THE SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES

by

Daniel M. Turpin

An Abstract
of a thesis in partial fulfillment
of the requirements for the degree of
Master of Science
in the Department of Psychology
University of Central Missouri

November, 2011
ABSTRACT

by

Daniel M. Turpin

The 21-Item Test was administered to 20 healthy elderly individuals dwelling in the community and 53 individuals residing in a nursing home. The 20 healthy elderly individuals had means for age, education, and Short Portable Mental Status Questionnaire (SPMSQ) of 76.25 years ($SD = 7.19$), 14.30 years ($SD = 3.89$), and .30 ($SD = 4.70$), respectively. The second group of 53 nursing home residents had means for age, education, and SPMSQ of 82.79 years ($SD = 8.56$), 11.14 years ($SD = 2.46$), and 3.17 ($SD = 2.46$), respectively. The clinical cut-offs for the forced-choice component of the 21-Item Test proposed by Iverson (1998) may be insensitive to age and cognitive impairment. Not a single individual scored in the range of biased responding. Thus, the 21-Item test is unlikely to produce false positive results in elderly patients with obvious cognitive deficits.
THE SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES

by

Daniel M. Turpin

A Thesis
presented in partial fulfillment
of the requirements for the degree of
Master of Science
in the Department of Psychology
University of Central Missouri
November, 2011
ACKNOWLEDGMENTS

First, I wish to express my heartfelt appreciation to Dr. Joseph Ryan for your commitment to my education and this project. You are an invaluable thesis chair, mentor, educator, and friend. I appreciate your patience and guidance through this project. My continuing education in the field of psychology is solely due to the knowledge and passion you expressed as my educator. Thanks for believing in me, making me laugh, and always keeping your office door open. Special acknowledgment is also given to Drs. Jon Smith and David Kreiner for their commitment to this thesis project. I would also like to thank Tina Walker for assisting in the interpretation of data and her willingness to always provide a helping hand. Finally, my deepest appreciation is further extended to my parents Michael and Pamela, and to my sister, Kim. I am truly blessed to have such a caring family. Thank you for not giving up on me and encouraging me every step of the way.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1: NATURE AND SCOPE OF THE STUDY</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>1</td>
</tr>
<tr>
<td>Need for the Study</td>
<td>2</td>
</tr>
<tr>
<td>Thesis of the Study</td>
<td>2</td>
</tr>
<tr>
<td>Procedural Overview</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER 2: REVIEW OF THE LITERATURE</td>
<td>4</td>
</tr>
<tr>
<td>Definitions of Malingering</td>
<td>4</td>
</tr>
<tr>
<td>Prevalence of Malingering</td>
<td>6</td>
</tr>
<tr>
<td>Floor Effect</td>
<td>9</td>
</tr>
<tr>
<td>Performance Curve</td>
<td>16</td>
</tr>
<tr>
<td>Magnitude of Error</td>
<td>19</td>
</tr>
<tr>
<td>Atypical Presentation</td>
<td>20</td>
</tr>
<tr>
<td>Symptom Validity Testing</td>
<td>22</td>
</tr>
<tr>
<td>CHAPTER 3: METHODOLOGY</td>
<td>33</td>
</tr>
<tr>
<td>Participants</td>
<td>33</td>
</tr>
<tr>
<td>Materials</td>
<td>36</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

Table | Page
---|---
1. Diagnostic Criteria for the Identification of Malingered Neurocognitive Dysfunction | 7
2. Malingering Probability Classifications Based on Vocabulary - Digit Span Difference Scores | 17
3. Diagnostic Breakdown of Participants in the Nursing Home Group | 34
4. Medications Prescribed in the Nursing Home Group | 35
5. The Short Portable Mental Status Questionnaire | 38
6. Scoring Sample for the Temporal Orientation Scale | 40
10. Performance Based on Statistical and Clinical Standards of Interpretation of the Community and Nursing Home Samples | 48
11. Correlations for the Total Sample on Demographic and Test Variables | 49
12. Predicted Forced-Choice Scores from SPMSQ Errors Scores | 51
13. 21-Item Test Forced-Choice Scores for the Nursing Home Sample by SPMSQ Classification of Intellectual Impairment | 52
CHAPTER 1
NATURE AND SCOPE OF
THE STUDY

Purpose of the Study

The National Academy of Neuropsychology recommends that an assessment of response bias be included in every neuropsychological evaluation (Bush et al., 2005). To accomplish this goal, most neuropsychologists utilize one or more symptom validity measures in their assessment batteries. These instruments are used to detect sub-optimal effort, lack of motivation to perform the task at hand, and/or outright malingering. Symptom validity assessment is most crucial in cases that involve litigation and/or disability determination, but is also important in routine assessment situations that result in diagnoses or treatments for individual patients. For example, in capital murder cases the defendant may be motivated to fake mental retardation or some other cognitive disability in order to escape life imprisonment or capital punishment. In cases of personal injury litigation following head trauma, a patient may exaggerate his or her acquired deficits in order to obtain a large cash settlement from an insurance company. Finally, there are many instances in clinical practice where persons referred for intellectual and neuropsychological assessment have little investment in good performance and therefore put forth minimal effort resulting in poor scores that do not reflect their actual levels and patterns of ability. This is maybe especially true among persons of advanced age who frequently find psychological and neuropsychological testing to be of little relevancy to their current life situations. Therefore, the purpose of this study is to determine the specificity of a well known but little researched symptom validity test (i.e., 21-Item Test (Iverson, Franzen, & McCracken, 1991) in two elderly samples.
Need for the Study

This investigation will be important for experienced clinicians who utilize the 21-Item Test as a measure of biased responding. Both samples used in this study were elderly participants who were at least 64 years of age. These individuals were administered the 21-Item Test in order to provide specificity data of elderly performance on this task. Thus, clinicians can compare these data to those of individuals who take the 21-Item Test in the future. Currently, there is no 21-Item Test normative data for the elderly population. This is the first study that investigates elderly performance on the 21-Item Test.

Thesis of the Study

This study was designed to assess the specificity of the 21-Item Test in two groups of elderly volunteers. One group consisted of healthy persons living in the community and the second group was composed of individuals, with or without cognitive dysfunction, residing in nursing homes. All participants were administered a brief cognitive battery, in addition to the 21-Item Test, which included the Temporal Orientation Scale (TOS; Benton, Sivan, Hamsher, Varney, & Spreen, 1994) and the Short Portable Mental Status Questionnaire (SPMSQ; Pfeiffer, 1975). Other variables of interest were participants’ age and years of education which were correlated with scores from the 21-Item Test and the brief cognitive battery. The study addressed the following hypotheses:

1. Normal elderly participants residing in the community will perform in a non-biased manner on the 21-Item Test.

2. Elderly participants residing in a nursing home (with or without cognitive dysfunction) will perform in a non-biased manner on the 21-Item Test.
3. Scores on the 21-Item Test for the total sample of participants residing in the community and in nursing homes will not be significantly associated with age, education, or scores on the Temporal Orientation Scale and Short Portable Mental Status Questionnaire.

Procedural Overview

Participants in this research were elderly volunteers residing in the community or in nursing homes. Community-dwelling elderly volunteered through a local church program designed for senior citizens. All participants were at least 64 years of age and were predominantly Caucasian. Each participant was given an informed consent form to read and sign. Nursing home administrators allowed the research to take place in their facilities. Medical diagnoses and medication regiments were provided by nursing home staff. Demographic information was also obtained. Each participant answered questions from the SPMSQ. Each participant answered questions regarding orientation to time from the TOS and completed the free recall and forced-choice components of the 21-Item Test.
Neuropsychological assessment is used in a variety of business, educational, clinical, and forensic settings. In any of the aforementioned settings, the assessment professional must be aware of each examinee’s test-taking motivation. For example, in the forensic context, an individual may feign brain impairment to evade criminal prosecution, whereas in civil litigation an individual may falsely claim to have a memory deficit secondary to a minor head injury sustained in a traffic accident. Over the last decade, the literature has been inundated with research concerning malingering, which may be due to the many life-altering conclusions that can result from neuropsychological assessment and the actual prevalence of malingering itself. The following literature review will focus on the definition of malingering, its prevalence, and five techniques used for its detection.

Definitions of Malingering

The malingering literature may seem cumbersome because there are various definitions and diagnostic criteria available (e.g., Slick, Sherman, & Iverson, 1999). The most commonly used definition is that provided by the *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* (DSM-IV; American Psychiatric Association, 1994). In the DSM-IV, 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, 1994, p. 683). This definition is widely accepted but is not without flaw. First, the definition relies too heavily on the idea of volition and reward
without providing specific criteria to help guide the neuropsychological examiner. Discerning someone’s true volition or incentive for committing a certain act is much more complex than it may seem; as Slick et al. (1999) state:

“Dichotomous diagnostic criteria such as external versus psychological incentives, and volitional versus unconscious behavior are easy to write into definitions, but in practice, judging the degree to which a behavior is volitional is fraught with uncertainty, and disentangling which incentive is primary in cases where external and internal incentives coexist - as is often the case - can be exceedingly difficult (p. 546).”

Also, the DSM-IV does not provide a clear and concise set of diagnostic criteria for malingering. The current DSM-IV criteria for malingering are as follows:

1. Medicolegal context of presentation (e.g., the person is referred by an attorney to the clinician for examination)
2. Marked discrepancy between the person’s claimed stress or disability and the objective findings
3. Lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen
4. The presence of Antisocial Personality Disorder (American Psychiatric Association, 1994)

The DSM-IV’s definition of malingering provides little to no direction in terms of malingering in the context of a neuropsychological evaluation (Slick et al., 1999). As one can see, the DSM’s guidance for malingering is limited and has caused neuropsychological experts to propose many alternate definitions and criteria. Slick et al. (1999) define Malingering of Neurocognitive
Dysfunction (MND) and divide it into three categories. These categories consist of: Possible MND, Probable MND, and Definite MND. The examinee must satisfy specific criteria to fall into one of the three categories. The Slick et al. (1999) diagnostic categories and criteria of MND are presented in Table 1. Slick et al. (1999) define MND as the:

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

For the purposes of this study, we used the aforementioned definition of malingering as proposed by Slick et al. (1999). It is reasonably complete, and the criteria are detailed enough to guide the examiner in a way that is systematic and understandable. Although the definition itself does use the concepts of volition and incentive, the criteria proposed are detailed and relatively easy to apply.

**Prevalence of Malingering**

Obtaining accurate base rate information on malingering is a difficult task because probable malingerers tend to successfully evade detection (Slick, Tan, Strauss, & Hultsch, 2004). Malingering base rates are also hard to acquire because malingering itself may be misdiagnosed as legitimate injury. As Slick et al. (1999) state, “the process of diagnosing malingering remains difficult and largely idiosyncratic” (p.545).
Table 1

*Diagnostic Criteria for the Identification of Malingered Neurocognitive Dysfunction (MND)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite MND</td>
<td>Substantial external incentive</td>
</tr>
<tr>
<td></td>
<td>Definite negative response bias</td>
</tr>
<tr>
<td></td>
<td>Behavior not completely accounted for by other factors</td>
</tr>
<tr>
<td>Probable MND</td>
<td>Substantial external incentive</td>
</tr>
<tr>
<td></td>
<td>≥ 2 types of evidence from neuropsychological measures or 1 type of</td>
</tr>
<tr>
<td></td>
<td>evidence from neuropsychological measures and ≥ 1 types of evidence from</td>
</tr>
<tr>
<td></td>
<td>self-report measures.</td>
</tr>
<tr>
<td></td>
<td>Behavior not completely accounted for from definite negative response</td>
</tr>
<tr>
<td></td>
<td>bias or other factors.</td>
</tr>
<tr>
<td>Possible MND</td>
<td>Substantial external incentive</td>
</tr>
<tr>
<td></td>
<td>Evidence from self-report measures</td>
</tr>
<tr>
<td></td>
<td>Behavior not accounted for by other factors (e.g. psychiatric or</td>
</tr>
<tr>
<td></td>
<td>neurological), or criteria for previous categories are met except for</td>
</tr>
<tr>
<td></td>
<td>by other factors. In instances where these are met, etiologies not</td>
</tr>
<tr>
<td></td>
<td>ruled out should be specified.</td>
</tr>
</tbody>
</table>

*Note.* Table adapted from Slick et al. (1999). MND = Malingered Neurocognitive Dysfunction.
Despite these setbacks, studies have provided valuable information revealing the frequency of malingering in real life situations.

