The Rey Complex Figure is Moderately Useful as a Screen for Poor Effort Among Veterans with

Possible Mild Traumatic Brain Injury

by

Nazanin Maghsoodi

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of Bachelor of Arts

With Honors in Psychology from the

University of Michigan

2011

Advisor: Dr. Linas A. Bieliauskas

Abstract

Measures of effort assist neuropsychologists in determining the validity of assessment data. Formal tests of effort are often time-consuming, while embedded measures of effort within standard neuropsychological tests require no additional administration time. The Rey Complex Figure Test (RCFT) is a widely administered test of visual perception and memory, executive functioning, and visual motor and constructional skills. It may also be a useful measure of effort. The current project explores embedded performance cutoffs for effort on the RCFT. Data were collected from 353 consecutive veterans seen in a traumatic brain injury clinic who underwent neuropsychological testing, including the RCFT, and two measures of effort: the Test of Memory Malingering (TOMM) and Digit Span. Data from the 214 veterans established effort cut-offs on three RCFT measures: copy accuracy, copy time, and recall accuracy. Chi-Square analyses indicated that only age-corrected and recall accuracy scores below the 16th percentile were significantly related to effort. These scores were added into a single index and cross-validated on a second group of 139 veterans who had taken both the TOMM and Digit Span, failure on either of which was considered indicative of suspect effort. Twenty-four percent of the veterans failed either the TOMM or Digit Span. Among participants with zero, one and two RCFT scores below the 16th percentile, 11%, 28%, and 45% had poor effort on either the TOMM or Digit Span. The RCFT contains embedded measures that provide an indication of effort and can be used to identify individuals in need of more comprehensive assessment of symptom validity.

The Rey Complex Figure is Moderately Useful as a Screen for Poor Effort Among Veterans with

Possible Mild Traumatic Brain Injury

The fundamental assumption of neuropsychological tests is that the activity of the brain

manifests itself in behavior, and abnormal behavior is reflective of abnormal brain composition

or function. The patient undergoing neuropsychological assessment completes structured tasks

under controlled, standardized conditions. Performance on these tasks is assumed to reflect

brain-behavior relations and have a direct bearing on function. Inferences derived from these

assessments are then used for diagnosis, treatment planning, educational interventions, and

tracking treatment gains or disease progression over time. The cooperation of patients is essential

to this mutual exploration, a necessary partnership that is far from universal among medical tests.

For example, tests such as biopsies, electrocardiogram (EKG), or electroencephalogram (EEG)

do not require the level of sustained cooperation of a neuropsychological test battery, where the

examinee essentially joins the neuropsychologist as a co-investigator. Accordingly, the

underlying assumption of neuropsychological test data is that performance involves some

minimal threshold of cooperation and thus, reflects true cognitive ability. Research has shown,

however, that the assumption of examinee cooperation is sometimes unfulfilled, and test data

cannot always be interpreted at face value.

It is important to know whether test data reflects low examinee cooperation. Although

many practitioners colloquially use the term "malingering" as a catch-all, poor patient effort

comes in many different forms and is referred to by many different names. Terms such as poor

effort, incomplete effort, malingering, and the like, each refer to suboptimal cooperation on the

part of the patient and communicate the same core message: that test performances cannot be

interpreted as a straightforward representation of true ability. When there is an indication of poor

effort, the best that can be salvaged from the neuropsychological test data is that the manifest

performance reflects the lower limits of the true ability of the patient and not necessarily the

actual ability[1]. Such nuanced interpretations are only possible if poor effort is detected.

    Before discussing tests of effort, it is necessary to partial "poor effort" into its nosological

parts. Formal tests of effort can help ascertain the validity of the neuropsychological test results,

but they do not specify the reasons behind poor effort, which can be the end product of a variety

of motivations and situational factors. In the case of malingering, poor effort is intentional and

represents the conscious attempt to appear impaired for external gain. In instances of factitious

disorder, poor effort is also conscious, but the motive is to assume the sick role and elicit care-

taking from others, particularly treatment providers (Heilbronner, Sweet, Morgan, Larrabee, &

Millis, 2009). Many authors conceptualize a continuum between malingering and factitious

disorder that represents a variety of incentives, ranging from external to internal. Other instances

of poor effort are viewed as under unconscious control. The poor test performances from

individuals in these cases resemble psychosomatic illnesses, only with cognitive symptoms

(Delis & Wetter, 2007). Theorists have proposed a link between these psychosomatic reactions

and factitious disorders along a conscious-unconscious continuum. Although these categories are

useful for nosological purposes, and the continua are useful for conceptual purposes, the

behavior of actual examinees rarely conforms to such prototypical labels. In practice,

psychological and situational factors come into play.

    In some instances, patients may over-report symptoms and exaggerate deficits on

neuropsychological tests to catch the attention of the neuropsychologist and get what they

believe is proper treatment (Heilbronner et al., 2009). In other instances, poor effort may be due

to fluctuations in pain, depression, or mood. Greiffenstein, Baker, and Johnson-Greene (2002)

and Suhr, Tranel, Weffel, and Barrash (1997) argue that the false information provided by some litigants is more a function of the psychosocial dynamics of the litigation context, rather than due to any malicious attempt to deceive.

