research (Rogers et al., 1993), it has been considered an acceptable tradeoff given the potential gains in ecological validity (as compared to simulation and differential prevalence designs).

Final Conclusions

To summarize, the final conclusions are offered for use of the MFIT in criminal forensic settings:

1. The current results indicate the MFIT is accurate at identifying malingerers within criminal forensic settings. Although it is not more accurate than the base rate at achieving this goal, incorporating specific errors (between rows and extraneous elements) with the total score can increase correct classification, mainly that for actual patients. The large Type II error therefore necessitates the use of additional malingering measures. However, it should be noted that use of base rate predictions alone would not result in identification of any malingerers, as well multiple indices are used in making this determination.

2. The lack of effectiveness of the MFIT with this population can be attributed to three factors. First, the low estimated base rate for malingering leaves little room for error on any given measure. Second, the combined effects of hallucinations, disorientation, conceptual disorganization, and lowered intelligence results in increased false positives. Third, sensitivity for this measure is decreased substantially when dealing with malingerers who are less blatant or more sophisticated in their presentations.
3. Capitalization errors should not be counted against the total MFIT score, as this reduces the amount of Type I error.
4. One can be more confident in correctly identifying malingering when a) very low total scores (< 5) are obtained and b) there is a failure to recall at least three items from the first 2 rows, particularly when total scores fall below the traditional cutoff.
5. It has been noted that certain psychiatric symptoms (i.e., disorganized speech) are difficult to malinger for extended periods of time (Resnick, 1993). Given the association between this symptom and lower MFIT performance, it is recommended that lengthy behavioral observations be conducted before interpreting low MFIT scores. Likewise, the impact of low IQ on MFIT scores argues for the importance of determining premorbid intellectual functioning (e.g., failure to meet developmental milestones, school records indicating severe learning problems) before making clinical decisions.
6. Future MFIT research would benefit from a) obtaining psychiatric symptom ratings for pretrial non-malingerers and b) examining the effects of increased administration time on classification rates. Future malingering research would

likewise benefit from identifying or developing neurocognitive malingering assessments that are relatively impervious to the effects of low IQ and psychiatric illness.

7. Last, the current results indicate the MFIT has utility as a screening (rather than diagnostic) measure for malingering within criminal forensic settings, particularly when dealing with more blatant forms of deception.

References

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., revised). Washington, DC: Author.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Arnett, P. A., Hammeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-Item Test in neurological patients and dissimulators. The Clinical Neuropsychologist, 9(1), 17-26.

Back, C., Boone, K., Edwards, C., Parks, C., Burgoyne, K., & Silver, B. (1996). The performance of schizophrenics on three cognitive tests of malingering, Rey-15 Item Memory Test, Rey Dot Counting, and Hiscock Forced-Choice Method. Assessment, 3(4), 449-457.

Beaber, R. J., Marston, A., Michelli, J., & Mills, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. American Journal of Psychiatry, 142, 1478-1481.

Benton, A., & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. Archives of General Psychiatry, 4, 79-83.

Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. Journal of Clinical and Experimental Neuropsychology, 12(5), 715-728.

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.

Bernard, L. C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. Journal of Clinical Psychology, 49, 45-53.

Binks, P. G., Gouvier, W. D., & Waters, W. F. (1997). Malingering detection with the Dot Counting Test. Archives of Clinical Neuropsychology, 12(1), 41-46.

Dush, D. M., Simons, L. E., Platt, M., Nation, P. C., & Ayres, S. Y. (1994). Psychological profiles distinguishing litigation and nonlitigation pain patients: Subtle and not so subtle. Journal of Personality Assessment, 62, 299-313.

Faust, D., & Nurcombe, B. (1989). Improving the accuracy of clinical judgment. Psychiatry, 52, 197-208.

Gothard, S., Rogers, R., & Sewell, K. W. (1995). Feigning incompetency to stand trial: An investigation of the Georgia Court Competency Test. Law and Human Behavior, 19(4), 363-373.

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

Gouvier, W. D., Hayes, J. S., & Smiroldo, B. B. (1998). The significance of base rates, test sensitivity, test specificity, and subjects' knowledge of symptoms in assessing TBI sequela and malingering. In C. Reynolds (Ed.), Detection of malingering in head injury litigation. New York: Plenum.

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

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Comparison of multiple scoring methods for Rey's malingered amnesia measures. Archives of Clinical Neuropsychology, 11(4), 283-293.