An early study by Heaton, Smith, Lehman, and Vogt (1978) found prevalence as high as 64% among litigating individuals undergoing neuropsychological evaluation. The literature now contains empirical studies that report prevalence rates in sequential series of patients (Mittenberg, Patton, Canyock, & Condit, 2002). For example, Larrabee (2003) noted a malingering prevalence of about 40% in 1,363 compensation seeking instances of mild head injury.

Mittenberg et al. (2002) surveyed 131 neuropsychologists regarding the actual prevalence of malingering in their practices. The findings of probable malingerers were as follows: 29% of personal injury cases, 30% of disability cases, 19% of criminal cases, and 8% of medical cases. The base rate estimates of this study were calculated on 33,531 cases (Mittenberg et al., 2002). It is important to note that the three aforementioned studies on malingering prevalence used variable methods to identify malingerers. Heaton et al. (1978) detected malingering by analyzing variables on neuropsychological testing and the Minnesota Multiphasic Personality Inventory. Larrabee (2003) utilized the definition of Malingered Neurocognitive Dysfunction proposed by Slick et al. (1999). Finally, Mittenberg et al. (2002) obtained prevalence data by surveying practicing neuropsychologists regarding probable malingering cases.

Although base rate information on malingering may be difficult to collect, it is easy to see that malingering can be a significant problem in the realm of clinical neuropsychology. The bulk of referrals for neuropsychological evaluation stem from the forensic context (e.g. criminal prosecution, workers compensation, personal injury litigation), and those being referred may
receive substantial financial compensation if they can demonstrate either existing or faked cognitive impairment (Slick et al., 1999). With such high prevalence rates, it is no surprise that researchers and clinicians have developed strategies to assist with the detection of probable-malingerers. Throughout the last decade, the literature has been inundated with malingering research studies. Through this research, five strategies for detecting feigned impairment have emerged. The remainder of this review will focus on the following detection strategies for malingering: floor effect, performance curve, magnitude of error, atypical presentation, and symptom validity testing.

Floor effect can be observed when a normal person fails a measure that a severely impaired person could easily pass. Performance curve refers to a malingerer not taking into consideration discrepancies in item difficulty when determining which items to fail and thus performing in a nonsensical manner. Magnitude of error refers to the extent to which a response is inaccurate. Atypical presentation is defined as the malingerer performing differently on tests that measure similar abilities or the same test. Each detection method will be described in detail (Rogers, Harrell, & Liff, 1993).

**Floor Effect**

Research shows that malingering may effectively be observed when an individual fails extremely simple tasks. This approach is known as the “floor effect” (Rogers et al., 1993) and involves an apparently able person failing a task that even a severely impaired individual can perform correctly with little difficulty. For example, Wiggins and Brandt (1988) asked amnesic patients and simulated malingerers made up of university students and community volunteers 25 autobiographical questions. These were simple autobiographical questions such as “What is your
name?” and “What is your age?” (Wiggins & Brandt, 1988, p. 65). More simulators answered incorrectly than did the amnesic patients on 24 of the 25 questions. For example, when asked to state their name, 25 simulators answered incorrectly while every amnesic gave correct responses. Also, 35 simulators claimed they forgot their age while every amnesic responded correctly. Many neuropsychological measures demonstrate the floor effect. Two popular tests are the Rey 15-Item Test and the Digit Span subtest that appears in multiple editions of the Wechsler Scales of adult intelligence.

**Rey’s 15-Item Test.** This widely used test was developed in 1964 by the French neurologist Andre Rey. Originally, it was referred to as the Rey Visual Memory Test. However, due to its design it is now more commonly known as the Rey Memory Test or the Rey 15-Item Test. It will be referred to as the Rey 15-Item Test throughout the remainder of this review. This instrument was developed as a test to assess the validity of memory impairment. The test uses an 8.5 x 11 sheet of paper on which are printed 15 items (e.g. A B C in the first row, 1 2 3 in the second row, a b c in the third row, ○ □ △ in the fourth, and I II III in the last row). After a 10-second exposure, the 15 items are removed and the examinee is asked to draw the items from memory (Yamaguchi, Ryan, & Kreiner, 2001).

The underlying principle is that the characters and figures are designed to be extremely easy to remember. However, the instructions give the examinee the impression that the task is extremely difficult because he or she is told that all 15 items have to be remembered (Bernard & Fowler, 1990). Although the examinee may feel overwhelmed having to memorize 15 items in 10 seconds, the patterns are so repetitive and easy that only a few pieces of information must actually be registered (Lee, Loring, & Martin, 1992).
Empirical studies demonstrating the simplicity of the Rey 15-Item Test are numerous. For instance, Lee et al. administered the Rey 15-Item to 100 individuals with temporal lobe epilepsy (TLE) and obvious symptoms of memory impairment. Forty-two percent of the sample obtained a perfect score of 15 on the task, and 96 percent recalled 8 items or above. Similarly, Bernard and Fowler (1990) administered the test to 18 patients with documented brain damage. There were nine with head injury, five with cerebral vascular accidents, two with brain tumors, one with acute hydrocephalus, and one with Alzheimer’s disease. In terms of group performance, adequate scores were obtained on the Rey 15-Item Test as evidenced by a mean score of 11.4 ($SD = 2.7$) designs correctly reproduced. Finally, Goldberg and Miller (1986) administered the Rey 15-Item Test to 50 psychiatric inpatients (aged 18 to 63 years) and 16 intellectually deficient individuals (aged 18 to 53 years) and found that only individuals with severe impairment recalled fewer than three of the five rows. Thus, malingering should be considered when an individual fails to remember at least nine of the 15 items.

The main concept behind the Rey 15-Item Test is that the malingerer can demonstrate inadequate performance on an easy task that “all but the most severely brain damaged or retarded patients perform easily” (Lezak, 1983, p. 618). The Rey 15-Item Test becomes a useful detection strategy for malingering when cut-off scores are utilized. Cut-off scores give the neuropsychologist an idea of how examinees should be performing relative to other groups (normals, demented persons, brain injured, etc), and performance below these scores should raise a red flag for probable malingering. As one can see in the two aforementioned studies, cut-off scores are usually obtained by observing performance in samples with cognitive dysfunction. In the previously mentioned studies, the most discriminating cut-off scores with brain damaged
patients and those with memory loss due to temporal lobe epilepsy were determined. The
researchers reported that a score between seven and nine items correct is useful. Hence, if an
individual with apparently normal cognitive functioning obtained a score less than nine, this
might reflect feigned memory impairment.

It is important to note that there are differing recommendations regarding the correct cut-
off scores for the Rey 15-Item Test. Lezak (1983), Bernard, and Fowler (1990), and Goldberg
and Miller (1986) found a cut-off score of nine to be discriminating. However, Lee, Loring, and
Martin (1992) claim that, to avoid false positives, the cut-off should be seven based on their
findings with temporal lobe epilepsy patients suffering from memory loss.

The Rey 15-Item Test has been frequently researched in terms of its specificity and
sensitivity. Specificity is defined as “the proportion of persons that do not have the condition
who are predicted by the test to not have it” (Smith, Cerhan, & Ivnik, 2003, p. 280). Sensitivity is
defined as “the proportion of persons who have the condition that are predicted by the test to
have it” (Smith et al., 2003, p. 280). Guilmette, Hart, Giuliano, and Leininger (1994) obtained
specificity and sensitivity data based on different cut-off scores. The Rey 15-Item Test was
administered to 20 brain-damaged participants, 20 inpatient psychiatric participants, and 20
normal persons instructed to feign memory impairment. Using a cut off score of nine, they found
a specificity of 70% and a sensitivity of 25%. Similarly, for a cut off of 8 the specificity was
72.5% and sensitivity was 15%. Finally, a specificity of 75% and a sensitivity of 5% were found
for a cut-off of seven.

Another study of seven individuals claiming to have cognitive impairment due to closed
head trauma, found a specificity of 71% and a sensitivity of 57% for the Rey 15-Item Test
The previously mentioned studies indicate that the Rey 15-Item Test has moderate specificity, but low sensitivity. Thus, the Rey 15-Item Test will produce many false negatives. However, when it is performed poorly (< 7 correct) by an individual with apparently normal cognitive functions, the probability of malingering is high.

Due to its simplicity, the Rey 15-Item Test is a prime example of an examination that demonstrates the floor effect. Rogers et al. (1993) explain that the floor effect is a strategy that may “give rise to a troubling ethical question in its administration” (p.260). The authors discuss the work of Drob and Berger (1987) that suggests that stressing the difficulty involved in recalling fifteen items in the instructions of the test is misleading to the examinee. Rogers et al. (1993) also question the ethical issue of deceiving a possible malingerer. They discuss the question, “Is misleading suspected malingerers either ethically proper or clinically necessary?” (p.261). It is beyond the scope of this review to provide an answer, but the inquiry itself would be an interesting research topic. Regardless of the ethical issues raised by the use of the floor effect approach, it continues to be a strategy utilized with the Wechsler Adult Intelligence Scales Digit Span subtest.

**Digit Span.** In this test, the individual is read aloud a list of random numbers at a rate of one number per second. The test starts with two numbers and increases in length until the examinee commits recall errors. The longest string of digits to recall is nine. After the examinee is read each set of numbers, he or she recalls them in order. This is the Forward Digit Span component of the subtest. The second component, Backward Digit Span, involves recalling a series of random numbers in reverse order. The longest string of digits to recall is eight. This
subtest measures immediate auditory recall and is viewed as a measure of working memory (Wechsler Adult Intelligence Scale – Revised, Wechsler, 1981).

Currently, a large body of literature reports that suboptimal performance on the Digit Span subtest may be an indicator of malingering. Similar to the Rey 15-Item Test, the forward part of Digit Span is an example of the floor effect strategy due to its simplicity. A score in the normal range for Forward Digit Span has been described as seven ± 2 digits (Iverson & Franzen, 1996). Digit Span scores for amnesic individuals (e.g., those with encephalitis or Korsakoff’s syndrome) have been reported to fall within this range (Baddeley & Warrington, 1970). For example, Butters’ and Cermak’s (1980) sample of 9 Korsakoff patients yielded a mean Digit Span scaled score of 9.44 (which translates to about 6 or 7 digits forward). According to Iverson and Franzen (1996), it is important to note that “although these patients typically perform within the normal range, they often fall on the lower end of the distribution (i.e., 5 to 6 digits recalled)” (p.41). In fact, in an earlier study by Black (1986) the mean Digit Span forward score was 5.9 ($SD = 1.4$) and the mean Digit Span backward score was 4.0 ($SD = 1.3$) in a sample of 162 brain-damaged individuals.