Issues of effort are particularly relevant with mild head injuries, where much of the research on psychosocial factors and effort has been done. Many studies indicate that in the great majority of individuals with mild head injuries, cognitive symptoms subside within three months of the injury (McCrea, 2008). There are many patients, however, who present with unusually severe deficits or with late-developing post-concussive syndrome more than one year later, claiming that their symptoms are due to their head trauma. One reason for distorting their symptoms, self-reports, and neuropsychological test performances could be for secondary gain. Psychosocial dynamics might also be at play. Many individuals with mild head injuries, for example, give an unrealistically positive depiction of their past functioning. This phenomenon, known as "the good old days," is evident among individuals in forensic contexts, but it is also seen among individuals with no apparent external incentives (Lange, Iverson, & Rose, 2010; Suhr & Gunstad, 2002, 2005). Thus, mTBI patients tend to attribute ordinary problems to a head injury because they fail to recall similar problems in "the good old days."[2] Neuropsychological test performances can also be affected by the expectations that individuals have of the typical problems associated with their diagnoses. Suhr and Gunstad (2005) demonstrated that individuals with histories of head injuries perform poorly on neuropsychological tests when attention is called to these events. In summary, patients' motivations and the situational contexts are important factors to consider when interpreting neuropsychological performances in general and effort tests specifically. Because patient motivation is usually not known, in this project, the more operationally definable and appropriately agnostic term, "poor effort," is used.

Examinee effort exerts a powerful influence over performance, and in some contexts, effort is more predictive of test performance than are factors universally assumed to impact scores such as brain injury, disease, and demographic characteristics (Green, Rohling, Lees-Haley, & Allen, 2001). Many sources urge that effort be assessed in practically all neuropsychological assessments (Bush et al., 2005). Historically, psychologists in clinical, research, and forensic settings estimated their patients' effort on neuropsychological tests using their own subjective impressions, based on clinical observations and qualitative analyses of standard neuropsychological test scores. However, such impressions can be misleading. Studies by Faust, Hart, and Guilmette (1988) and Ruff, Wylie, and Tennant (1993) demonstrated that neuropsychologists using their subjective impressions performed at chance levels in identifying individuals feigning cognitive deficits. In more recent times, neuropsychologists, particularly those working in forensic contexts, have been tasked with differentiating between those with genuine versus feigned brain injuries by using objective effort measures with established validity. Explorations of formal measures of effort began in earnest and now are well-supported by the research literature. Today, many effort tests are routinely used to establish the validity of neuropsychological test results.

Neuropsychological Assessment of Effort

Because it is often difficult to tell whether patients are putting forth their full effort, psychologists developed procedures to detect insufficient effort. Perhaps the first formal neuropsychological test of effort to gain wide usage was the Rey Fifteen-Item Memory Test (FIT; Lezak, 1983). This test requires examinees to view and then subsequently recall fifteen items that are presented visually on a card for ten seconds. The test was designed to appear difficult, but because the fifteen items are composed of five logical clusters of three (e.g., A-B-C,

1-2-3), the test is actually deceptively easy. The test became a staple of many neuropsychological

batteries, partly due to its intuitive appeal. Unfortunately, the test has poor sensitivity in most

contexts (Schretlen, Brandt, Krafft, & Van Gorp, 1991), and poor specificity among older

populations (Spencer et al., under review). Nevertheless, the test still enjoys wide usage, mostly

because of its brevity. In most cases, the test can be administered in less than one minute.

Following the development of the FIT, lengthier and more precise measures of effort have been

developed including the Green Word Memory Test (Green, Allen, & Astner, 1996), the Victoria

Symptom Validity Test (Slick, Hopp, Strauss, & Thompson, 1997), and the Test of Memory

Malingering (TOMM; Tombaugh, 1996, 1997). These measures offer far greater diagnostic

power, but at the cost of protracted administration time. Many of these tests take upwards of 20

minutes to administer, a cost which might be prohibitive in situations where there is a premium

on efficiency, such as in cognitive screenings or in brief examinations.

The cost-benefit analyses inherent in selecting effort tests create an ongoing dilemma for

the practicing neuropsychologist who must choose between efficiency and accuracy. One could

forgo efficiency and make an accurate assessment of effort, but risk sacrificing the opportunity

for an adequate and thorough assessment of cognitive abilities. On the other hand, one could

make a brief, cursory, assessment of effort but risk that the results are possibly of suspect

validity. One compromise that has been proposed is to use scores on traditional cognitive tests as

a gauge of effort, also known as the "embedded measure" approach. This strategy capitalizes on

some of the principles of traditional effort tests and has the advantage of requiring no additional

administration time (Larrabee, 2007). Most effort tests are effective because they appear to be

more difficult than they actually are. Thus, individuals feigning cognitive problems and who fail

to appreciate how easy the test is, perform poorly in an attempt to appear disordered (Strauss,

Sherman, & Spreen, 2006). These tests are validated in two main types of research. The first

approach examines the degree to which the tests differentiate between individuals with genuine

neurologic disease, versus healthy individuals instructed to simulate illness. The second type of

research design involves examining improbably poor scores among individuals with bona-fide

neurologic disease. Often, these approaches are blended so that evaluators know the scores

which are most likely to have been produced by individuals feigning problems and least likely to

have been produced by individuals with genuine neuropathology (Larrabee, 2007).

