Analysis of Group Differences and Predictors of Hooper Visual Organization Test Scores

Michael R. DeVries
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Clinical Psychology Commons, Cognition and Perception Commons, and the Counseling Psychology Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1027

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
ANALYSIS OF GROUP DIFFERENCES AND PREDICTORS OF HOOPER VISUAL ORGANIZATION TEST SCORES

by

Michael R. DeVries

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Counselor Education and Counseling Psychology
Dr. Patrick Munley, Adviser

Western Michigan University
Kalamazoo, Michigan
August 2005
The Hooper Visual Organization Test (VOT) is described in the manual as a screening instrument that measures the ability to organize visual stimuli (Hooper, 1983). The VOT is identified as being particularly sensitive to neurological impairment. Studies to determine the criterion and construct validity of the VOT have examined its usefulness in distinguishing between individuals with neurological impairment from those with other disorders. Few studies have included samples from normal, psychiatrically impaired, and neurologically impaired populations in determining the VOT’s usefulness in identifying neurologically impaired individuals. Furthermore, as neuropsychology has moved away from the understanding of neurological impairment as a unitary concept, research on the VOT has focused on understanding the cognitive constructs that are related to VOT scores.

The current study utilized an archival sample of test protocols of 146 individuals without psychiatric or neurologic impairment, 92 individuals with psychiatric impairment, and 100 individuals with neurological impairment to evaluate the utility of the VOT in differentiating between the three groups through quantitative and qualitative scoring. Archival test protocols included collateral neuropsychological test data including the Symbol Digit Modality Test (SDMT) scores, Wechsler Adult Intelligence Scale –
Revised Performance IQ (PIQ) scores, and Raven Matrices Short Form scores. Scores from these measures were used to predict VOT scores after controlling for age and years of education. Additionally, individual VOT items were evaluated for their suitability and ability to distinguish between the three groups. The results showed that VOT mean scores were significantly higher in the normal group than in the psychiatrically and neurologically impaired groups. The mean scores for the latter two groups were not significantly different from each other. When entered together in the regression analysis after controlling for age and years of education SDMT written scores, PIQ scores, and Raven Matrices Short Form scores were each significant unique predictors of VOT scores. Sixteen of the thirty total items showed significant differences between the three groups. Findings are discussed and suggestions made for possible revisions to the VOT.
ACKNOWLEDGEMENTS

Above all, I would like to thank God. Through Him all things are possible.

I would like to thank my wife Angie, daughter Joelle, and son Kellen. Their patience was extraordinary. My wife has heard the phrase “I have to go do my dissertation” countless times. She held her tongue and said “Ok” when I needed it. I thank you dear for your sacrifice. To Joelle and Kellen, though they did not understand, they too were patient. No child of three should know the word dissertation. I am sorry for the times I had to put you off.

To my committee, I thank you for your time, effort, and patience. My committee Chair, Patrick Munley Ph. D. suffered endless revisions and telephone calls. He made himself available more than anyone could ask. The credit for the completion of this dissertation is largely his. I could not have asked for a better advisor. Ennis Berker, Ph. D. provided the inspiration for this project. His class ignited my interest in neuropsychology and literally changed the course of my life. His time, energy, and expertise made this project possible. The third member of my committee, Kelly McDonnell, Ph. D., was willing to participate in my committee despite the nature of the project being outside her typical research interests. She provided valuable insight through fresh perspective and enthusiasm that enhanced the quality of the project.

Michael R. DeVries
TABLE OF CONTENTS

ACKNOWLEDGEMENTS.............................................................................................. ii
LIST OF TABLES............................................................................................................. vii
CHAPTER

I. INTRODUCTION........................................................................................................... 1

   Background..................................................................................................... 1

   Statement of the Problem................................................................. 9

   Purpose of the Study............................................................... 11

   Research Questions.............................................................................. 13

      Research Question 1............................................................... 13

      Research Question 2............................................................... 14

      Research Question 3............................................................... 14

   Significance of the Study........................................................... 16

   Definition of Terms........................................................................... 17

      Visual-spatial Abilities.......................................................... 17

      Normal Individuals................................................................... 18

      Neurological Impairment...................................................... 18

      Psychiatric Impairment......................................................... 18