The claim that suppressed performance on Digit Span forward can identify malingerers is reinforced by several studies. For example, Binder and Willis (1991) reported reduced Digit Span performance in a sample of litigating patients with mild head traumas who had been identified as deliberately performing poorly on a forced-choice symptom validity task known as the Portland Digit Recognition Test (Binder, 1993; Iverson & Tulsky, 2003). Poor performance on a forced-choice symptom validity test is definitive evidence of intentional suboptimal effort.
Iverson and Tulsky (2003) provided cumulative percentages of low Digit Span scores in multiple clinical groups. Two samples from this study included patients with Traumatic Brain Injury and patients with Korsakoff’s syndrome. No patients from either sample scored below five on Digit Span forward. Thus, scores below five on Digit Span forward may be used to detect malingering. Also, Iverson and Franzen (1994) found that Digit Span performance distinguished Federal prison inmates and students instructed to malingering from individuals with head injuries. Iverson and Franzen (1996) also found that experimental-malingerers performed much worse than patients with memory impairment on the Digit Span subtest. This study utilized scales scores which have a mean of 10 and a standard deviation of 3. Iverson and Franzen (1996) used a cutting age-corrected scaled score of four (which roughly translates to a forward raw score of 3 or 4). With this score, the researchers correctly identified 100% of the patients with memory impairment and 77.5% of the malingerers. The researchers used the same cut-off score and correctly identified “82.5% of experimental malingerers, 100% of normal controls, and 95% of patients with closed-head injuries” (Iverson & Franzen, 1996, p. 48). A cut-off score of three has also been used to detect suboptimal effort. For example, Franzen and Martin (1996) administered the Digit Span forward task of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) to 37 clinical psychology graduate students instructed to malingering a memory deficit. A cut-off score of three was used for Digit Span forward performance and correctly identified 29 participants as malingerers even though only 2 participants were identified as malingerers by a Symptom Validity Test (e.g., the 21-item Test). Digit Span difference scores have also been used to separate malingerers from impaired individuals. For example, Mittenberg, Theroux-Fichera, Zielinski, and Heilbronner (1995) administered the WAIS-R to one group of
67 non-litigating head-injured patients and another group of 67 individuals who were instructed to feign head-trauma symptoms. Each group was matched on Full Scale IQ, age, and occupation. Using scores from these two groups, Mittenberg et al. (1995) took the Vocabulary and Digit Span subtests of the WAIS-R and calculated scaled score difference scores (Vocabulary – Digit Span) that coincide with malingering probabilities. These probabilities are presented in Table 2.

**Performance Curve**

Early literature by Goldstein (1945) gave the initial description of the performance curve strategy and based it on the numbers of “easy items failed and difficult items passed” (p.105). This strategy is based on the assumption that a malingerer will not account for discrepancies in item difficulty in determining which items to fail. If this assumption is correct, then detecting malingerers is as simple as comparing their performance to authentic patients (Rogers et al., 1993). Presumably, the performance curve strategy is successful when an examinee’s performance on neuropsychological measures contradicts itself. That is, if an individual achieves a score of 3 to 4 on an immediate recall portion of a memory measure and then produces zero correct answers during recognition, there is an inconsistency that makes no sense and may be indicative of malingering. Recognition memory is an easier task than free recall. Thus, recognition memory and free recall score differences are analyzed during interpretation of the Rey Word Recognition Test (WRT; Rey, 1941).

**Rey Word Recognition Test.** This WRT is a tool to detect malingered memory complaints (Nitch, Boone, Wen, Arnold, & Alfano, 2006). During this test, the participant is first read 15 words. Then, he or she is presented with a printed page containing the 15 words paired with 15 distractor items and is instructed to circle the words previously read aloud. The examiner
Table 2

*Malingering Probability Classifications Based on Vocabulary - Digit Span Difference Scores.*

<table>
<thead>
<tr>
<th>Vocabulary – Digit Span Difference Score</th>
<th>Probability of Malingering</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>.99</td>
</tr>
<tr>
<td>9</td>
<td>.95</td>
</tr>
<tr>
<td>8</td>
<td>.90</td>
</tr>
<tr>
<td>7</td>
<td>.85</td>
</tr>
<tr>
<td>6</td>
<td>.80</td>
</tr>
<tr>
<td>5</td>
<td>.75</td>
</tr>
<tr>
<td>4</td>
<td>.70</td>
</tr>
<tr>
<td>3</td>
<td>.65</td>
</tr>
<tr>
<td>2</td>
<td>.60</td>
</tr>
<tr>
<td>1</td>
<td>.55</td>
</tr>
<tr>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>-1</td>
<td>.45</td>
</tr>
<tr>
<td>-2</td>
<td>.40</td>
</tr>
<tr>
<td>-3</td>
<td>.35</td>
</tr>
<tr>
<td>-4</td>
<td>.30</td>
</tr>
<tr>
<td>-5</td>
<td>.25</td>
</tr>
<tr>
<td>-6</td>
<td>.20</td>
</tr>
</tbody>
</table>

*Note.* Table adapted from Mittenberg et al. (1999).
then compares the participant’s recognition score to a free recall trial of a new 15-item word list. If the recognition score is equal to or less than the free recall score, the participant’s performance may reflect intentional failure because it is more difficult to freely recall words than it is to recognize them (Nitch et al., 2006). Numerous studies have shown that the WRT is highly effective at detecting feigned cognitive deficits. For example, Greiffeinstein, Baker, and Gola (1994) administered the WRT to a sample consisting of 33 Traumatic Brain Injury (TBI) patients, 30 patients with Persistent Post-Concussive Syndrome (PPCS), and 43 probable malingerers. The results of this study revealed that probable malingerers’ scores on the WRT were an average of nearly 3 points less than individuals with TBI and an average just over 3 points less than the PPCS group. This study also reported WRT specificity as high as 93% (Greiffenstein et al., 1994). The literature highlights other neuropsychological instruments that implement the performance curve strategy. For example, Lezak (1983) recommends applying the performance curve strategy to Rey’s Dot Counting Test.

**Ungrouped Dots.** Rey’s (1941) Dot Counting Test (DCT) actually consists of two parts: Ungrouped Dots and Grouped and Ungrouped Dots. Both tasks are similar and seem to be adequate examples of implementing the performance curve strategy. The Ungrouped Dots task was designed for individuals complaining of specific perceptual errors (e.g. vision) or cognitive impairment (Lezak, 1983). This initial task uses six 3 x 5 inch cards that each have a specific number of dots. The dots are arranged in a random sequence. These cards, labeled cards 1-6, contain 11, 19, 15, 23, 27, and 7 dots, respectively. After each card is exposed, the examinee is instructed to count and report the number of dots as fast as possible. Response times are compared to empirically validated times by brain injured subjects and normal adult participants.
SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES

(Lezak, 1983). If the patient is putting forth optimal effort, response times will gradually increase as the number of dots increase. As Lezak (1983) states, “more than one pronounced deviation from this pattern raises the likelihood that the patient is not acting in good faith” (p. 800). The previously mentioned procedure is very similar to the Grouped and Ungrouped Dots task.

**Grouped and Ungrouped Dots.** The Grouped and Ungrouped Dots test adds six more cards to the exam. However, this new set of cards mostly contains dots that are arranged in groupings that form shapes (e.g. squares, rectangles, and diamonds). One card in this new set consists of two squares made up of five dots and two ungrouped dots. These cards are labeled 7 to 12 and contain 12, 20, 16, 24, 28, and 8 dots, respectively. Again, the participant is instructed to count and report the number of dots after exposure to each card as quickly as possible. When compared to the Ungrouped Dots component, time used to count the dots on this task should be much less because the dots are grouped. Evaluation of this component consists of simply looking at the difference in times between the Grouped cards and the Ungrouped cards. The performance curve strategy becomes apparent when the patient takes longer to count the grouped dots, or there is little to no difference in time between the two trials (Lezak, 1983).

**Magnitude of Error**

Another malingering detection strategy is the magnitude of error approach. This is where the clinician looks at the quantitative or qualitative aspects of incorrect responses (Nitch et al., 2006). The underlying concept behind magnitude of error is that malingerers are detectable due to their approximate answers (Bash & Alpert, 1980). As Rogers et al. (1993) suggest, taking a closer look at incorrect responses may have been influenced by research on the Ganser syndrome.
(a dissociative disorder in which the individual gives grossly erroneous answers to questions, for instance, “How many legs does a dog have?” and the response would be “three”). Rogers et al. (1993) hypothesize that “the use of the magnitude of error strategy might yield different patterns of incorrect responses among simulators from approximate answers to those which are grossly wrong” (p. 261).

Martin, Franzen, and Orey (1998) applied the magnitude or error strategy by studying 30 college students given the option to simulate malingering or perform their best. Their sample also included 30 moderate to severe head injury patients not in litigation, and 7 suspected malingers. These researchers observed magnitude of error by using multiple choice forms of recognition memory tasks adapted for the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987) Logical and Visual Reproduction Memory subtests. Simulated and suspected malingers were more apt to choose lower probability multiple choice items. Interestingly enough, the simulated and suspected malingers also frequently chose answers that the head injury patients and controls did not (Martin, Franzen, & Orey, 1998).

**Atypical Presentation**

Numerous researchers and clinicians look for atypical performance patterns to detect symptom feigning. That is, they try to observe inconsistencies in test performance. For example, a client may do poorly on a test of word knowledge but do well on a test of similar abilities when tested later in the assessment session. Mittenberg, Azrin, Millsaps, and Heilbronner (1993) compared scores on the WMS-R of 39 outpatients with head-injuries and 39 individuals instructed to malinger head injury symptoms. In this study, the atypical performance strategy was observed when the General Memory Index was compared with the Attention/Concentration
Index. Mittenberg and colleagues (1993) found that the majority of simulated malingerers earned an average Attention/Concentration Index two standard deviations below the normal range, which was well below the mean score of the General Memory Index. In this case, the atypical performance strategy detected malingering because “it is not likely that an individual with severely impaired attention span could demonstrate low average memory abilities, because one cannot remember what is not first attended to. Attention capacity limits memory function so that severely impaired attention should result in memory impairment that is at least equally severe” (Mittenberg et al., 1993, p. 37). In other words, Attention Concentration scores less than those of General Memory are suggestive of suboptimal effort.

Doerr and Carlin (1991) illustrate the atypical performance detection strategy by describing a litigation case of a 35 year old male who suffered a mild traumatic brain injury (MTBI) after being in a motorcycle accident. The patient underwent a comprehensive neuropsychological evaluation one month from the time of injury and a similar evaluation two years post-injury. Results from the second evaluation showed marked deterioration in 21 of 33 neuropsychological tests. However, there was no evidence of additional neurological complaints or illness between evaluations. Considering improvement in neuropsychological functioning is the rule in MTBI, malingering was suggested by observing atypical performance.

Despite the aforementioned literature, the clinician must use caution when using atypical presentation as a malingering detection strategy. As Pankratz (1988) observed, discrepancy of presentation and symptoms common with functional disorders are not unusual in patients with brain injury. However, atypical presentation, if used wisely, is potentially an important tool for malingering detection.
Symptom Validity Testing

Clinical neuropsychologists are responsible for determining the validity of clinical information (symptoms, signs, etc) collected during the course of a neuropsychological evaluation (Bush et al., 2005). Symptom validity may be defined as “the accuracy or truthfulness of the examinee’s behavioral presentation (signs), self-reported symptoms (including their cause and course), or performance on neuropsychological measures” (Bush et al., 2005, p. 420). Specialized measures have been developed and utilize symptom validity as a detection strategy to detect false presentation of symptoms, insufficient effort on ability tests, and outright malingering. These measures, commonly called Symptom Validity Tests (SVTs), are currently a major focus of neuropsychological malingering research (Bianchini, Mathias, & Greve, 2001).

Symptom validity testing was first introduced by Pankratz (1983). His early studies were designed to detect exaggerated perceptual deficits (e.g. hearing loss) by presenting numerous trials in which individuals were asked to pick between two choices (e.g. the presence of sound, or its absence). The stimulus (e.g. a tone) would only be presented half of the time. The underlying principle of this technique is that it provides a 50% probability of being correct if the respondent is merely guessing (Pankratz, 1983). If the patient scores below-chance, this is a marker that he or she does indeed “recognize the correct stimulus and is responding with bias against this item” (Bianchini et al., 2001, p. 21). Pankratz (1983) named this method “symptom validity testing” and, in addition to using it to assess perceptual deficits, adapted it to inquiries of feigned memory complaints.

When assessing memory complaints, Pankratz (1983) presented patients with one red light and one white light. During each trial, one of the lights came on for two seconds. Then, the
patient completed 15 seconds of Symbol Digit Modalities Test (Smith, 1968) as a distractor task. After the distractor task, the patient was to simply report which of the lights came on. Thus, the patient was forced to choose either the red light or white light. The Pankratz (1983) procedure laid the ground work for forced-choice tests. The forced-choice paradigm is the common underlying principle by which most SVTs are designed. Simplicity is also an aspect of a useful or effective SVT. The test should be so simple that both normal participants and cognitively deficient individuals can obtain near perfect scores (Bianchini et al., 2001).