Effort assessments using measures embedded within the existing tests would benefit from

research using these approaches. For instance, many forms of neurologic disease are not

associated with deficits in simple attention, a fact that is likely underappreciated by most lay

people. Accordingly, deficits in simple attention-span tests, such as Digit Span (Wechsler, 2008),

are associated with poor effort (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Whitney,

Davis, Shepard, Bertram, Adams, 2009). Neuropsychologists, who are aware of this research and

are using this test, would suspect poor effort when examinees score poorly on this test.

Otherwise, they might interpret scores in a more straightforward manner, assessing attention and

working memory capacity, a purpose for which the test was intended.

Poor performance on an embedded measure is generally insufficient to independently

determine whether an individual is feigning problems. In the case of poor attention-span, deficits

might result from delirium from infection, or heightened anxiety, or severe global cognitive

impairments. For this reason, the suspicion of poor effort must be confirmed by other sources.

For instance, with low Digit Span scores, suspicion would rise among individuals with normal

lab values or among those who had intact performance on complex measures that require simple

attention.

Conceptually, the embedded measure approach can be taken with practically any neuropsychological test. In fact, in addition to simple attention tests, the embedded measure approach has been applied to tests of recognition memory (Kim et al., 2010; Millis, Putnam, Adams, & Ricker, 1995), problem-solving (Greve, Bianchini, Mathias, Houston, & Crouch, 2002), and processing speed (Iverson, Lange, Green, & Franzen, 2002). In each case, aspects of performances on certain measures of these tests are uncommonly problematic among those with neurologic illness and commonly "problematic" among individuals either simulating illness or who have genuine capability but are not producing a genuine effort. Among the tests that have been examined as an embedded measure of effort is the Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995).

Rey Complex Figure Test as a Measure of Effort

The Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995) is generally used to screen for a variety of cognitive abilities. The RCFT involves presentation of a complex geometric design to an examinee, which they copy as accurately as possible. Examiners generally score the accuracy of the copy, the time taken to copy the figure, and the organization of the copying approach. Resultant scores are measures of visuospatial processing, processing speed, and executive functioning, respectively. Subsequently, and generally without forewarning, the examinee is asked to reproduce the drawing from memory as a gauge of incidental visual recall. Research on the reliability and validity of the RCFT generally depict it as a rough measure of a variety of cognitive abilities. The RCFT is practically never used in isolation, and it is often used in cognitive screenings to generate hypotheses that can be evaluated with additional testing. In surveys of practicing neuropsychologists, the RCFT is among the most widely used neuropsychological instruments (Rabin, Barr, & Burton, 2005).

Because of its popularity and inclusion of multiple performance measures, the RCFT has attracted attention as a potential embedded measure of effort. Strauss and colleagues (2006) have noted that, at first glance, most examinees assume that the figure is more difficult to copy than it actually is. Individuals simulating cognitive impairment might overestimate the difficulty of the figure copy when approximating the performance of someone with genuine impairments. Furthermore, the RCFT is a measure of memory, an ability in which problems are frequently exaggerated among those feigning cognitive problems. Using a pattern-analysis approach, Meyers, Galinsky, & Volbrecht (1999) indicated that some patterns of memory errors, such as performing worse on recognition than on recall or performing very poorly on recognition, immediate recall, and delayed recall, indicate poor effort. The authors pointed out that, aside from severely impaired individuals, certain aspects of performance, such as attention and memory storage failure, are rarely impaired in most people.

Although the RCFT appears to have some attractive features of an effort test, only a handful of studies have empirically tested its possibilities. Knight and Meyers (1995) found that individuals simulating cognitive impairment perform poorly on copy accuracy, copy speed, delayed recall, and on a test of recognition of components of the figure (cited in Meyers & Meyers, 1995). They found that both normal and brain-damaged individuals rarely fail to recognize certain items. Subsequently, Lu, Boone, Cozolino, and Mitchell (2003) confirmed that the recognition trial of the RCFT, along with copy and immediate recall, is useful in identifying those with poor effort from among individuals with genuine memory impairments and healthy control participants. The authors created a composite score consisting of scores for copying accuracy, true positive recognition, and atypical recognition error scores that yielded a sensitivity of 74% (Lu et al., 2003).

These studies highlight the potential of the RCFT as a measure of effort, but methodological limitations have yet to be addressed. First, the studies by Knight and Meyers (1995) and Lu et al. (2003) employed modest sample sizes. Furthermore, the researchers used multiple measures to derive optimum cutoffs to classify participants. In the absence of cross-validation in independent samples, cutoffs derived from these analyses are vulnerable to sample-specific characteristics, and the resultant classification rates are likely inflated. Clinicians working with individual patients and who apply the cutoffs used in these studies may have a general impression of who is and is not simulating from a qualitative standpoint, but the psychometrics will be of limited use.