Griffin, G. A. E., Normington, J., & Glassmire, D. (1996). Qualitative dimensions in scoring the Rey Visual Memory Test of Malingering. Psychological Assessment, 8, 383-387.

Guilmette, T. J., Hart, K. J., Giuliano, A. J., & Leininger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock Forced-Choice procedure. The Clinical Neuropsychologist, 8(3), 283-294.

Hankins, G. C., Barnard, G. W., & Robbins, L. (1993). The validity of the M Test in a residential forensic facility. The Bulletin of the American Academy of Psychiatry and the Law, 21(1), 111-121.

Hayes, J. S., Hale, D. B., & Gouvier, W. D. (1997). Do tests predict malingering in defendants with mental retardation? The Journal of Psychology, 131(5), 575-576.

Hayes, J. S., Hale, D. B., & Gouvier, W. D. (1998). Malingering detection in a mentally retarded forensic population. Applied Neuropsychology, 5(1), 33-36.

Hayes, J. S., Hilsabeck, R. C., & Gouvier, W. D. (1999). Malingering traumatic brain injury: Current issues and caveats in assessment and classification. In N. R. Varney & R. J. Roberts (Eds.), The Evaluation and Treatment of Mild Traumatic Brain Injury (pp. 249-290). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hays, J. R., Emmons, J., & Lawson, K. A. (1993). Psychiatric norms for the Rey 15-Item Visual Memory Test. Perceptual and Motor Skills, 76, 1331-1334.

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

Lee, G. P., Loring, D. W., & Martin, R. C. (1992). Rey's 15-Item Visual Memory Test for the detection of malingering: Normative observations on patients with neurological disorders. Psychological Assessment, 4, 43-46.

Leng, N. R. C., & Parkin, A. J. (1995). The detection of exaggerated or simulated memory disorder by neuropsychological methods. Journal of Psychosomatic Research, 39(6), 767-776.

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

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

Main, C. J., & Spanswick, C. C. (1995). 'Functional overlay' and illness behaviour in chronic pain: Distress or malingering? Conceptual difficulties in medico-legal assessment of personal injury claims. Journal of Psychosomatic Research, 39(6), 737-753.

Malloy, P., & Duffy, J. (1994). The frontal lobes in neuropsychiatric disorders. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology, Vol. 9 (pp. 203-232). Elsevier Science.

Malloy, P. F., & Richardson, E. D. (1994). Assessment of frontal lobe functions. Journal of Neuropsychiatry and Clinical Neurosciences, 6, 399-410.

Millis, S. R., & Kler, S. (1995). Limitations of the Rey Fifteen-Item Test in the detection of malingering. The Clinical Neuropsychologist, 9(3), 241-244.

Morey, L. C. (1991). The Personality Assessment Inventory professional manual. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (1996). An interpretive guide to the Personality Assessment Inventory (PAI). Odessa, FL: Psychological Assessment Resources.

Morgan, S. F. (1991). Effect of true memory impairment on a test of memory complaint validity. Archives of Clinical Neuropsychology, 6, 327-334.

Norusis, M. J. (1988). SPSS/PC+ advanced statistics. Chicago: SPSS, Inc.

Pachana, N. A., Boone, K. B., & Ganzell, S. L. (1998). False positive errors on selected tests of malingering. American Journal of Forensic Psychology, 16(2), 17-25.

Pedhazur, E. J. (Ed.) (1982). Multiple regression in behavioral research (2nd Ed.). New York: Holt, Rinehart, and Winston.

Resnick, P. J. (1993). Defrocking the fraud: The detection of malingering. Israel Journal of Psychiatry and Related Sciences, 30(2), 93-101.

Resnick, P. J. (1997). Malingered psychosis. In R. Rogers (Ed.), Clinical assessment of malingering and deception, 2nd ed. (pp. 47-67). New York: Guilford Press.

Rey, A. (1941). L'examen psychologie dans las cas d'encephalopathie traumatique. Archives de Psychologie, 23(112), 286-340.

Rey, A. (1964). L'examen clinique en psychologie. Paris: Presses Universitaires de France.

Rogers, R. (1986). Conducting insanity evaluations. New York: Van Nostrand Reinhold.

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

Rogers, R., Bagby, R. M., & Gillis, J. R. (1992). Improvements in the M Test as a screening measure for malingering. Bulletin of the American Academy of Psychiatry and the Law, 20(1), 101-104.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. Law and Human Behavior, 22(3), 273-285.