Symptom Validity Test research has come a long way since 1983, and is now at the forefront of the malingering literature. What is remarkable about SVTs is their effectiveness in terms of catching probable malingerers. Basically, SVTs are designed so that a well-motivated individual can easily achieve normal scores (Larrabee, 2003). The SVT derives its power from the fact that below chance performance means the individual deliberately avoided correct responses (Bianchini et al., 2001). This avoidance is precisely what defines the behavior of a malingerer.

However, many malingerers are sophisticated and do not produce below-chance scores. This is why SVTs are also standardized on a variety of clinical groups who have no motivation to feign cognitive impairment. Then, cut-off scores for groups of traumatic brain injury, demented patients, and other clinical disorders might prove useful when a below-chance outcome is not obtained. If an individual’s score is well below that of patients with obvious cognitive impairment, his or her effort may be considered suboptimal. The information a neuropsychologist can obtain from an SVT, combined with other neuropsychological measures and observational data, allows him or her to make an informed diagnosis of malingering. The
remainder of this literature review will describe frequently used and empirically supported SVTs such as the Digit Memory Test (DMT; Hiscock & Hiscock, 1989), Test of Memory Malingering (TOMM; Tombaugh, 1996), Computerized Assessment of Response Bias (CARB; Conder, Allen, & Cox, 1992), and the 21-Item Test (Iverson, Franzen, & McCracken, 1991).

**Digit Memory Test.** Hiscock and Hiscock (1989) developed the Digit Memory Test (DMT). Currently, a plethora of empirically validated SVTs are based on the concept used in the DMT. The task is comprised of 72 trials (three 24-item blocks). Following the presentation of a five-digit string of numbers, there is a two-choice recognition task (Bianchini et al., 2001). According to Bianchini et al. (2001) the “passive delay period between target presentation and recognition increases with blocks of 5, 10, and 15-s intervals” (p. 21). To evaluate the performance on the DMT, one must compare the number of correct responses to a cut-off score as well as chance and below chance levels of performance based on the binomial distribution.

The Hiscock and Hiscock DMT revolutionized symptom validity testing in that it made the examinee choose a five-digit number string, which increased the validity of the test as an actual measure of memory (Hiscock & Hiscock, 1989). Also, as Bianchini et al. (2001) point out, using increasing delay intervals between presentation and recognition develops the false impression that the task gets more difficult as time goes on. This illusion is beneficial because malingers often suppress their performance on difficult tasks and may be reluctant to do so on easier ones.

Prigatano and Amin (1993) provided further validation for the DMT by studying impaired individuals and suspected malingers. Their sample consisted of “27 patients with unequivocal cerebral dysfunction, 5 patients with postconcussional syndrome, 6 suspected
maligners and 10 normal controls” (Prigatano & Amin, 1993, p. 537). Patients with cerebral dysfunction and memory impairment scored between 95% and 100% correct, whereas the suspected malingering group obtained a score of only 74% correct. Prigatano and Amin (1993) noted that the suspected malingerers did not perform significantly below chance, still their DMT performance clearly was indicative of suboptimal effort.

**Computerized Assessment of Response Bias.** The Computerized Assessment of Response Bias (CARB) is an effort measure developed by Conder, Allen, and Cox in 1992 (as cited in Green & Iverson, 2001). The CARB uses a digit recognition format to detect poor effort during the course of a neuropsychological examination. During the administration of the CARB, the examinee is told he or she will see

“111 successive 2-s presentations of a five-digit number on the computer screen, followed by delays, during which the screen displays the message “Count backwards from 20”, before presentation of 2 five-digit numbers, one of which is the previously presented digit string” (Green & Iverson, 2001, p. 494).

Then, by pressing a right or left shift button, the examinee simply chooses which string he or she believes was presented initially. For each set of 37 items, the delay times are “3, 6, and 9 s” (Green & Iverson, 2001, p. 494). The CARB’s administration time is approximately 5 minutes (Gervais, Rohling, Green, & Ford, 2001).

The underlying principle of the CARB is its simplicity. In fact, most participants achieve perfect or near-perfect scores (Allen, Iverson, & Green, 2002). Internal consistency of the CARB for patients putting forth maximum effort is very high. However, the internal consistency varies for those who do not try their best on the measure (Allen et al., 2002). Allen et al. (2002) did not
report actual internal consistency figures. De Armas (1996) assessed the CARB’s test-retest reliability using a sample of male domestic violence offenders who were attending an intervention program. The time between tests was one week and participants were asked to do their best. The reliability coefficient of \( r = .97 \) was impressively high (Allen et al., 2002).

**Test of Memory Malingering.** The TOMM was developed because the author, a practicing neuropsychologist, wanted an SVT that was “sensitive to malingering but insensitive to neurological impairments” (Tombaugh, 1996, p. 261). Cognitive psychology has found that humans can store and retrieve a great deal of visual information. Tombaugh (1996) used this knowledge base and utilized exclusively visual stimuli when he created the TOMM. The visual stimuli on the TOMM are drawings of everyday items (e.g. a spoon, birds, a hat, etc.). According to Tombaugh (1996), the TOMM is “a 50-item recognition test for adults that includes two learning trials and a retention trial. During the two learning trials, the patient is shown 50 line-drawings (target pictures) of common objects for 3 seconds each, at 1-second intervals” (p. 1). First, the examinee is read instructions and then he or she begins trial one. After the presentation of 50 target pictures, the examinee is individually presented with 50 panels consisting of two choices each. One choice is the target item and one is a distractor item. Twenty minutes later, during Trial 2, the process is repeated. The original design of the TOMM presented individuals with four choices. However, the number of choices was later reduced to two to make the “actual level of chance responding consistent with the perceived level of chance responding” (Tombaugh, 1996, p. 11).

Tombaugh (1996) collected empirical data to validate the TOMM by administering the test to four samples. The first sample was a group of 405 community-dwelling healthy
individuals ranging in age from 16 to 84 years ($M = 54.8, SD = 20.2$). Recruitment for the sample took place in psychology classes, in public shopping malls, at social organizations, and at workplaces. One-hundred and ninety participants were male and 215 were female. This sample was administered the four-choice format of the TOMM. These individuals found the TOMM to be extremely easy, with 94% of targets correctly identified on Trial 1. On Trial 2 and the retention trial, over 99% of targets were correctly identified.

The second sample consisted of 70 community-dwelling volunteers in the age range of 17 to 73 years ($M = 37.8, SD = 14.2$). Forty-four of the participants were male and 26 were female. Each participant was administered the two-choice format of the TOMM. For Trial 1, the sample identified 95.6% of the targets correctly. For Trial 2 and the retention trial, the participants correctly identified over 99% of the targets.

The third sample was comprised of 138 individuals, the majority of whom were brain damaged. This group was divided into five subgroups based on the following diagnoses: aphasia, traumatic brain injury, general cognitive dysfunction, dementia, and no cognitive dysfunction. As Tombaugh (1996) states, “The robustness of the TOMM is particularly evident in Trial 2 – the Dementia group obtained 92% correct and the remaining four groups responded at greater than 97% accuracy” (p. 13). The other four groups performed highly accurately as well, by correctly identifying of 97% of the targets on Trial 2.

Finally, the fourth sample consisted of 27 college students instructed to malinger a memory problem and 22 college students in a control group who were asked to take the test in the standard manner and try their best. On Trial 2, the control group ($M = 49.9; SD = 0.2$)
performed significantly better than the simulated malingerers ($M = 35.3; SD = 9.4$) (Tombaugh, 1996). These results demonstrate that the TOMM can clearly detect feigned memory impairment.

Also, Rees, Tombaugh, Gansler, and Moczyngski (1998) conducted five validation experiments for the TOMM and found it to be a reliable method for detecting malingering. Rees et al. (1998) also claim the TOMM has high levels of specificity and sensitivity and should be used as a symptom validity test for malingering.

High specificity and sensitivity rates for the TOMM were also reported in a 2004 study by Powell, Gfeller, Hendricks, and Sharland. Powell et al. (2004) administered the TOMM to 80 university student volunteers who were randomly assigned to the following groups: a control group instructed to perform optimally, a symptom-coached group instructed to feign memory problems after being educated about traumatic brain injury symptomatology, and a test-coached group instructed to feign memory problems after being educated about test-taking strategies to avoid detection. Results showed a TOMM specificity of 100% and a sensitivity of 94.2% (Powell et al., 2004). Greve, Bianchini, and Doane (2006) also concluded that the TOMM had excellent classification accuracy. Greve et al. (2006) administered the TOMM to 161 traumatic brain injury patients, 41 of which were classified as Probable or Definite Malingerers according to the Malingered Neurocognitive Dysfunction criteria proposed by Slick, Sherman, and Iverson (1999). Greve et al. (2006) found a specificity of 95% when using a cut-off score of 46 on Trail 2 and the retention trial. This study also found a sensitivity of over 45% for the TOMM.

**The 21-Item Test.** The 21-Item Test is based on the symptom validity strategy made popular by Pankratz (1983). It is an adaptation of a word-list task popularized by Brandt and colleagues (Brandt, Rubinsky, & Lassen, 1985). The 21-Item Test was developed by Iverson,
Franzen, and McCraken (1991) and is used to detect biased responding. Biased responding may be defined as test performance that does not reveal the respondent’s true ability level (Iverson, 1998). The test is comprised of two lists of common nouns. There is the original word list (e.g. the target list) and a second word list including the target words and 21 foil words used in the recognition component. The original list is read to the participants immediately followed by a recall phase. Then, the examinees are presented with a “two-alternative, forced-choice recognition task in which they are required to select the word from the target list” (Iverson, 1998, p. 7). To clarify, the examinees are to select the word he or she believes was read aloud from the original list. The word lists contain 7 rhyming word pairs, 7 words that are semantically unrelated and 7 words that are semantically similar (Iverson, 1998). Items are presented at a rate of approximately one every 1.5 seconds. Overall, the word lists should take about 30-33 seconds to read.

Similar to most forced-choice tests, individuals’ scores on the 21-Item Test should be compared to the probable range of chance performance (Iverson, 1998). Guessing alone on this measure should result in scores between 7 to 14 correct on the recognition component, 90% of the time. If an examinee scores above or below this level, his or her responses are not likely to be a result of guessing. Hence, a score of 6 or less is indicative of suppressed performance. The literature on the 21-Item Test is meager, but there are some empirical studies that have provided cut-off scores from specific control and patient groups.

In an analog malingering study, Iverson et al. (1991) used the 21-Item Test to differentiate college students instructed to mangle from students performing their best and patients with documented memory impairment performing their best. They applied a cut-off
score that minimized false positives on the forced-choice task which resulted in “a correct classification rate of 100% for normal controls and memory-impaired subjects, 65% for experimental-malignerers, and 88% overall” (Iverson, 1998, p. 10).

Iverson, Franzen, and McCracken (1994) extended the previously mentioned research by administering the test to 60 patients undergoing neuropsychological evaluation, 60 psychiatric inpatients, and 60 community-dwelling volunteers. Fifty-percent of the psychiatric and community participants were instructed to malinger, and the remaining participants were told to perform their best. Discriminate function analysis correctly classified 90% of the participants into nonmalingering and malingering groups. In another study by Iverson and Franzen (1996), the 21-Item Test was given to twenty psychiatric patients and twenty undergraduate college students. In one phase, participants were instructed to put forth their best effort, and in the other phase they were instructed to malinger memory impairment. Iverson (1998) explains the results of this study by stating, “Applying a cutting score that would eliminate false positives on the forced-choice component of the test resulted in a correct classification rate of 100% for normal control, and memory impaired subjects, 22.5% for experimental malingerers, and 69% overall” (p. 10). Results from these three studies were also compared to chance. As stated, guessing alone would result in a score between 7 to 14, 90% of the time. Also previously mentioned, scoring 6 or below constitutes a below-chance performance which is a marker for malingering. However, no control or memory-impaired individual scored below chance. The 21-Item Test has also been empirically studied in clinical samples.