The second limitation from the studies that have examined the RCFT as a measure of effort is that it includes analyses of recognition performances. Recognition memory is assessed by the standard Meyers and Meyers (1995) protocol, but this version of the RCFT involves a 30-minute delay. Such a delay prolongs administration time, and so delayed recall and recognition are not routinely administered situations where rapid assessment is necessary.

In the current study, we will explore the power of RCFT copy accuracy, time to copy, and immediate (3-minute) recall as embedded measures of effort, using a large sample size of mild traumatic brain injury patients assessed at the Veteran's Affairs Ann Arbor Hospital. Scores on these RCFT measures will be examined in relation to scores on two established measures of effort to derive cutoff scores for poor effort on the RCFT. These cutoff scores will then be cross-validated on an independent sample. We hypothesize that scores derived for the RCFT will identify examinees displaying poor effort.

Method

The purpose of the current study was to assess whether performance measures from the Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995) could be used to measure effort. To answer this question, we compared aspects of performance on the RCFT (copy accuracy, time to copy, and three-minute recall accuracy) with performance on two standard indices of effort, the Digit Span subtest of the Wechsler Adult Intelligence Scales (WAIS-IV; Wechsler, 2008) and The Test of Memory Malingering (TOMM; Tombaugh, 1996).

Participants

Participants were 448 veterans referred for neuropsychological assessment in a traumatic brain injury clinic. Participants were consecutive referrals to the clinic who were included for data analyses if they had taken the RCFT and Digit Span subtest of the WAIS-IV (Wechsler, 2008). Participants were excluded if they had non-TBI CNS pathology (such as stroke and epilepsy), if they had a loss of consciousness exceeding 30 minutes, or post traumatic amnesia greater than 24-hours. A total of 95 veterans were excluded. The remaining 353 veterans were divided into two groups. Participants in Group One ($N = 214$) completed the Digit Span as their only effort measure. The TOMM was added to the TBI clinic's neuropsychological battery later on during the data collection process. Participants in Group Two ($N = 139$) completed both the Digit Span and TOMM.

Materials

Neuropsychological assessment. The data in this study was collected from veterans seen at the Ann Arbor Veterans Affairs Hospital TBI Clinic in south-east Michigan. The neuropsychological tests, measuring executive functioning, attention and concentration, memory, psychiatric symptoms, and effort, were part of a routine assessment given to all veterans receiving care at the clinic.

Rey Complex Figure Test (RCFT). The RCFT is a complex design containing 18 elements that generates three primary scores: copy accuracy, copy time, and recall accuracy after a three-minute delay filled with distraction. The accuracy with which patients copied and recalled the RCFT was scored on a 36-point scale for accurate drawing and correct placement of the figure's components (See Meyers & Meyers, 1995). Examinees were covertly timed as they copied the figure. We used the norms from Meyers and Meyers (1995) to determine performance cut-offs.

Test of Memory Malingering (TOMM). The TOMM (Tombaugh, 1996) is a forced-choice symptom validity test that involves up to three trials. In trials one and two, participants were presented with 50 consecutive target line drawings, which they were instructed to remember. After each learning trial, participants were presented 50 pages, with each containing one of the 50 original drawings and a new drawing. Participants were asked to identify the original drawing. The third trial, called the retention trial, occurs after a ten minute delay, and does not include a learning trial. Although the test appears difficult, it is usually performed flawlessly by both impaired and unimpaired individuals who put forth good effort. We defined failure on the TOMM as a score below 45 out of 50 on either the second learning trial or the retention trial. The test was discontinued when patients scored 45 or above on trial one, as previous research has indicated that such individuals very rarely fail the test (Rees, Tombaugh, Gansler, & Moczynski, 1998). Research has indicated that the TOMM is moderately sensitive to poor effort, and rarely misclassifies cooperative, though genuinely impaired, individuals as having poor effort (Larrabee, 2007; Rees et al., 1998).

Digit Span Subtest of the Wechsler Adult Intelligence Scales (Digit Span). The Digit Span Subtests is a standard neuropsychological test that can also be used as an embedded

measure of effort (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Greiffenstein, Baker and Gola, 1994; Suhr & Barrash, 2007). An age-corrected scaled score of six or less on Digit Span was used as a criterion to indicate poor effort on the RCFT (Larrabee, 2007).

Results

Descriptive Statistics

We used SPSS software for data analyses. Of the 353 participants, 90.7% were right-handed and 96.0% were male. Participants' Digit Span age-corrected scaled score spanned the full range, from 1 to 19 ($M = 8.6$, $SD = 2.5$). Participants scored from 15 to 36 points on the copy portion of the RCFT ($M = 33.4$, $SD = 3.2$) out of a possible range of 0 to 36 points. RCFT time to copy ranged from 47 to 420 seconds ($M = 166.3$, $SD = 63.9$). Scores on RCFT recall ranged from 0 to 35 ($M = 20.1$, $SD = 6.4$) out of a possible range of 0 to 36 points. Before the main analysis could be performed, the two samples were compared for demographic variables. Ages ranged from 20 to 78 for all participants, and those in Group One ($M = 30.1$, $SD = 8.7$) were significantly younger than those in Group Two ($M = 35.6$, $SD = 12.8$), $F(1, 351) = 23.1$, $p < .001$. Education ranged from 7 to 20 years, and there was no significant difference between Group One ($M = 12.8$, $SD = 1.5$) and Group Two ($M = 12.9$, $SD = 1.6$). There was no significant difference on Digit Span scores between Groups One and Two, $F(1, 351) = 1.7$, ns.