Rogers, R., Gillis, J. R., Dickens, S. E., & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the Structured Interview of Reported Symptoms. Psychological Assessment, 3, 89-96.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. Clinical Psychology Review, 13, 255-274.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. Law and Human Behavior, 22(4), 353-367.

Rogers, R., Sewell, K. W., Cruise, K. R., Wang, E. W., Ustad, K. L. (1998). The PAI and feigning: A cautionary note on its use in forensic-correctional settings. Assessment, 5(4), 399-405.

Rogers, R., Sewell, K. W., & Goldstein, A. M. (1994). Explanatory models of malingering. Law and Human Behavior, 18(5), 543-552.

Rubinsky, E. W., & Brandt, J. (1986). Amnesia and criminal law: A clinical overview. Behavioral Sciences and the Law, 4(1), 27-46.

Schacter, D. L. (1986a). On the relation between genuine and simulated amnesia. Behavioral Sciences and the Law, 4(1), 47-64.

Schacter, D. L. (1986b). Amnesia and crime: How much do we really know? American Psychologist, 41(3), 286-295.

Schretlen, D., Brandt, J., Kraft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-Item Memory Test to detect malingered amnesia. Psychological Assessment, 3(4), 667-672.

Schretlen, D., Van Gorp, W. G., Wilkins, S. S., & Bobholz, J. H. (1992). Cross-validation of a psychological test battery to detect faked insanity. Psychological Assessment, 4, 77-83.

Simon, M. J. (1994). The use of the Rey Memory Test to assess malingering in criminal defendants. Journal of Clinical Psychology, 50(6), 913-917.

Smith, G. P. (1997). Assessment of malingering with self-report instruments. In R. Rogers (Ed.), Clinical assessment of malingering and deception, 2nd ed. (pp. 351-370). New York: Guilford Press.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. Journal of Clinical and Experimental Neuropsychology, 15(4), 578-590.

Ventura, J., Lukoff, D., Nuechterlein, K. H., Liberman, R. P., Green, M. F., & Shaner, A. (1993). Manual for the Expanded Brief Psychiatric Rating Scale. International Journal of Methods in Psychiatric Research, 3, 221-224.

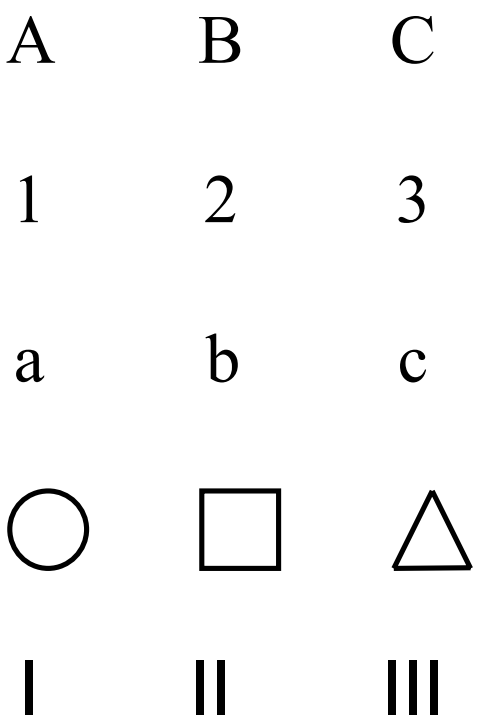
Warrington, E. K. (1984). Recognition Memory Test: Manual. Berkshire, UK: NFER-Nelson.

Wiggins, E. C., & Brandt, J. (1988). The detection of simulated amnesia. Law and Human Behavior, 12(1), 57-78.

Zachary, R. A. (1991). Shipley Institute of Living Scale: Revised manual. Los Angeles: Western Psychological Services.

Appendix A

The Rey 15-Item Memory Test



Appendix B

The Negative Impression Management Scale

Please read each statement and decide if it is an accurate statement about you. Use the following scale to rate each item:

0 = False, not at all true

1 = Slightly true

2 = Mainly true

3 = Very true

____ 1. Sometimes I cannot remember who I am.

____ 2. I have visions in which I see myself forced to commit crimes.

____ 3. Since the day I was born I was destined to be unhappy.

____ 4. I have three or four completely different personalities inside of me.

____ 5. People don't understand how much I suffer.

____ 6. Every once in a while I totally lose my memory.

____ 7. Sometimes my vision is only in black and white.

____ 8. I don't have any good memories from my childhood.

____ 9. I have severe psychological problems that begin very suddenly.

Appendix C

The Brief Psychiatric Rating Scale

Name/ID # _____ Date _____ Rater _____
Hospital/Location _____ Period of assessment _____

NA	1	2	3	4	5	6	7
Not assessed	Not Present		Very Mild	Mild	Moderate	Moderately Severe	Severe	Ext. Severe

Rate items 1-14 on the basis of patient's self-report during interview. Mark "NA" for symptoms not assessed. Note items 7, 12, and 13 are also rated on observed behavior during the interview. PROVIDE EXAMPLES.