Research from 212 non-litigating patients seen for neuropsychological evaluations due to traumatic brain injury, dementia, cerebral vascular accidents, seizure disorder, and learning
disabilities was evaluated by Boone, Schauss, Franzen, Iverson, Epperly, and Duncan (1994). Five individuals scored below the forced-choice cut-off score indicating response bias (8 or less). Iverson (1998) emphasizes that the sample consisted of persons with possible dementia or other severe cognitive dysfunction. Despite these few below-chance performances, research has shown that a score below 9 is extremely rare. Arnett and Franzen (1997) administered the 21-Item Test to 149 inpatients from a substance abuse program. They then divided the participants into two groups based on performance on the Delayed Memory Index of the WMS-R. One group was the high quintile group, and one was the low quintile group. There was a remarkable difference between the Delayed Memory Index scores of the high quintile group ($M = 118.9; SD = 7.1$) and the low quintile group ($M = 72.8; SD = 6.6$). However, the 21-Item Test forced-choice scores of the high quintile group ($M = 18.3; SD = 1.6$) and the low quintile group ($M = 17.0; SD = 2.2$) were comparable. These findings suggest that even patients with marked memory-impairment can perform well on the 21-Item Test (Arnett & Franzen, 1997).

A review of literature on the 21-Item Test found that data from several different groups are available. These groups were comprised of patients, normal controls, and experimental-malingerers. Currently, there are no studies that use the 21-Item Test with elderly normals and demented elderly. In fact, Iverson (1998) states, “Elderly subjects with dementia were not included in these samples because they were deemed an inappropriate comparison group for detecting response bias. These elderly subjects are more likely to score below the cutoffs of 12 and 9 (although a score below 9 remains rare)” (p. 14).

The aforementioned research on the 21-Item Test provided mean ages of participants for each sample utilized. Of those mean ages, there was a range from 20 to 49. This set of mean ages
was bimodal, with the modes being 36 and 38. The median was 36. As stated, Iverson (1998) did not include data from elderly participants because he felt they were not an appropriate comparison group in the detection of biased responding. Iverson (1998) also concluded that elderly individuals are more likely to score in at least the suspicious range. Despite Iverson’s (1998) conclusions, the current study utilized two samples of elderly participants. The decision to use elderly samples was based on the simplicity of the 21-Item Test and the absence of 21-Item Test data on this population. This is the first investigation of elderly performance on the 21-Item Test.

The forgoing literature review leads to the following hypotheses: (a) Normal elderly participants in the community will perform in a non-biased manner on the 21-Item Test, (b) Elderly participants residing in a nursing home (with or without cognitive dysfunction) will perform in a non-biased manner on the 21-Item Test, and (c) Scores on the 21-Item Test for the total sample of participants residing in the community and in nursing homes will not be significantly associated with age, education, or scores on the Temporal Orientation Scale and Short Portable Mental Status Questionnaire.
Participants

**Total sample.** A total of 73 (51 women, 22 men) elderly individuals living in Mid-Missouri participated in the study. All participants were 64 years of age or older. Means for age and education were 81.00 years ($SD = 8.67; Range = 64 to 99$) and 12.20 years ($SD = 3.17; Range = 8 to 22$), respectively. Seventy-two participants were Euro Americans and one was African American. Fifty-three resided in local retirement/nursing homes and 20 were living independently in the community and were attending a local church program for elderly men and women. The former sample will be referred to as the “nursing home group” and the latter sample will be referred to as the “community group.”

**Nursing home group.** This sample consisted of 53 (37 women, 16 men) participants; 1 was African American and 52 were Euro American. Means for age and education were 82.79 years ($SD = 8.56, Range = 64 to 99$) and 11.41 years ($SD = 2.46, Range = 8 to 19$). These participants resided in nursing/retirement homes in mid Missouri. Table 3 provides the diagnoses for these participants. Table 4 presents a listing of medications prescribed for this group.

**Community group.** This group was comprised of 20 (14 women, 6 men) elderly volunteers living independently in the community who were members of a local church activity program. Means for age and education were 76.25 years ($SD = 7.19$) and 14.30 years ($SD = 3.89, Range 8 to 22$), respectively. All individuals were Euro Americans. Participants were generally healthy, although many reported medical conditions typically associated with the aging process.
Table 3

*Diagnostic Breakdown of Participants in the Nursing Home Group (n = 53)*

<table>
<thead>
<tr>
<th>Clinical Diagnosis</th>
<th>n</th>
<th>Clinical Diagnosis</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective Psychosis</td>
<td>1</td>
<td>Late Effects CVA</td>
<td>3</td>
</tr>
<tr>
<td>Alcoholic Psychosis</td>
<td>1</td>
<td>Lung Cancer</td>
<td>1</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>9</td>
<td>Malaise</td>
<td>8</td>
</tr>
<tr>
<td>Angina</td>
<td>2</td>
<td>Non-Organic Psychosis</td>
<td>5</td>
</tr>
<tr>
<td>Anxiety Disorder</td>
<td>5</td>
<td>Neurotic Disorder</td>
<td>13</td>
</tr>
<tr>
<td>BiPolar I</td>
<td>3</td>
<td>Organic Brain Syndrome</td>
<td>2</td>
</tr>
<tr>
<td>Cerebral Artery Occlusion</td>
<td>2</td>
<td>Organic Psychosis</td>
<td>1</td>
</tr>
<tr>
<td>Cerebral Vascular Accident</td>
<td>8</td>
<td>Parkinson’s Disease</td>
<td>1</td>
</tr>
<tr>
<td>Chronic Bronchitis</td>
<td>1</td>
<td>Personality Disorder</td>
<td>2</td>
</tr>
<tr>
<td>COPD</td>
<td>6</td>
<td>Pneumonia</td>
<td>1</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>10</td>
<td>Pre-Senile Dementia</td>
<td>3</td>
</tr>
<tr>
<td>Convulsions</td>
<td>1</td>
<td>PVDS</td>
<td>1</td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>1</td>
<td>Renal Failure</td>
<td>1</td>
</tr>
<tr>
<td>Dehydration</td>
<td>1</td>
<td>Schizoafffective Disorder</td>
<td>1</td>
</tr>
<tr>
<td>Dementia</td>
<td>5</td>
<td>Senile Dementia</td>
<td>1</td>
</tr>
<tr>
<td>Depression</td>
<td>2</td>
<td>Sleep Disturbance</td>
<td>4</td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>26</td>
<td>Symbolic Dysfunction</td>
<td>7</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>1</td>
<td>Syncope</td>
<td>1</td>
</tr>
<tr>
<td>Hypertension</td>
<td>23</td>
<td>Transient Ischemic Attack</td>
<td>2</td>
</tr>
</tbody>
</table>

*Note.* The total number of diagnoses exceeds 53 because many participants had multiple diagnoses. COPD = Chronic Obstructive Pulmonary Disease, CVA = Cerebrovascular Accident, and PVDS = Pulmonary Vascular Disease.
Table 4

Medications Prescribed in the Nursing Home Group

<table>
<thead>
<tr>
<th>Medication</th>
<th>n</th>
<th>Medication</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldomet</td>
<td>1</td>
<td>Paxil</td>
<td>3</td>
</tr>
<tr>
<td>Atenolol</td>
<td>1</td>
<td>Plavix</td>
<td>1</td>
</tr>
<tr>
<td>Aricept</td>
<td>8</td>
<td>Prozac</td>
<td>2</td>
</tr>
<tr>
<td>Ativan</td>
<td>1</td>
<td>Remeron</td>
<td>1</td>
</tr>
<tr>
<td>Benzapril</td>
<td>1</td>
<td>Seroquel</td>
<td>6</td>
</tr>
<tr>
<td>Buspar</td>
<td>2</td>
<td>Synthroid</td>
<td>2</td>
</tr>
<tr>
<td>Celexa</td>
<td>4</td>
<td>Tramadol</td>
<td>1</td>
</tr>
<tr>
<td>Cymbalta</td>
<td>1</td>
<td>Trazadone</td>
<td>1</td>
</tr>
<tr>
<td>Depakote</td>
<td>1</td>
<td>Ultram</td>
<td>2</td>
</tr>
<tr>
<td>Effexor</td>
<td>2</td>
<td>Wellbutrin</td>
<td>1</td>
</tr>
<tr>
<td>Exelon</td>
<td>1</td>
<td>Zoloft</td>
<td>5</td>
</tr>
<tr>
<td>Lexapro</td>
<td>6</td>
<td>Zyprexa</td>
<td>2</td>
</tr>
<tr>
<td>Lidoderm</td>
<td>1</td>
<td>Xanax</td>
<td>1</td>
</tr>
<tr>
<td>Lithium</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopressor</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mellaril</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Namenda</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurontin</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitroglycerin</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Total medications exceeds 53 because many participants were taking multiple medications.
None were diagnosed with a significant medical condition (e.g., cancer of the lung), dementia (e.g., Alzheimer’s disease, or Parkinson’s disease), or major psychiatric illness (e.g., major depression, bipolar disorder, etc.)

**Materials**

*Patients’ medical records.* For those residing in the nursing/retirement homes, administrators examined the participants’ medical charts. The underlying purpose was two-fold: to check participation-eligibility and to obtain diagnoses and current medications. The administrators made sure each participant was at least 64 years of age and not adjudicated incompetent. The administrators then recorded participant diagnoses and current medication(s) on the demographic form (see Appendix).

*Short Portable Mental Status Questionnaire.* The Short Portable Mental Status Questionnaire (SPMSQ; Pfeiffer, 1975) was used to screen participants for intellectual deterioration and dementia. This measure is composed of 10 questions that address memory, orientation, knowledge of current events, and the performance of serial mathematical operations. Gross cognitive status is reflected by the total number of errors: 0 to -2 = intact, -3 to -4 = mildly impaired, -5 to -7 = moderately impaired, and -8 to -10 = severely impaired. Norms of the SPMSQ are based on 997 community-dwelling elderly, 141 elderly individuals referred for psychiatric evaluation, and 102 elderly individuals residing in nursing homes.

Validation studies for the SPMSQ in elderly samples have yielded encouraging results. For example, one study found the SPMSQ to have good diagnostic precision for the identification of organic dysfunction and fairly good concurrent validity with popular neuropsychological measures such as the Bender Gestalt Test and Digit Span subtest of the
Wechsler Scales (Wolber, Romaniuk, Eastman, & Robinson, 1984). Also, Erkinjuntti, Sulkava, Wilkstrom, and Autio (1987) evaluated the specificity and sensitivity of the SPMSQ in an elderly sample of community residents (n = 119) and medical inpatients (n = 282). In the community group the sensitivity to cognitive impairment of the SPMSQ was 66.7% with a specificity of 100%. For the medical inpatients, the SPMSQ yielded a sensitivity of 86.2% and specificity of 99.0%. Also, the SPMSQ has been found to be comparable to the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and both measures correlate with Computed Tomography (CT) and Electroencephalogram (EEG) results (Salokangas, Loikkanen, & Santala, 1990). Finally, Pfeiffer (1975) conducted a 4-week test-retest stability study of the SPMSQ which yielded reliability coefficients of .82 and .83 in two elderly samples of 59 total participants.

During the administration of the SPMSQ, questions were read out loud to the examinee who responded orally. The SPMSQ provides space next to each question for the examiner to indicate if a response is correct and to record an examinee’s verbatim response. The SPMSQ also has a scoring guide with specific instructions to correct for education level. The instrument is presented in Table 5.