Group One: Establishing Cutoffs

Preliminary analyses were employed on the 214 veterans in Group One (Digit Span only) to establish effort cutoffs on three RCFT measures: copy accuracy, copy time, and recall accuracy. To establish optimum RCFT performance cutoff scores for poor effort, RCFT performances on relevant indices were compared to those who did and did not meet criteria for poor effort on the Digit Span subtest. These cutoffs were later applied to Group Two for cross-

validation. Insufficient effort on Digit Span was operationalized as an age-corrected scaled score

of six or less. Several published effort cutoffs have been suggested for Digit Span (Suhr &

Barrash, 2007). This cutoff was selected because it is relatively sensitive to poor effort (Axelrod,

Fichtenberg, Millis, & Wertheimer, 2006; Larrabee, 2007).

Next, the data were inspected for the correspondence between poor effort on Digit Span

and RCFT performance. Following the normative data presented in Meyers and Meyers (1995),

scores on the RCFT were standardized into five categories of percentiles for ease of

interpretation. Those scoring at and above the 16th percentile on the RCFT measures

disproportionately passed Digit Span. Rey scores were then transformed into a dichotomous

variable of pass/fail (scores at and above the 16th percentile versus scores below 16%). Next,

RCFT cutoffs were examined to determine if each measure was significantly related to

insufficient effort on Digit Span. Chi-Square analyses were performed for each performance

measure on the RCFT against pass/fail scores on Digit Span to identify whether the distribution

of the categorical variables was significant or random.

Among the 153 individuals scoring at and above the 16th percentile on RCFT copy

(passing), 86.9% passed the Digit Span effort measure. Of the 61 individuals who scored below

the 16th percentile (failing) RCFT copy, 31.1% displayed insufficient effort on Digit Span.

Among the 199 individuals passing the time segment of the RCFT, 82.4% passed the Digit Span

effort measure. Of the 11 individuals who failed RCFT time to copy, 36.4% also failed the Digit

Span effort measure. Of the 138 patients with passing scores on RCFT recall, 89.1% displayed

good effort on Digit Span, and 31.6% of the 76 people who failed the RCFT recall displayed

insufficient effort on Digit Span. Chi-Square analyses indicated that only age-corrected copy, χ^2

$(1, N = 214) = 9.6$, $p = .002$, and recall, $\chi^2 (1, N = 214) = 14.1$, $p < .001$, accuracy scores below

the 16th percentile were significantly related to effort. In summary, the preliminary analyses

suggested that cutoff scores below the 16th percentile on the RCFT copy and recall were effective

for identifying those performing with insufficient effort.

The copy and recall portions of the RCFT are typically administered in clinical settings,

and in this study, these RCFT measures were most strongly predictive of poor effort in Group

One. Therefore, we examined the degree to which the combination of scores from these

measures predicted performances on traditional measures of effort. We categorized participants

based upon whether they failed both copy and recall, failed either copy or recall, or failed

neither. These possible outcomes were scored as zero, one, and two, respectively. The possible

performance outcomes for the copy and recall combination were applied on the same participants

in Group One and compared with pass and fail scores on the Digit Span. Chi-Square analyses

indicated that they were moderately predictive in classifying those with and without poor effort

on Digit Span, χ^2 (2, $N = 214$) = 19.6, $p < .001$. Table 1 shows proportions of those individuals in

Group 1 passing and failing Digit Span (cutoff of 6/7) for each of the three possible performance

outcomes for the copy and recall combination.

Group Two: Cross-Validation

The three possible performance outcomes for copy and recall were cross-validated on an

independent sample, Group Two. Because individuals in Group Two completed both the Digit

Span and TOMM, their effort was categorized using both liberal and strict criteria for poor

effort. A liberal (inclusive) criterion for having poor effort is failing *either* the Digit Span or

TOMM. The strict criterion for poor effort is failing *both* the Digit Span and TOMM. Chi-Square

analysis on the liberal effort criterion indicated that the RCFT cutoff score on the copy and recall

portions was significant in capturing those with insufficient effort, χ^2 (2, $N = 139$) = 13.4, $p =$

.001. Chi-Square analysis on the strict effort criterion indicated that the RCFT cutoff score on the copy and recall portions was significant in capturing those displaying poor effort, χ^2 (2, $N = 139$) $= 23.6$, $p < .001$. Similar to Group One, proportions of those with suspect effort were identified among those with adequate copy and recall, poor copy or recall, and poor copy and recall. Because data were analyzed for both liberal and strict criteria for poor effort, a range of probabilities for poor effort was created for each of the three possible scores indicating performance on these two subtests. Once the base rate of poor effort was calculated for each measure, charted values were created so that clinicians may be able to assess the validity of examinees' neuropsychological test results and make decisions about further testing and diagnosis (see Table 2).