1.	Somatic Concern	NA	1	2	3	4	5	6	7
2.	Anxiety	NA	1	2	3	4	5	6	7
3.	Depression	NA	1	2	3	4	5	6	7
4.	Suicidality	NA	1	2	3	4	5	6	7
5.	Guilt	NA	1	2	3	4	5	6	7
6.	Hostility	NA	1	2	3	4	5	6	7
7.	Elevated Mood	NA	1	2	3	4	5	6	7
8.	Grandiosity	NA	1	2	3	4	5	6	7
9.	Suspiciousness	NA	1	2	3	4	5	6	7
10.	Hallucinations	NA	1	2	3	4	5	6	7
11.	Unusual Thought Content	NA	1	2	3	4	5	6	7
12.	Bizarre Behavior	NA	1	2	3	4	5	6	7
13.	Self-neglect	NA	1	2	3	4	5	6	7
14.	Disorientation	NA	1	2	3	4	5	6	7

Rate items 15-24 on the basis of observed behavior or speech of the patient during the interview.

15.	Conceptual Disorganization	NA	1	2	3	4	5	6	7
16.	Blunted Affect	NA	1	2	3	4	5	6	7
17.	Emotional Withdrawal	NA	1	2	3	4	5	6	7
18.	Motor Retardation	NA	1	2	3	4	5	6	7
19.	Tension	NA	1	2	3	4	5	6	7
20.	Uncooperativeness	NA	1	2	3	4	5	6	7
21.	Excitement	NA	1	2	3	4	5	6	7
22.	Distractibility	NA	1	2	3	4	5	6	7
23.	Motor Hyperactivity	NA	1	2	3	4	5	6	7
24.	Mannerisms and Posturing	NA	1	2	3	4	5	6	7

Sources of information

_____ Patient
_____ Parents/Relatives
_____ Mental Health Professionals
_____ Chart

Explain here if validity of assessment is questionable

_____ Symptoms possibly drug-induced
_____ Underreported due to lack of rapport
_____ Underreported due to negative symptoms
_____ Patient uncooperative
_____ Difficult to assess due to formal thought disorder
_____ Other _____

Appendix D

Demographics Form

Study # _____ Years of Education _____

Age _____ Legal Status _____

Race _____

Charge(s)

_____

_____

Maximum possible sentence:_____

prior arrests: _____ Felony _____ # convictions

_____ Misdem _____ # convictions

_____ City _____ # convictions

Diagnosis:

AxisI_____

Axis II: _____

Axis III: _____

h/o head injury w/ LOC?

documented h/o neurological condition (include source and results of CT/MRI/EEG)?

h/o seizure disorder? (include treatments, last documented seizure)

prior diagnosis of malingering? (state by whom)

Previous suspicion of malingering? (state by whom and what reasons/evidence)

Provide any documented h/o mental retardation (give IQ when available, source):

Provide any documented h/o mental illness (include diagnoses, treatments, duration, source)

# of prior hospitalizations:

TEST DATA:

FSIQ: _____ circle one WAIS-R WAIS-III SILS RSPM

4-subtest WAIS-R

Referred for malingering evaluation?

If yes, was it for psychosis, cognitive (including memory) or both?

Behavioral observations by tx. team or psychologist:

Was malingering confirmed through other tests?

What were they malingering?

Other tests given:

____ SIRS _____ total scales in definite range (RS SC IA BL SU SEL SEV RO)

____ total scales in probable range (RS SC IA BL SU SEL SEV RO)

___ total scales in indefinite range (RS SC IA BL SU SEL SEV RO)

_____ probability of malingering

_____ M Test _____ total C items _____ total S items

_____ total M items

_____ Dot Counting _____ time for ungrouped

_____ time for grouped

_____ total number of ungrouped correct

_____ total number of grouped correct

_____ Digit Memory _____ level of performance

Other confirming sources of malingering (e.g., improbably poor performance on neuropsych tests, MMPI or PAI scales, etc.):

Appendix E

Consent Form

1. Title: Qualitative Scoring of the Rey 15-Item Memory Test in a Forensic Population.

2. Where: Eastern Louisiana Mental Health System Forensic Division.

3. Contacts: If you have any questions, you may contact the following individuals Monday through Friday between the hours of 7:30 a.m. and 6:00 p.m.