**Temporal Orientation Scale.** This measure (TOS; Benton, Sivan, Hamsher, Varney, & Spreen, 1994) was administered to each participant. It consists of questions regarding the year, month, day of the month, day of the week, and time of day. Incorrect responses are assigned penalty points with possible scores ranging from -113 (worst score) to zero (perfect score).
Table 5

*The Short Portable Mental Status Questionnaire (SPMSQ; Pfeiffer 1975)*

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is the date today?</td>
</tr>
<tr>
<td>2. What day of the week is it?</td>
</tr>
<tr>
<td>3. What is the name of this place?</td>
</tr>
<tr>
<td>4. What is your telephone number or street address?</td>
</tr>
<tr>
<td>5. How old are you?</td>
</tr>
<tr>
<td>6. When were you born?</td>
</tr>
<tr>
<td>7. Who is the President of the United States of America now?</td>
</tr>
<tr>
<td>8. Who was President just before him?</td>
</tr>
<tr>
<td>9. What was your mother’s maiden name?</td>
</tr>
<tr>
<td>10. Subtract 3 from 30 and keep subtracting 3 from each new number, all the way down.</td>
</tr>
</tbody>
</table>

*Note.* The following are the classifications of cognitive functioning based on age: 0 – -2 = intact, -3 to -4 = mildly impaired, -5 to -7 = moderately impaired, and -8 to -10 = severely impaired.
Norms are based on 434 healthy persons and yield the following classifications in terms of adequacy of temporal orientation: 0 to -2 = normal, -3 = borderline, -4 to -7 = moderately defective, and -8 or more = severely defective. A scoring sample is presented in Table 6.

**21-Item Test.** This instrument is a symptom validity measure that was administered to each participant. Its purpose is to detect suboptimal effort or deliberate failure. The task uses two word lists comprised of nouns. The initial word list is read to the examinee and is referred to as the target list. The other list includes 21 word pairs each consisting of a target word from the original list and a distractor word that is not on the original list. The original list is read at a rate of 1.5 s per word. Following the reading of the first list, the participant is asked to freely recall as many items as he or she possibly can. Finally, the participants are given a forced-choice recognition component in which they are read word pairs and asked to choose which word from each pair they previously heard.

Scoring is achieved by counting all correct items on the forced-choice component of the test. There are two ways to interpret the scores. The first is using statistical criteria. The statistical criteria consist of the following cut-offs: \( \geq 15 = \) above chance, \( 7 – 14 = \) chance, and \( \leq 6 = \) below chance. The second way to interpret scores is to use Iverson’s (1998) clinical cut-offs which are “based on known distributions of performance from specific control and patient groups” (p. 19). The clinical cut-off scores are as follows: 12 = suspicious, 9 – 11 = highly suspicious, and \( \leq 8 = \) biased (Iverson, 1998).
### Table 6

**Scoring Sample for the Temporal Orientation Scale**

<table>
<thead>
<tr>
<th>Response (Errors)</th>
<th>Correct Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Jan</td>
<td>Feb</td>
</tr>
<tr>
<td>Day</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>Year</td>
<td>2008</td>
<td>2009</td>
</tr>
<tr>
<td>Week Day</td>
<td>Wed</td>
<td>Wed</td>
</tr>
<tr>
<td>Time</td>
<td>2:30</td>
<td>2:15</td>
</tr>
</tbody>
</table>

*Note.* The total score for this participant is 17. Scoring for the TOS is as follows:

**Error Points**

- **Month:** 5 for each month up to a maximum of 30. Full credit if within 15 days of correct date.
- **Day of month:** 1 for each day up to a maximum of 15
- **Year:** 10 for each year up to a maximum of 60. Full credit if within 15 days of correct date.
- **Day of week:** 1 for each day up to a maximum of 3.
- **Time of day:** 1 for each 30 min. from correct time up to a maximum of 5.
Design and Procedure

Clinical setting. Prior to starting the investigation, a graduate student examiner received specialized training on test administration and interpretation. This individual had successfully completed formal graduate courses in assessment, including a class dealing with the assessment of intelligence in adults. All measures used were administered according to standard instructions. Also, permission to enter two nursing/retirement home facilities was granted from the administrators. The administrators made an announcement to all appropriate residents of the investigator’s intentions, and participants were solicited. Finally, informed consent (see Appendix A) was obtained from each volunteer (or his or her guardian) prior to participation and confidentiality remained a priority throughout the study.

To acquire diagnoses and medication information from each nursing home participant, the nursing home administrators reviewed the volunteers’ medical charts. The administrators recorded patient medical information on the demographic form (see Appendix B). Next, each patient was scheduled for an assessment appointment. At the time of assessment, participants were seen individually in a preselected testing room provided by the nursing home. The same rooms were utilized in each of the two nursing homes for all participants tested. The rooms were well lit and comfortable in temperature. No breaks were offered because the time of testing did not exceed ten minutes. Rapport was established and administration of the assessment measures began.

Examination procedure. Participants were brought to their appointments by nursing home staff. Once the participant was comfortably seated in the examination room and rapport had been established, he or she was asked to provide demographic information such as age, date
of birth, and education level. Next, the SPMSQ was administered, followed by the TOS, and 21-Item Test.

In one case, a participant was too cognitively impaired to complete the SPMSQ. The individual was thanked for her participation and escorted back to her room by nursing home staff. Her test responses were not included in the database. It was determined that this resident was in the later stages of Alzheimer’s disease and was having one of her “bad days.” The participant was virtually non-responsive on the day of testing.

Community setting. The church activity program had a meeting of community-dwelling elderly every month. The investigator was given the opportunity to address the group and provide a synopsis of the study and ask for volunteers. The elderly church members were then given the opportunity to volunteer and make an appointment to participate. This process would later prove to be problematic and will be addressed further in the discussion chapter.

All testing took place in a private room at the church. The room was a favorable testing environment in which rapport was easy to establish and maintain. First, participants completed the informed consent (see Appendix A) followed by the demographic form (see Appendix B). Participants were asked to provide any general medical conditions they were aware of as well as current medication regimens. Next, the SPMSQ, TOS, and 21-Item Test were administered. At the conclusion of testing, the participant was thanked and allowed to leave.
Means and standard deviations for age, education, TOS, SPMSQ, and 21-Item Test forced-choice component for the total sample are presented in Table 7 and separately for the community and nursing home samples in Tables 8 and 9, respectively.

For each participant in the community sample, temporal orientation was intact. Also, every participant had normal cognitive functioning (-1 to 0 on the SPMSQ). On the 21-Item Test forced-choice component, all 20 participants scored in the range consistent with Optimal Effort ($\geq 15$ correct). These findings support hypothesis 1 and indicate that the present sample of community-dwelling elderly performed in a non-biased manner on the 21-Item Test.

For participants in the nursing home sample, the SMPSQ indicated that cognitive functions were grossly intact in 23 (43.4%) individuals, mildly impaired in 15 (28.3%), and moderately-to-severely impaired in 15 (28.3%). On the TOS, 26 (49.1%) individuals were in the normal range, two (3.8%) were borderline, two (3.8%) were moderately impaired, and 22 (41.5%) were severely impaired. Using the statistical criteria (i.e. probable range of chance performance) for interpreting the 21-Item Test forced-choice component, 48 (90.6%) of the participants had scores suggesting optimal effort ($\geq 15$), five (9.4%) had random scores (7-14), and no one (0%) had a score below chance ($\leq 6$). When applying the clinical interpretative rules of Iverson (1998) to detect biased responding, only 1 (2%) individual had a score classified as suspicious ($\leq 12$), two (4%) had scores that were highly suspicious (9 to 11), and no one (0%) had a biased score ($\leq 8$). These findings provide substantial support for hypothesis 2 and indicate that the vast majority of participants residing in nursing homes performed in a non-biased manner on the 21-Item Test. This finding also suggests that even elderly persons with cognitive
Table 7


<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>81.00</td>
<td>8.67</td>
<td>64 – 99</td>
</tr>
<tr>
<td>Education</td>
<td>12.20</td>
<td>3.17</td>
<td>8 – 22</td>
</tr>
<tr>
<td>SPMSQ (Errors)</td>
<td>-2.38</td>
<td>2.47</td>
<td>-9 – 0</td>
</tr>
<tr>
<td>TOS</td>
<td>-16.49</td>
<td>28.80</td>
<td>-97 – 00</td>
</tr>
<tr>
<td>21-Item Forced Choice</td>
<td>18.11</td>
<td>2.45</td>
<td>9 – 21</td>
</tr>
</tbody>
</table>

*Note. SPMSQ = Short Portable Mental Status Questionnaire; TOS = Temporal Orientation Scale.*
Table 8

*Means, Standard Deviations, and Ranges for Demographic Variables, SPMSQ Error Scores, Temporal Orientation Scores, and 21-Item Forced-Choice Scores of the Community Sample (n=20).*

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>76.25</td>
<td>7.20</td>
<td>66 – 90</td>
</tr>
<tr>
<td>Education</td>
<td>14.30</td>
<td>3.89</td>
<td>8 – 22</td>
</tr>
<tr>
<td>SPMSQ (Errors)</td>
<td>-0.30</td>
<td>0.47</td>
<td>-1 – 0</td>
</tr>
<tr>
<td>TOS</td>
<td>0.00</td>
<td>0.00</td>
<td>00 – 00</td>
</tr>
<tr>
<td>21-Item Forced Choice</td>
<td>19.65</td>
<td>1.34</td>
<td>17 – 21</td>
</tr>
</tbody>
</table>

*Note. SPMSQ = Short Portable Mental Status Questionnaire; TOS = Temporal Orientation Scale.*
Table 9


<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>82.79</td>
<td>8.56</td>
<td>64 – 99</td>
</tr>
<tr>
<td>Education</td>
<td>11.41</td>
<td>2.46</td>
<td>8 – 19</td>
</tr>
<tr>
<td>SPMSQ (Errors)</td>
<td>3.17</td>
<td>2.46</td>
<td>0 – 9</td>
</tr>
<tr>
<td>TOS</td>
<td>-22.83</td>
<td>31.74</td>
<td>-97 – 00</td>
</tr>
<tr>
<td>21-Item Forced Choice</td>
<td>17.53</td>
<td>2.57</td>
<td>9 – 21</td>
</tr>
</tbody>
</table>

Note. SPMSQ = Short Portable Mental Status Questionnaire; TOS = Temporal Orientation Scale; 52 participants completed the TOS.
impairment should be expected to perform at chance or above levels on the 21-Item Test. Table 10 summarizes performance for the community and nursing home samples using the statistically based criteria as well as the clinical interpretive approach recommended by Iverson (1998).

For the total sample, Pearson Product-Moment Correlation coefficients were calculated between the 21-Item Test forced-choice score and the variables of age, education, TOS, and SPMSQ error score. These correlations appear in Table 11. As can be seen, there were significant associations between the forced-choice component of the 21-Item Test and the variables of age, TOS, and SPMSQ error scores. These results do not support hypothesis 3.

To investigate the meaning of these significant correlations, a stepwise multiple linear regression analysis was run in which age, TOS, and SPMSQ error scores were entered as independent variables and the forced-choice component of the 21-Item Test was entered as the dependent variable. The SPMSQ error score was the only variable that emerged as a significant predictor of the forced-choice scores, $t(3,68) = 4.43, p < .0001$.

Next, a bivariate linear regression analysis was run with SPMSQ error scores as the independent variable and Forced-choice scores as the dependent variable. This analysis was conducted to evaluate the prediction of forced-choice scores based on the degree of cognitive impairment reflected by SPMSQ error scores. The regression equation for predicting performance on the forced-choice component of the 21-Item test was:

$\text{Predicted forced-choice score} = .499(\text{SPMSQ error score}) + 19.298$.

Elderly individuals who are cognitively impaired tend to have lower forced-choice scores than elderly individuals who are not cognitively impaired. Accuracy in predicting the forced-choice scores was moderate. The correlation between the SPMSQ error score and the forced-choice
Table 10

Performance Based on Statistical and Clinical Standards of Interpretation of the Community and Nursing Home Samples.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Interpretation</th>
<th>Community (N=20)</th>
<th>Nursing Home (N=53)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>≥ 15 Above Chance</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>7-14 Chance</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>≤ 6 Below Chance</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Clinical</td>
<td>12 Suspicious</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9-11 Highly Suspicious</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>≤ 8 Biased</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note.* The total number of nursing home participants is 53. However, there are 56 nursing home participants in this table because multiple individuals’ scores fell in both Statistical and Clinical interpretation classifications.
**Table 11**

*Correlations for the Total Sample on Demographic and Test Variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age</th>
<th>Ed</th>
<th>TOS</th>
<th>21-FC</th>
<th>SPMSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>−</td>
<td>-.28*</td>
<td>-.27*</td>
<td>-.28*</td>
<td>-.28*</td>
</tr>
<tr>
<td>Ed</td>
<td>-.26*</td>
<td>−</td>
<td>.08</td>
<td>.20</td>
<td>-.04</td>
</tr>
<tr>
<td>TOS</td>
<td>.27*</td>
<td>.08</td>
<td>−</td>
<td>.24*</td>
<td>.67**</td>
</tr>
<tr>
<td>21-FC</td>
<td>-.28*</td>
<td>.20</td>
<td>.24*</td>
<td>−</td>
<td>.50**</td>
</tr>
<tr>
<td>SPMSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Errors</td>
<td>.28</td>
<td>.04</td>
<td>.67**</td>
<td>.50**</td>
<td>−</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

*Note.* Ed = Education; TOS = Temporal Orientation Scale; 21-FC = 21-Item Test Forced-Choice Component; SPMSQ = Short Portable Mental Status Questionnaire.
score was .502. Approximately 25% of the variance in the forced-choice component was accounted for by its linear relationship with performance on the SPMSQ. Table 12 provides predicted forced-choice scores based on the SPMSQ error scores.