Discussion

Neuropsychological tests, which infer brain and cognitive states from behavioral performances, are profoundly compromised without the cooperation of examinees. Interpretations of cognitive capacity, obtained from neuropsychological tests such as the RCFT, can be made only when adequate effort is assured. Despite the necessity of effort tests, many neuropsychologists do not spend the time to routinely administer measures of effort, perhaps sacrificing accuracy in the process (Lees-Haley, Smith, Williams, & Dunn, 1996). The rationale for such omissions in most instances is likely because of efficiency, as many effort tests are time-consuming to administer. A long line of research has demonstrated that administering brief effort tests provide a poor compromise and are inadequate for most purposes. The current finding that performance on the RCFT, a test traditionally viewed as a measure of visuoconstructional ability, memory and executive function, is also sensitive to poor effort is valuable information for neuropsychologists who desire efficiency and accuracy. The RCFT, and indeed all valid

embedded measures, requires no additional administration time, and the results strongly indicate those individuals for whom the time and resources of formal effort testing would likely yield meaningful results.

RCFT performances, particularly when used in conjunction with other effort indicators, provide information on the *relative* risk of inadequate effort, and these suspicions can be confirmed or refuted with additional testing. Using independent samples, this study concluded that, although good performances on RCFT copy and immediate (3-minute) recall could be used to identify individuals unlikely to fail additional measures of effort, these scores could not definitively identify those who failed additional effort tests. Our results indicated that, even when strict operational criteria for poor effort were applied[4], among those with inadequate RCFT copy and recall accuracy, approximately one-third of individuals displayed poor effort. The others failing the RCFT might have comprised a group of individuals who either had an aberrant poor performance on the RCFT or had genuine cognitive impairments.

Certain aspects of embedded effort measures derived from standard neuropsychological tests have advantages over formal symptom validity tests. Embedded measures take less administration time, and, compared with well-known formal effort measures, they are potentially more resistant to coaching (Berry & Schipper, 2008; Brennan et al., 2009). In a sense, discovering the range of expected performances for given clinical conditions on standard neuropsychological tests can be a way for tests to perform "double duty": not only can they measure the integrity of different cognitive domains, but they can also check the validity of the test results during the session (Meyers, Galinsky, & Volbrecht, 1999; Meyers & Volbrecht, 2003). In addition, effort can fluctuate across a testing session, and patients who intentionally seek to feign cognitive deficits usually pick certain types of cognitive deficits to simulate

(Greiffenstein, Gola, & Baker, 1995; Meyers & Volbrecht, 1998). Although a single effort measure cannot assess effort across the time span and range of cognitive abilities assessed within the testing session, embedded effort measures found within specific tests allow neuropsychologists to assess effort in differing ways during multiple time points over the course of the evaluation.

Considering that embedded measures present the possibility for many tests in the neuropsychological battery to assess for effort, research on embedded measures will allow neuropsychologists to aggregate evidence and make appropriate decisions, as poor effort should never be concluded or excluded on the basis of performance on a single test (Heilbronner et al., 2009).

Additional research is needed to explore the use of the RCFT as an effort measure. Because of the dearth of visuospatial effort measures, it is opportune that a test as widely used as the RCFT can now be used as a relatively effective measure of effort. Furthermore, the RCFT has potential because of its deceptively easy copy task and its inclusion of a memory trial, as poor memory is a deficit often mimicked by those who malinger.

In the current project, we employed large samples of patients to whom we administered a relatively short version of the RCFT. The few studies that have explored the use of the RCFT as a measure of effort have explored the copy, immediate recall, recognition and delayed recall with small sample sizes and no cross-validation of their cut-off criterion. Additionally, most prior research has focused on recognition and delayed recall performance, which takes 30 additional minutes of administration time. Our study found that immediate recall and copy are useful embedded measures, and our findings are supported by a large sample size and are cross-validated on an independent sample.

Unlike other studies that have used confirmed malingering patients or healthy individuals instructed to simulate impairment, we defined poor effort according to performances on other effort tests. This methodology has disadvantages because it implies that failure on one effort test congruently aligns with failures on other tests. The TOMM and Digit Span are likely imperfect in accounting for all those whom actually displayed poor effort, as many studies show that each measure tends to give the benefit of the doubt to patients, and thus has good specificity and moderate sensitivity with regard to insufficient effort (Larrabee, 2007). As a result, the current analyses may underestimate the true incidence of poor effort because the cutoff score we derived on the RCFT captures only those individuals who displayed poor effort on the TOMM and Digit Span. It is quite possible that some proportion of the remaining individuals performed adequately on these tasks but displayed poor effort on a significant part of the remainder of the battery. For example, a poor association between two effort tests may mean that one or both of the tests are poor measures of effort. On the other hand, the tests may be identifying different individuals of the same target population and may, in fact, be complementary effort measures. The fact that our results imperfectly but significantly account for those showing poor effort is, however, encouraging for two reasons. First, we would not want the RCFT to identify exactly the same individuals performing with poor effort on the TOMM and Digit Span, because then it would be redundant with those tests. In the case of effort, aggregating distinctive information increases diagnostic accuracy by increasing sensitivity but not considerably affecting specificity, so it is worthwhile that the RCFT, TOMM, and Digit Span measure overlapping but also unique domains (Hunsley & Meyer, 2003; Larrabee, 2003; Victor, Boone, Serpa, Buehler, Ziegler, 2009). Most importantly, our findings justify additional research, because inasmuch as we know that a certain cutoff on the RCFT has successfully but solely captured those individuals

displaying poor effort on tests like the TOMM and Digit Span, we can predict that a study using simulation or malingering samples would enable us to identify even more of those individuals who are performing with suboptimal effort on the RCFT.