      VOT........................................................................................... 19

      WAIS-R.................................................................................... 19

      SDMT....................................................................................... 20

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven Matrices Short Form</td>
<td>20</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>22</td>
</tr>
<tr>
<td>Factors Thought to Influence Scores on the VOT</td>
<td>22</td>
</tr>
<tr>
<td>Age</td>
<td>23</td>
</tr>
<tr>
<td>Education</td>
<td>28</td>
</tr>
<tr>
<td>Intelligence</td>
<td>31</td>
</tr>
<tr>
<td>Naming Ability</td>
<td>35</td>
</tr>
<tr>
<td>Visual-spatial Abilities</td>
<td>40</td>
</tr>
<tr>
<td>Validity</td>
<td>43</td>
</tr>
<tr>
<td>Generalized Impairment</td>
<td>43</td>
</tr>
<tr>
<td>Lesion Location and Laterality</td>
<td>55</td>
</tr>
<tr>
<td>Neurological versus Psychiatric Impairment</td>
<td>59</td>
</tr>
<tr>
<td>Item Analysis</td>
<td>63</td>
</tr>
<tr>
<td>Summary</td>
<td>67</td>
</tr>
<tr>
<td>III. METHOD</td>
<td>71</td>
</tr>
<tr>
<td>Overview</td>
<td>71</td>
</tr>
<tr>
<td>Archival File Review</td>
<td>72</td>
</tr>
<tr>
<td>Participants</td>
<td>74</td>
</tr>
<tr>
<td>Instruments</td>
<td>77</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

<table>
<thead>
<tr>
<th>Test Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hooper Visual Organization Test (VOT)</td>
<td>77</td>
</tr>
<tr>
<td>Wechsler Adult Intelligence Scale-Revised (WAIS-R)</td>
<td>79</td>
</tr>
<tr>
<td>Symbol Digit Modalities Test (SDMT)</td>
<td>81</td>
</tr>
<tr>
<td>Raven Progressive Matrices Short Form</td>
<td>84</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>85</td>
</tr>
</tbody>
</table>

IV. RESULTS AND ANALYSES.

<table>
<thead>
<tr>
<th>Analysis Type</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics</td>
<td>87</td>
</tr>
<tr>
<td>Correlations</td>
<td>89</td>
</tr>
<tr>
<td>Primary Analyses</td>
<td>91</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>91</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>97</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>102</td>
</tr>
</tbody>
</table>

V. DISCUSSION

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Questions</td>
<td>110</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>113</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>118</td>
</tr>
<tr>
<td>Item Analysis</td>
<td>121</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>129</td>
</tr>
<tr>
<td>Limitations</td>
<td>132</td>
</tr>
<tr>
<td>Directions for Future Research</td>
<td>134</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

REFERENCES .................................................................................................................... 137
APPENDIX .......................................................................................................................... 144
LIST OF TABLES

1. Entire Sample Descriptive Statistics................................................................. 88
2. Group Mean Scores on Neuropsychological Measures........................................ 88
3. Two-way Pearson Correlations Between Demographic and Criterion Variables.... 90
4. Isolate Responses by Group.................................................................................. 95
5. Bizarre Responses by Group.................................................................................. 95
6. Isolate and Bizarre Response Means and Standard Deviations.......................... 96
7. Percent Full-credit Responses for Individual Items by Group............................... 99
8. Correlations Between Items and Total Scale Score with Item Removed.............. 103
9. Whole Sample Two-tailed Pearson Correlations Between Criterion and Predictor Variables.......................................................................................................................... 104
10. Pearson Correlations Between VOT Scores and Predictor Variables by Group.... 105
11. Multiple Regression Model for Age, Education, SMDT Written, Performance IQ and Raven Matrices SF as Predictors of Raw Hooper VOT Scores for the Whole Sample.......................................................... 109
12. Hierarchical Regression Model Summary for the Whole Sample....................... 109
CHAPTER I