Using the nursing home sample, a one-way between-groups analysis of variance (ANOVA) compared the mean scores on the 21-Item Test forced-choice component for individuals whose levels of intellectual functioning were classified as intact, mildly impaired, or moderately to severely impaired according to the SPMSQ error score. This test was statistically significant, $F(2, 50) = 5.90, p < .005, \eta^2 = .19$. Post hoc Bonferroni tests were conducted to evaluate differences among the means. Presented in Table 13 are the means, standard deviations, and ranges of the 21-Item Test forced-choice scores based on levels of intellectual impairment derived from SPMSQ error scores. A significant difference in the means existed between the intact group and both the mildly impaired and moderately to severely impaired groups, with individuals in the intact intellectual functioning group scoring above the other groups. Means for the mildly impaired and moderately to severely impaired groups did not differ significantly.
Table 12

*Predicted Forced-Choice Scores from SPMSQ Errors Scores*

<table>
<thead>
<tr>
<th>SPMSQ Error Score</th>
<th>Predicted Forced-Choice Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>-1</td>
<td>19</td>
</tr>
<tr>
<td>-2</td>
<td>18</td>
</tr>
<tr>
<td>-3</td>
<td>18</td>
</tr>
<tr>
<td>-4</td>
<td>17</td>
</tr>
<tr>
<td>-5</td>
<td>17</td>
</tr>
<tr>
<td>-6</td>
<td>16</td>
</tr>
<tr>
<td>-7</td>
<td>16</td>
</tr>
<tr>
<td>-8</td>
<td>15</td>
</tr>
<tr>
<td>-9</td>
<td>15</td>
</tr>
<tr>
<td>-10</td>
<td>14</td>
</tr>
</tbody>
</table>

*Note.* SPMSQ = Short Portable Mental Status Questionnaire
Table 13

21-Item Test Forced-Choice Scores for the Nursing Home Sample (n = 53) by SPMSQ

Classification of Intellectual Impairment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Impairment Level</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forced-Choice</td>
<td>Intact (0 – 2)</td>
<td>18.78</td>
<td>1.78</td>
<td>13 – 21</td>
</tr>
<tr>
<td></td>
<td>Mild (3 – 4)</td>
<td>16.80</td>
<td>3.43</td>
<td>9 – 21</td>
</tr>
<tr>
<td></td>
<td>Moderate to Severe (&gt; 5)</td>
<td>16.33</td>
<td>1.76</td>
<td>12 – 19</td>
</tr>
</tbody>
</table>

Note. SPMSQ = Short Portable Mental Status Questionnaire
CHAPTER 5
DISCUSSION

The purpose of the present study was to investigate the performance of two elderly samples on the 21-Item Test. The samples included individuals residing in nursing homes and individuals living independently in the community. This is the first study to investigate the performance of elderly individuals on the 21-Item Test. The specific focus was on the specificity of the forced-choice component of the 21-Item Test.

**Hypothesis One**

The first hypothesis stated that normal elderly participants living in the community would perform in a non-biased manner on the forced-choice component of the 21-Item Test. This hypothesis was confirmed. The normal elderly participants performed at a high level with scores ranging from 17 to 21. Thus, the specificity of the 21-Item Test forced-choice component was 100%. This figure is comparable to scores achieved by groups of non-malingering college students, psychiatric patients, and adult community volunteers whose mean scores ranged from 17.1 to 18.8 (Iverson, 1998). These groups were also considerably younger than the present sample with ages ranging from 20 years to 39 years (Iverson, 1998).

When interpreting these findings, it is important to note that the current sample of elderly volunteers might be considered atypical. These individuals enjoyed optimal health since none had chronic illnesses and none were taking medications for serious neurological or major physical impairments. All medical problems that existed were to be expected with the normal aging process. It is also noted that these participants were highly educated. The mean for educational level was 14.30 years ($SD = 3.89$, *Range* 8 to 22). Research has shown that high levels of education in the elderly are an unusual occurrence. For example, the 1995
standardization sample of the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997) showed that less than 30% of individuals age 65 and older had more than 13 years of education (Kaufman & Lichtenberger, 1999). However, 45% of the community sample in the present study had more than 13 years of education.

Individuals in this sample were in the middle to upper class in terms of socioeconomic status. Research by Antonovsky has shown that individuals with higher socioeconomic status are generally in better health than those with lower socioeconomic status (as cited in Adler et al., 1994). Another important factor is that all of these community-dwelling elderly were active in a local church social program. Each participant met at this social program with other elderly individuals twice a month. They also actively took part in regular worship services at the church. Thus, these elderly were fully engaged with the community and enjoyed an extensive social network.

These individuals were eager volunteers. According to Horn and Donaldson (1976), individuals who do not volunteer initially for research may achieve lower scores on ability tests than those who do volunteer (as cited in Lichtenberger, Kaufman, & Lai, 2002). Also, Rosenthal and Rosnow (1975) found that individuals who volunteer for research are both more intelligent and more sociable than those who do not volunteer (as cited in Kreiner, Alvarado, & Shoekley, 1997). Thus, one may assume that these community-dwelling individuals eagerly volunteered because they were confident in their abilities. Two final observations have to deal with the fact that all participants in this sample were Caucasian and lived in a rural community. The results may have differed had the sample been composed of inner city residents, including members of minority groups and those with lower socioeconomic status. Finally, the 21-Item Test was
designed for litigation assessment to detect poor test taking motivation and the possibility that an examinee was attempting to fake or malingering a memory disorder. Because the present sample of elderly were not in litigation and were not pursuing a monetary award for acquired disability following an accident or head injury, there was absolutely no reason for them to deliberately perform poorly. Finally, because the participants welcomed the interaction with the examiner and viewed the testing session as a “social encounter,” their motivation to perform their best was likely assured.

Hypothesis Two

The second hypothesis stated that elderly participants residing in a nursing home (with or without cognitive dysfunction) would perform in a non-biased manner on the forced-choice component of the 21-Item Test. This hypothesis was confirmed. No participant scored in the range classified as biased (≤ 8) or were deliberately choosing incorrect answers (≤ 6) (Iverson, 1998). However, one individual scored in the suspicious range (≤ 12) and two scored in the highly suspicious range (9 to 11). One individual had a clinical diagnosis of dementia and achieved a score on the Short Portable Mental Status Questionnaire (SPMSQ, Pfeiffer, 1975) indicative of mild cognitive impairment. This same individual earned a severely defective score on the TOS (Benton et al., 1994). Proper orientation to time is an essential adaptive capacity, without which an individual is incapable of self-determination or securing the satisfaction of his or her basic needs. Another individual showed moderate cognitive impairment on the SPMSQ and a third was moderately impaired on the SPMSQ. The latter individual also exhibited moderate temporal disorientation. Thus, the suspicious scores of these individuals were not likely the result of bias or frank malingering, but instead most probably reflect obvious cognitive and
adaptive impairments. Although no scores were in the biased range, suspicious scores are to be expected in a sample of individuals with severe cognitive dysfunction. Iverson (1998) found that non-litigating patients who scored in or below the suspicious range most likely had severe cognitive impairment. Also, Gontkovsky and Souheaver (2000) found that individuals with brain-damage living in a Veterans Affairs Medical Center performed more poorly on the 21-Item Test than medical controls at the same institution. Moreover, none of their brain-damaged patients scored in the range suggesting biased responding (Gontkovsky & Souheaver, 2000).

Not only were there no biased scores, the nursing home sample’s average forced-choice component score was comparable to that of other samples reported in the literature. The forced-choice component mean in the present sample of nursing home residents was 17.53 ($SD = 2.57$, $Range = 9-21$) and Iverson’s (1991) sample of 20 memory-impaired individuals had a mean forced-choice score of 16.10 ($SD = 2.73$, $Range = 9-20$) and a mean age of 49.2 years ($SD = 18.12$). Iverson’s (1991) study also included 20 normal young adult controls whose mean forced-choice score was 18.25 ($SD = 1.55$). The current sample’s mean was higher than Gontkovsky and Souheaver’s (2000) sample of medical controls from a Veterans Affairs Medical Center whose mean was 16.4 correct ($SD = 2.1$). Overall, the nursing home sample performed well on the forced-choice component of the 21-Item Test with scores comparable to other samples of non-malingering memory-impaired and normal participants who were considerably younger.

This relatively good performance in the current nursing home sample may be attributed to several factors. First, the forced-choice format was designed to be simple, undemanding, and even transparent as to its purpose. In fact, Franzen and Martin (1996) designed a study that utilized samples of graduate level psychology students and PhD level psychologists. Participants
in these samples were instructed to malinger a memory deficit on a battery of tests which included the 21-Item Test. This research showed that malingerers with psychological knowledge can figure out that the 21-Item Test is too easy to fail. Once an individual has ascertained this fact, he or she can perform in a manner that accurately portrays memory impairments without revealing an intention to appear impaired (Franzen & Martin, 1996). Although these individuals were feigning impaired memory, the level of impairment on the 21-Item Test forced-choice component was not severe enough to be considered indication of exaggeration or malingering (Franzen & Martin, 1996). It is important to note that, based on differences in age and education, the current elderly participants’ level of cognitive functioning was certainly not as high as the participants tested by Franzen and Martin (1996). Second, no participants in the current sample were in litigation or in conflict with nursing home administration and thus had no incentive to perform poorly. Third, the mental ability of nursing home residents is often under constant scrutiny which may have motivated these individuals to perform their best. Also, it is important to note that these individuals seemed to enjoy the interaction with the examiner and their optimal performance may have been a reflection of this positive rapport.

Specificity is defined as “the proportion of persons that do not have the condition who are predicted by the test to not have it” (Smith et al., 2003, p. 280). The current findings show excellent specificity (100 %) for the 21-Item Test in the elderly population when using the clinical and statistical interpretations. Thus, no participants were classified as responding in a biased manner (≤ 8) or as demonstrating frank malingering (≤ 6). This perfect specificity indicates that if an examinee earns a forced-choice score less than or equal to six, he or she is very likely to display gross cognitive impairment which can be observed clinically or the score
reflects suboptimal effort, uncooperativeness, or a deliberate attempt at feigning. Even though three individuals had suspicious scores, 94% of participants scored above the suspicious range \( \leq 12 \) proposed by Iverson (1998).

Sensitivity is defined as “the proportion of persons who have the condition that is predicted by the test to have it” (Smith et al., 2003, p. 280). One drawback of the current study is that it did not address issues of sensitivity. However, research suggests that the sensitivity of the 21-Item Test is low (Vickery, Berry, Inman, Harris, & Orey, 2001). For example, Frederick, Sarfaty, Johnston, and Powel (1994) administered the 21-Item Test to a sample of 269 college students with a mean age of 22.1 years \( (SD = 6.9) \). This sample consisted of 92 individuals instructed to give their best effort and 177 individuals instructed to fake cognitive impairment. Of the 177 individuals instructed to feign cognitive impairment, 94 were described as naïve malingerers (i.e., participants not given strategies for faking believable deficits) and 83 were described as informed malingerers (i.e., participants who were given strategies for faking believable deficits). Approximately 39% of the naïve malingerers and 23% of the informed malingerers were identified as responding in a biased manner on the 21-Item Test forced-choice component. Thus, Frederick et al. (1994) described the sensitivity of the 21-Item Test as limited.