Primarily due to the abundance of false positive errors, the RCFT should not be used in isolation to conclude poor effort. In essence, individuals who perform well on the RCFT are probably putting forth adequate effort, and the group who performs poorly on the RCFT is likely comprised of a combination of individuals who are putting forth poor effort, those with genuine impairments, and those with normal relative weaknesses for visual-constructional processing. An indication of poor performance on the RCFT signifies to the clinician that additional testing is needed to determine the reason for their poor performance. Aggregating across multiple embedded measures of effort throughout a battery not only leads to greater accuracy, but also provides an ongoing monitoring of effort (Larrabee, 2008; Nelson et al., 2003; Slick, Sherman, & Iverson, 1999). Effort measures are critical in the neuropsychological battery to ensure validity and the informed and confident diagnoses of neuropsychologists. Their responsible use and interpretation will result in more appropriate and just care for patients.

References

Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting

incomplete effort with digit span from the Wechsler Adult Intelligence Scale, third

edition. *The Clinical Neuropsychologist, 20*, 513-523.

Berry, D. R., & Schipper, L. J. (2008). Assessment of feigned cognitive impairment using

standard neuropsychological tests. In R. Rogers, R. Rogers (Eds.), *Clinical assessment of

malingering and deception (3rd ed.)* (pp. 237-252). New York, NY US: Guilford Press.

Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. (2009). The

vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist*,

23(2), 314-328.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., & ... Silver, C.

H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN

Policy & Planning Committee.

Delis, D. C. & Wetter S. R. (2007). Cogniform Disorder and Cogniform Condition: proposed

diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22*,

pp. 589–604.

Faust, D., Hart, K. J., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children

to fake believable deficits on neuropsychological testing. *Journal of Consulting and

Clinical Psychology*, 56(4), 578-582.

Green, P., Allen, L., & Astner, K. (1996). The Word Memory Test: A users guide to the oral and

computer-administered forms, US version 1.1. Durham, NC: Cognisyst.

Green P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on

test scores than severe brain injury in compensation claimants. *Brain Injury*, 15(12),

1045-1060.

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., Baker, W., & Johnson-Greene, D. (2002). Actual versus self-reported

scholastic achievement of litigating postconcussion and severe closed head injury

claimants. *Psychological Assessment*, 14(2), 202-208.

Greiffenstein, M. F., Gola, T., & Baker, W. J. (1995). MMPI-2 validity scales versus domain

specific measures in detection of factitious traumatic brain injury. *The Clinical

Neuropsychologist*, 9(3), 230-240.

Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2002).

Detecting malingered performance with the Wisconsin Card Sorting Test: A preliminary

investigation in traumatic brain injury. *The Clinical Neuropsychologist*, 16(2), 179-191.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American

Academy of Clinical Neuropsychology consensus conference statement on the

neuropsychological assessment of effort, response bias, and malingering. *The Clinical

Neuropsychologist*, 23(7), 1093-1129.

Hunsley, J., & Meyer, G. J. (2003). The Incremental Validity of Psychological Testing and

Assessment: Conceptual, Methodological, and Statistical Issues. *Psychological

Assessment*, 15(4), 446-455.

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. (2002). Detecting exaggeration and

malingering with the Trail Making Test. *The Clinical Neuropsychologist*, 16(3), 398-

406.

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., & …Zeller,

M. A. (2010). The Warrington Recognition Memory Test for Words as a measure of

response bias: Total score and response time cutoffs developed on "real world" credible

and noncredible subjects. *Archives of Clinical Neuropsychology*, 25(1), 60-70.

Lange, R. T., Iverson, G. L., & Rose, A. (2010). Post-concussion symptom reporting and the

"good-old-days" bias following mild traumatic brain injury. *Archives of Clinical

Neuropsychology*, 25(5), 442-450.

Larrabee, G. J. (2003). Detection of Malingering Using Atypical Performance Patterns on

Standard Neuropsychological Tests. *The Clinical Neuropsychologist*, 17(3), 410-425.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits.* Oxford: Oxford

University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of

malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22(4),

666-679.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic

neuropsychological test usage: An empirical survey. *Archives of Clinical

Neuropsychology*, 11(1), 45-51.

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

Press.

Lu, P.H., Boone, K.B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth

Complex Figure Test and the Meyers and Meyers Recognition Trial in the Detection of

Suspect Effort. *The Clinical Neuropsychologist*, 17(3), 426–440.