INTRODUCTION

Background

The Hooper Visual Organization Test (VOT) (Hooper 1983) is generally regarded as a test of perceptual organization and visual-spatial abilities (Lezak, 1995; Spreen & Strauss, 1998). The VOT has been widely used in neuropsychological assessment since it was first developed in 1958. It was originally developed as a screening tool for brain damage, however, researchers have since questioned the use of the VOT to identify generalized brain injury (Rathbun & Smith 1982). Research on the VOT has suggested that scores are related to visual-spatial abilities and attempts have been made to clarify what abilities are related to VOT scores. Additionally, item analyses have been performed that suggest that the VOT might benefit from revision (Merten & Beal 2000, Verma, Pershad, & Khanna 1993). The purpose of this study is threefold. First, the study utilizes archival data to study group differences in VOT performance between normal, psychiatrically impaired, and neurologically impaired individuals to address the use of the VOT to identify neurological impairment. Secondly, the relationship between VOT scores and other neuropsychological measures is examined to suggest which cognitive abilities may be related to VOT performance. Finally, an item analyses was performed to examine individual item and scale characteristics. By investigating these three areas, this study is intended to provide increased understanding of the utility of this common neuropsychological measure and suggest areas for future research.
The VOT requires the examinee to identify items that have been drawn on stimulus cards. Each item is drawn in several pieces with the pieces disconnected from each other and separated as if they had been cut apart. The task requires the subject to identify each of the thirty items either verbally or in writing and can be used in individual and group settings.

Scoring of the VOT includes one point credit for correct responses with some responses receiving half credit. The manual also identifies four categories of responses that can be used for qualitative scoring (Hooper, 1983). These four categories include: Isolative, Perseverative, Bizarre, and Neologistic. An Isolative answer refers to only part of the drawing to develop a response. A Perseverative response is a repetition of a response from a previous item. A Bizarre response is defined as one that lacks a direct relationship to the test item. A Neologistic response is one that consists of nonsense letters to form a word with no meaning to the examiner. When using quantitative scoring methods, a cutoff score may be used. Hooper suggests a cutoff score of 24 out of 30 when a high false positive rate is acceptable such as when identifying patients for further assessment in a population where impairment is common. A more conservative cutoff score of 20 is suggested when impairment is thought to be less likely in a given population and a lower false positive rate is desired. The manual also provides age and education adjusted T-scores based on Mason and Ganzler’s (1964) sample of an all male Veterans Administration population.

The provided norms based on a Veteran’s administration population were thought to be a “fairly” representative sample of the adult population in terms of the ages represented (Hooper, 1983). The original norms were based on 30 junior high school
students, 166 college students, and 28 residents of a home for the aged. Studies by Boyd (1981) and Wentworth-Rohr, Mackintosh, and Fialkoff (1974) suggested a relationship between VOT scores and age and level of education. Mason and Ganzler (1964) used a regression equation to correct for these factors and found that they accounted for 21 percent of the variance in raw VOT scores. This regression equation was used to provide the age and education corrected T-scores. Lezak (1995) reported that the VOT does not correlate with gender or education but has a moderate correlation with mental ability. Furthermore, she reports that the correlation with age is not clear citing Ska, Poissant, and Joanette (as cited in Lezak, 1995) who found that subjects aged 55 to 84 showed neither a correlation with age or education nor with visuoperceptual tests including line orientation, form discrimination, and cube drawing tasks.

The reliability of the VOT has been found to be quite high. The original reliability studies completed by Hooper and quoted in the manual show a split-half correlation coefficient of .82 in a population of college students and of .78 in a clinical population (Hooper, 1983). Gerson (1974) completed additional reliability studies and found a split-half correlation coefficient of .80 in a more diverse clinical population. The test-retest reliability of the VOT was not reported in the manual, however, Lezak (1982) reported a coefficient of concordance of .86 after both 6 and 12 months. Levin, Llabre, and Reisman (1991) found the one year test-retest reliability to be .68 in elderly subjects. Sawrie, Chelune, Naugle, and Luders (1996) found the 8 month test-retest reliability to be .75 in intractable epileptics.

The question of the validity of the VOT has been the focus of several research studies. The VOT is thought by some to be a test of generalized brain injury and
dysfunction. Others believe that it is better used as a test of perceptual organization or visual-spatial skills. Much of the research on the VOT has focused on answering questions regarding its validity in these two areas. Another focus of research has been the effect of naming ability on VOT performance. At first glance it would seem that naming plays some role in a participant's ability to provide the correct responses to each item of the VOT. Some research on the VOT has addressed the question of whether the VOT is affected by a participant's naming ability and to what degree.