Also, Inman et al. (1998) found a 21-Item Test sensitivity of only 5%. This sample consisted of 32 patients with intracranial disorder, 27 head-injury patients, and 151 college students randomly assigned into a control group or a group of naïve or coached malingerers. Research on 21-Item Test sensitivity has only been conducted using individuals less than 50 years of age (Vickery et al., 2001). The sensitivity of the 21-Item Test in the elderly population is unknown. Future research should utilize elderly malingerers or elderly individuals instructed to respond in a biased
manner. For example, elderly participants may be instructed to feign a memory deficit based on a case vignette provided by the researchers. This case vignette may tell the story of an individual in litigation after suffering a mild traumatic brain injury due to an injurious circumstance (e.g., car accident, an employment-related head trauma, etc.). The elderly participant could be asked to feign symptoms of cognitive dysfunction he or she thought would be relevant to the injury described in the case vignette. It may also be possible to use actual malingerers identified by means other than scores on the 21-Item Test. For example, if an individual was given a validated SVT such as the TOMM or Portland Digit Recognition Test (PDRT; Binder & Willis, 1991) and scored in the range suggestive of suboptimal effort, he or she could also be administered the 21-Item Test during the same session.

Hypothesis Three

The third hypothesis stated that scores on the forced-choice component of the 21-Item Test for participants residing in the nursing home or in the community will not be significantly associated with age, education, or scores on the TOS and SPMSQ. This hypothesis was not confirmed.

Correlational analysis for the current total sample yielded interesting findings. For example, education was not significantly correlated with the forced-choice component of the 21-Item Test. This lack of association was not due to restriction of range in education since the total sample had 8 to 22 years of schooling. This finding shows that educational level should not be a major concern when administering the 21-Item Test. In fact, research has shown third and fourth grade participants are capable of completing the test without suspicion of biased responding (Collier, Glass, Ryan, & Duckworth, 2008). Thus, one may assume that basic literacy is the only
educational attainment necessary for successful completion of the 21-Item Test.

The current study showed a small but significant negative correlation between forced-choice scores and age. Thus, for the current sample, as age increased forced-choice scores decreased. Even with the relationship between age and forced-choice scores no one scored below chance and 94% scored above Iverson’s (1998) suspicious range (≤12). Other research has shown a significant correlation between age and forced-choice scores using children (Collier et al., 2008). However, these correlations are not strong enough to invalidate the test based on age. In the current study, only 8% of the variance between age and forced-choice scores was shared. The association in children was much higher, with 16% of the variance being shared (Collier et al., 2008). Although future research is needed, it is likely that forced-choice scores will decrease slightly as age increases. The decrease in forced-choice scores as age increases may be a result of cognitive decline brought on by the normal aging process. However, the age associated decline does not appear to be strong enough to invalidate the 21-Item Test forced-choice component or change an individual’s classification from normal to biased.

In terms of cognitive and adaptive functioning, the current study yielded positive correlations between forced-choice scores and scores on the TOS and the SPMSQ. This finding shows that as cognitive and adaptive functioning decreases, forced-choice scores also decrease. The correlations were positive because the data for the TOS and SPMSQ were entered as negative values. Even though forced-choice scores decreased as cognitive impairment decreased, cognitive impairment was at a severe level before forced-choice scores went down sufficiently to reach the suspicious classification. It is important to note that it is the clinician’s responsibility to determine the severity of cognitive impairment before administering the 21-Item Test.
someone is obviously demented, perhaps a SVT is not needed as part of the test battery since Iverson (1998) has noted that elderly persons with dementia are likely to score below the cut-offs for suspicious performance.

**Limitations of the Study**

The current study was not without limitations. For example, the total sample size of elderly participants was small, especially since the community group consisted of only 20 individuals. Also, the current sample consisted of mainly one ethnic group. There were 72 Caucasians and one African American female. Obviously, one individual does not represent the typical performance of the 21-Item Test for the African American population. Also, there were only 21 males in the total sample. However, it is important to note that there are many more elderly females than males according to census data (Wechsler, 2008). Still, future research should attempt utilize larger and more diverse samples in terms of gender and ethnicity.

The community sample was atypical because participants had a mean education of 14.30 years ($SD = 3.89$, *Range* 8 to 22), which is well beyond that of most elderly individuals. To illustrate this point, Kaufman and Lichtenberger (1999) report that the 1995 standardization sample of the WAIS-III (Psychological Corporation, 1997), which was based on United States Census data, shows less than 30% of elderly individuals have 13 or more years of education. However, 45 percent of the community sample in this investigation had more than 13 years of education. In addition to possessing high levels of education, the community sample was in optimal health. They had no diagnosed chronic illnesses and none were taking medication for neurological or major physical impairments. Finally, the community sample was made up of participants who resided in rural areas. Therefore, future investigations should include elderly
samples with population-appropriate education, those who have age-related minor health issues (e.g., high cholesterol, hypertension, or arthritis) and those taking medication, as well as individuals residing in urban areas.

Another limitation of the present investigation was that only specificity to response bias during administration of the 21-Item Test was measured, not sensitivity. Sensitivity is the proportion of individuals responding in a biased manner who are correctly identified as responding in a biased manner, whereas specificity is the proportion of individuals who do not respond in a biased manner who are correctly identified as responding in an unbiased manner. It would have been ideal to include a group of elderly persons who were deliberately trying to fake or exaggerate a memory deficit to obtain a disability settlement or some other benefit such as special privileges in the nursing home, medications, or even permission to avoid scheduled activities like group therapy or cleaning their rooms. However, such a sample is difficult to find. Therefore, future research should include elderly individuals who receive financial incentives to malinger or are instructed and paid to malinger. For example, a study might include elderly individuals who are instructed to demonstrate poor effort on neuropsychological tests using a scenario for the participants to read. The scenario may include a situation in which the elderly reader is trying to portray a memory deficit to evade criminal prosecution, or obtain workers compensation, social security disability, or government coverage of his or her medical expenses. The study could also have a similar design to research by Iverson et al. (1991) where individuals were asked to imagine a psychological evaluation situation in which their only option was to feign memory impairment. Regardless of the design utilized, using the 21-Item Test to detect biased responding by elderly individuals instructed to feign cognitive deficits could provide
useful data regarding the sensitivity of the 21-Item Test forced-choice component.

Another limitation of the current investigation is the use of the SPMSQ. It may not have been the best choice even though it has good diagnostic precision for the detection of organic dysfunction (Wolber et al., 1984) and has demonstrated moderate to high sensitivity and specificity (Erkinjuntti et al., 1987). For example, in a sample of 119 community-dwelling elderly the SPMSQ yielded 66.7% sensitivity to cognitive impairment and 100% specificity. Also, the SPMSQ yielded a sensitivity of 86.2% and a specificity of 99% in a sample of 282 medical inpatients (Erkinjuntti et al., 1987). Despite these positive findings, the SPMSQ is not as widely used as the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), a test that might have been the better choice. In fact, the MMSE is the most widely accepted cognitive screening examination (Sumerall, Heaton, & Ryan, 2008) and is a more comprehensive measure of cognitive functioning than the SPMSQ. The SPMSQ only measures the participant’s distant memory, temporal orientation, and attention concentration (Soderqvist, Stromberg, Ponzer, & Tidermark, 2006). Thus, it is possible that some patients with significant cognitive impairments were not identified by the SPMSQ.

The MMSE assesses “temporal orientation, geographic orientation, short-term memory, attention, visual confrontation naming, repetition, language comprehension, reading, writing, and copying designs” (Sumerall et al., 2008, p. 286). The downside to using the MMSE is that it requires considerably more time and examiner skill to administer than the SPMSQ. However, future researchers might consider using the MMSE in place of the SPMSQ.
A final limitation is that the majority of nursing home residents were on medication and some of these medications have the potential to produce cognitive impairment as a side-effect. Thus, their performance on the assessment measures may have been affected by medication side-effects. Although names of medication for each nursing home patient were obtained, side effects were not taken into account for the current study. It would not have been possible to control for medication side-effects in the present study due to the fact that so few individuals were taking the same drugs.

Although other research has obtained normative data on the 21-Item Test, the current study is the first to do so with elderly samples. The results of the current study indicate that healthy elderly individuals and elderly individuals with cognitive impairment perform well on the 21-Item Test. In fact, 100% of community-dwelling elderly and 90% of elderly individuals residing in a nursing home achieved optimal (≥15) scores on the forced-choice component of the test. Also, the current study showed that 21-Item Test forced-choice component scores decreased as age increased. However, this finding was not significant enough to invalidate the test based on age. Finally, the results of the current study indicate that scores on the forced-choice component of the 21-Item Test decreased as cognitive and adaptive functioning decreased. However, it is important to note that cognitive and adaptive functioning had to be very severe before forced-choice scores were affected. Based on the present study, it is concluded that in clinical assessment situations that demand the use of SVTs (e.g., personal litigation, disability determination) to support the validity of subjective memory complaints, any examinee who is capable of completing a battery of neuropsychological tests and interacting effectively with the examiner should earn a forced-choice score ≥7. If a score of ≤6 is obtained on the forced-choice
component of the 21-Item Test, the likelihood of symptom exaggeration and frank malingering must be seriously considered when formulating diagnostic conclusions.
SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES

REFERENCES


SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES


planning committee. *Archives of Clinical Neuropsychology* 20, 419-426. 


doi:10.1076/clin.12.1.84.1722


doi:10.1080/13854049508400486


SPECIFICITY OF THE 21-ITEM TEST IN TWO ELDERLY SAMPLES


CONSENT FORM

My name is Daniel Turpin, and I am a graduate student under the supervision of Dr. Joseph J. Ryan, working on my master’s degree in psychology at the University of Central Missouri (UCM). The research you are being asked to participate in is for my thesis project. I will be asking volunteers to answer a few questions about themselves (e.g., age, education, etc.) and then listen to and recall a brief list of everyday words. We are simply trying to collect information about how people 65 years of age or older perform on this test. Your (name of nursing home/senior citizens program director) has invited me into (name of nursing home/senior citizens program) because he/she feels this is a worthwhile project. When this research is completed, I will have acquired enough valuable information to start writing my thesis, a requirement for completion for the master’s degree.

There are no serious risks to you if you volunteer for this study. The task will not exceed 20 minutes and this includes completing this consent form and answering a few personal questions. Testing will take place in a room assigned by the nursing home director or senior citizens program director. The results of testing will be provided to you if you desire to know your score (i.e., the number of words you recalled correctly).

Participation in this study is completely voluntary. You have the right to withdraw from this study at any time, without consequence. Also, there are no penalties for declining to participate. Finally, you have the right to withdraw your data at the end of this session.

Benefits from this study include being a participant in psychological research. Benefits also include interaction with a visitor from UCM.

From the beginning of the study, all test data will be kept in a secure, locked cabinet that may be accessed only by Dr. Ryan or me. Once all the volunteers have been tested, the data will be entered into a computer program for storage and analysis at UCM. At that time, all identifying information will be eliminated and each participant will be identified by a number only.

Your signed consent is required before you can participate in this study. If you have any questions about your rights as a research participant, please contact UCM's Human Subjects Protection Program (660 543-4621). You may also contact Joseph J. Ryan (660-543-4185 or ryan@ucmo.edu), Daniel Turpin (660-238-0043 or dxt63990@ucmo.edu), or (name and contact information of nursing home/senior citizens program director).

By signing below, you are indicating that you have read this form and agree to be tested by Daniel Turpin, a UCM graduate student. Please indicate your consent by signing this form. You may keep a copy of this consent form.

Participant Signature: ______________________________ Date: __________________

Name (please print): ______________________________
APPENDIX B

Demographic Information

Identification Number: _______________

Education: _______________

Age: _______________

Sex: Male Female

Ethnicity: __________________________

Current living situation:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Current Diagnosis: __________________________

Current Medication(s): __________________________