McCrea, M. (2008). Mild traumatic brain injury and postconcussion syndrome: the new evidence

base for diagnosis and treatment. Oxford: Oxford University Press.

Meyers, J. E., Galinsky, A. M., & Volbrecht, M. (1999). Malingering and mild brain injury: How

low is too low?. *Applied Neuropsychology*, 6(4), 208-216.

Meyers, J. E., & Meyers, K. R. (1995). Rey Complex Figure Test and Recognition Trial:

Professional manual. Lutz, FL: Psychological Assessment Resources.

Meyers, J. E., & Volbrecht, M. (1998). Validation of memory error patterns on the Rey Complex

Figure and Recognition Trial. *Applied Neuropsychology*, 5(3), 120-131.

Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection

methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18(3), 261-

276.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal

Learning Test in the detection of incomplete effort in neuropsychological

evaluation. *Psychological Assessment*, 7(4), 463-471.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships

Between Eight Measures of Suspect Effort. *The Clinical Neuropsychologist*, 17(2), 263-

272.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical

neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA

Division 40 members. *Archives of Clinical Neuropsychology,* 20, 33-66.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation

experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*,

10(1), 10-20.

Ruff, R. M., Wylie, T., & Tennant, W. (1993). Malingering and malingering-like aspects of mild

closed head injury. *The Journal of Head Trauma Rehabilitation*, 8(3), 60-73.

Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-
   Item Memory Test to detect malingered amnesia. *Psychological Assessment: A Journal of*
   Consulting and Clinical Psychology, 3(4), 667-672.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). VSVT: Victoria Symptom Validity
   Test, version 1.0, professional manual. Odessa, FL: Psychological Assessment
   Resources, Inc.

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

Spencer, R. J., Seymour, C., Wright, S. L., Ford, A. I., Vanorsdall, T., & Bieliauskas, L. A.
   (under review). Factors predicting Rey Fifteen-Item Test performance among veterans
   undergoing inpatient medical rehabilitation.

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

Suhr, J. A., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor
   speed tasks as indicators of malingering. In G. L. Larrabee (Ed.), *Assessment of*
   malingered neuropsychological deficits. (pp. 131-170). New York: Oxford University
   Press.

Suhr, J. A., & Gunstad, J. (2002). 'Diagnosis threat': The effect of negative expectations on
   cognitive performance in head injury. *Journal of Clinical and Experimental*
   Neuropsychology, 24(4), 448-457.

Suhr, J. A., & Gunstad, J. (2005). Further exploration of the effect of 'diagnosis threat' on

cognitive performance in individuals with mild head injury. *Journal of the International Neuropsychological Society*, 11(1),

Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, 19(4), 500-514.

Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM).* North Tonawanda, NY: Multi Health Systems.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260-268.

Victor, T. L., Boone, K. B., Serpa, J., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, 23(2), 297-313.

Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale – Fourth Edition.* San Antonio, TX: Pearson.

Whitney, K. A., Davis, J. J., Shepard, P. H., Bertram, D. M., & Adams, K. M. (2009). Digit Span Age Scaled Score in middle-aged military veterans: Is it more closely associated with TOMM failure than Reliable Digit Span? *Archives of Clinical Neuropsychology*, 24(3), 263-272.

Footnotes

[1]It is easier to feign disability than to simulate extraordinary ability.

[2]This effect of over-estimating past functioning also holds true for education achievement (Greiffenstein et al., 2002).

[3]An example would be of a patient who could only repeat three digits, but whose memory performance was above average. It does not follow that someone could retain more information than they can learn.

[4]Most operational criteria for determining poor effort require failure of two or more tests. The analyses in Group Two included *only* two effort tests, and therefore, it can be reasoned that inclusion of more effort measures would have resulted in more participants being classified as performing with poor effort.

Author Note

Nazanin Maghsoodi, Department of Psychology, University of Michigan, Ann Arbor.

Table 1

Performance of Group 1 on RCFT copy and recall among those failing Digit Span (cut-off of 6/7).

RCFT Score	Digit Span	
	Pass (Good Effort)	Fail (Poor Effort)
2 (Pass copy and Recall)	102 (92.7%)	8 (7.3%)
1 (Fail copy or Recall)	52 (73.2%)	19 (26.8%)
0 (Fail copy and Recall)	21 (63.6%)	12 (36.4%)
Total	175 (81.8%)	39 (18.2%)

Table 2

Cross-validation of RCFT copy and recall among those in Group 2 failing strict and liberal

effort measures involving Digit Span and TOMM.

RCFT Score	Effort, Liberal		Effort, Strict	
	Pass (Good Effort)	Fail (Poor Effort)	Pass (Good Effort)	Fail (Poor Effort)
2 (Pass copy and Recall)	57 (89.1%)	7 (10.9%)	63 (98.4%)	1 (1.6%)
1 (Fail copy or Recall)	33 (71.7%)	13 (28.3%)	44 (95.7%)	2 (4.3%)
0 (Fail copy and Recall)	16 (55.2%)	13 (44.8%)	20 (69.0%)	9 (31.0%)
Total	106 (76.3%)	33 (23.7%)	127 (91.4%)	12 (8.6%)