VOT performance is thought by some to be particularly sensitive to brain injury and the test was originally developed as a screening tool to assess neurological impairment (Hooper, 1983; Spreen & Strauss, 1998). Hooper (1983) suggests that there are two types of measures of cognitive functioning. General measures are sensitive to deficits without regard for their location and are best used as screening tools for unknown or diffuse injury. Measures of specific deficits are sensitive to lesions in certain areas of the brain and are best used to address questions of the localization of an injury and an individual's pattern of strengths and weaknesses in areas of ability. Though not specifically stated by Hooper, general measures may tap several areas of functioning while specific measures may be sensitive to a more defined area of functioning such as naming ability alone. Hooper suggests that empirical support for the VOT being more rightly considered either a test of general or of specific neurological impairment was not significantly developed at the time the manual was revised in 1983. Spreen and Strauss (1998) noted that the validity of the VOT as a test for screening for brain damage has been the subject of some debate and the validity of the test in determining specific deficits and the localization of brain lesions may be of greater importance.
The debate over whether the VOT is most appropriately used as a test of general versus specific deficits is seen in the articles by Boyd (1981, 1982a) and Rathbun and Smith (1982). Boyd (1981) reported a 75 percent correct classification rate between brain-damaged patients and healthy controls when a cutoff score of 25 out of 30 was used on the VOT. Boyd referenced Wang (1977) who stated that the VOT is not merely a test of visual-spatial deficits or localized right hemisphere damage but is better used as a general measure of dysfunction. Wang had found that the VOT was effective in discriminating between “organic” and “normal” patients or those with and without brain damage respectively. Rathbun and Smith (1982) responded to Boyd (1981) stating that the VOT is not appropriate for general screening of neurological deficits as those functions thought to affect VOT performance are often spared in left hemisphere and right frontal lesions. While acknowledging both common uses, two well-respected neuropsychological references regarding assessment tools take differing stances to some degree. Lezak (1995) notes that 11 failed items on the VOT typically indicates the presence of brain damage. Spreen and Strauss (1998) comment that they do not use the VOT to test for the presence of brain damage but for an evaluation of deficits in perceptual organization.

Some authors have compared the VOT to other measures that are thought to be sensitive to neurological impairments. In assessing an inpatient substance abuse population, McCaffrey, Krahula, Heimberg, Keller, and Purcell (1988) found that the VOT was least sensitive of the three tests used when compared to the Symbol Digit Modalities Test (SDMT) (Smith, 1991) and the Trail Making Test (Reitan, 1992). The latter two instruments were also found to have greater correlations between them than did
the VOT with either test. Sterne (1973) was noted to have found that the Trail Making Test part B was more likely to contribute to the identification of brain injured patients than was the VOT. Since the Trail Making Test and the SDMT have a higher correlation (approximately .60) McCaffery et al. suggest they share convergent validity. This would imply that the SDMT would also be a better indicator of brain injury than the VOT.

In the VOT manual, Hooper (1983) cites studies by Walker (1956, 1957) suggesting that the VOT is effective in discriminating between neurologically impaired, schizophrenic, and normal populations, specifically with the addition of a qualitative analysis of errors. Walker found that schizophrenic patients made more “Isolate” responses, those that refer to only one part of the VOT item, than did neurologically impaired patients. Gerson (1974) found that when the level of general intelligence was controlled for, the VOT was not effective in differentiating between schizophrenic patients and normal controls. Hooper (1983) concluded that the VOT was found to be effective in differentiating neurologically impaired individuals in normal and psychiatric populations but was not appropriate for screening for schizophrenia.