Dr. David Hale	James Martin
ELMHS Forensic Division	ELMHS Forensic Division
Ph. (225) 634-2661 ext. 65	Ph. (225) 634-2661 ext. 72

4. Purpose of the Study: This is a research study which will look at the different types of errors that are made by patients in a forensic hospital on the Rey 15-Item Memory Test. We want to see if patients who are not malingering make different types of errors than patients who are suspected of malingering.

5. Participants: This study is open to all people who have been found "Not Guilty by Reason of Insanity" and committed by the courts to the Forensic Division. People who are currently experiencing severe psychological problems, who have been found to be exaggerating their psychological and intellectual problems, or who have a history of severe brain injury will be unable to participate.

6. Number of Participants: 30 patients at the Forensic Division will be enrolled in this study.

7. Procedures: This study will only use those tests which are normally used at the Forensic Division. You will first be asked to complete a short test of your memory. Then you will be asked about any psychological problems you may or may not have, such as feeling nervous, sad, or hearing voices. Last, you will be asked to complete some tests which ask you true or false questions, questions about how you generally see yourself, and tests which ask you to count dots and solve problems. The whole study will only take one hour of your time.

8. Benefits: You will receive $2.00 for completing the entire study. The other benefit is you will be helping us understand if using a different scoring method makes a test more useful for forensic patients.

9. Risks: There are no risks for doing this study. However, if you should feel uncomfortable you can refuse to answer questions or stop the study at any time.

10. Right to Refuse: You do not have to do this study if you do not want to. You can also stop doing the study at any time. If you decide not to do the study or decide

to stop the study, your decision will not affect your treatment at this facility or get you in trouble with the staff. You will not be punished in any way or lose points on the Level system.

11. Privacy: The information we get from you will only be used for this study, and is not to be used by the courts or for your treatment at Forensic Division. This is an anonymous study, which means your name will not be placed on any forms except the consent form, and this is to be kept in a different file. That way, no one will know who you are or how you did on these tests. Only the people listed above will be able to look at your answers. After the study, we will destroy all of these forms, so no one will know you did the study.

12. Financial Information: You will paid two (2) dollars upon completion of the entire study. This money will be deposited directly to your patient account.

13. Alternatives: The alternative is not to participate in this study.

14. Withdrawal: Since your participation in this study is voluntary, you have the right to stop at any time. This will not affect your treatment at Forensic Division or place on the Level system. However, only those people who complete the entire study will be paid the two dollars.

15. Removal: Participants who become disruptive, aggressive, or very psychotic during the study will be removed from the study.

16. Signatures: The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigators. If I have questions about subjects' rights or other concerns, I can contact Charles E. Graham, Chairman, LSU Institutional Review Board, (225) 388-8692. I agree to participate in the study described above and acknowledge the researcher's obligation to provide me with a copy of this consent form if signed by me.

Signature of the Patient Volunteer Date

"The study subject has indicated to me that he is unable to read. I certify that I have read this consent form to the subject and explained that by completing the signature line above, the subject has agreed to participate."

Signature of Reader Date

Signature of Witness Date

Signature of Investigator Date

Vita

James Martin was born in Athens, Georgia, on July 7, 1966. He spent his childhood and adolescent years with his mother, father, and two younger brothers in Augusta, Georgia, and Gainesville, Georgia. Upon graduating from high school in 1984, he entered the University of Georgia, where he earned a Bachelor of Science degree in psychology in 1988. From there he pursued a Master of Science in psychology through Augusta State University. After receiving the degree in 1991, he pursued work in the areas of neuropsychological testing and then psychopharmacological research with patients with schizophrenia. Realizing the importance of further education, he re-entered graduate school at the Illinois Institute of Technology, where he began social-cognitive research with schizophrenia under the guidance of David Penn, Ph.D. This work continued when he transferred with Dr. Penn to Louisiana State University, and expanded into working with inpatients in the state forensic hospital. His interests now include working with patients with schizophrenia in both forensic and non-forensic contexts, utilizing new cognitive-behavioral therapies for treatment-resistant psychotic symptoms, and further examining factors affecting neuropsychological test performance among patients with schizophrenia. He currently resides in Gainesville, Georgia, with his Cajun-bred wife (also a doctor in psychology) of four years and two daughters, ages 3 and 12, with whom he enjoys outdoor activities.