The determination that the VOT requires visual reasoning ability is based on studies which found that it correlates with other measures of visual-spatial abilities such as the Block Design and Picture Completion Subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Greve, Lindberg, Bianchini, & Adams, 2000; Paul et al., 2001; Seidel, 1994). Studies attempting to determine the validity of the VOT as a measure of visual-spatial abilities have often addressed the question of the effect of naming ability on VOT performance as well. Paul et al. (2001) found through regression analysis that over 60% of the variance on the VOT was accounted for by performance on the Block Design.
Subtest of the WAIS-R while performance on the Boston Naming Test (BNT) did not contribute significantly. They examined a population of patients with Vascular dementia who scored most poorly on the BNT compared to other cognitive measures. Ricker and Axelrod (1995) compared performance on the VOT with WAIS-R Subtests, the Multilingual Aphasia Examination Visual Naming Test (VNT), and a naming test composed of the reassembled items from the VOT. They report that hierarchical regression analysis revealed that a perceptual organization factor accounted for 48% of the variance in VOT performance. Confrontational naming (VNT) was said to account for 11% of the variance while performance on the reassembled VOT item-naming task was not shown to be significantly related to VOT performance. Paolo, Cluff, and Ryan (1996) used multiple regression analysis to compare the effects of perceptual organization and naming abilities on VOT performance using the WAIS-R and BNT. They found that perceptual organization abilities accounted for 44% of the variance while BNT performance accounted for 5%. Additionally, they compared participants with impaired naming abilities with those with normal naming performance by matching them by their scores on perceptual organization. They found no significant difference between the impaired and intact naming groups. The authors of these studies concluded that the VOT is primarily a test of visual-spatial and perceptual organization abilities and is not significantly affected by naming abilities (Paolo et al., 1996; Paul et al., 2001; Ricker & Axelrod, 1995).

Greve et al. (2000) found that perceptual processes uniquely accounted for 35% of the variance in VOT scores while naming accounted for 10%. They noted that while a smaller percentage, naming did account for a significant portion of the variance.
Practically, it was suggested that the VOT items are simple enough in their naming requirements that naming ability does not impact performance.

Johnstone and Wilhelm (1997) postulated that the VOT should be considered a measure of global visual-spatial intelligence. In their factor analysis it loaded on a similar factor to the Performance IQ Subtests of the WAIS-R. The VOT was found to have similar correlations to Performance IQ subtests as the individual Subtests do to each other suggesting that it measures a similar construct. The VOT was not found to load on a separate factor, which the authors suggest, might have indicated that it was measuring a distinct construct such as visual-spatial integration.

With the association between VOT performance and visual-spatial abilities it is not surprising that many have attempted to determine if a certain lesion location appears to most affect VOT scores. Nadler, Grace, White, Butters, and Malloy (1996) noted that “Despite the well documented relationship between visuospatial abilities and right hemisphere function, the literature has not supported a right hemisphere association with HVOT performance” (p. 223). Wang (1977) found a trend toward lower scores on the VOT in patients with lesions in the right hemisphere. Nadler et al. (1996) found that geriatric patients with right hemispheric Cerebral Vascular Accidents (CVA) scored more poorly on the VOT than did those with left hemisphere CVA. They also performed a qualitative analysis of the responses made by the two groups of patients and found that patients with left hemisphere lesions made more language based errors and those with right hemisphere lesions made more unformed or no response errors. They determined that patients with right CVA typically scored worse on the VOT though those with left CVA also had impaired performance. Furthermore, the two could be differentiated by a
qualitative analysis of the errors. Boyd (1981) found no differences in scores based on laterality of lesion. Commenting on Boyd’s article, Rathbun and Smith (1982) suggested that functions necessary for VOT performance are often spared in cases of left hemisphere and right frontal lesions and impaired in individuals with right posterior lesions. Fitz et al. (1992) found differences in VOT scores between those with right parietal lesions compared to those in other areas when scores were adjusted for age and education. Lewis et al. (1997) used an African American sample and determined that VOT performance was primarily affected by lesions to the right anterior region of the brain. They stated that while this finding supports the hypothesis that the VOT is differentially sensitive to regional location of brain lesions, it challenges the findings that it is most sensitive to right posterior damage.

Statement of the Problem

The Hooper Visual Organization Test (VOT) is generally regarded as a test of perceptual organization and visual-spatial abilities (Lezak, 1995; Spreen & Strauss, 1998). Despite the reported usefulness of the VOT in determining visual-spatial deficits, some suggest it would benefit from revision, partially due to a lack of sufficient psychometric analysis of the test at the time of its development. For example, it has been suggested that certain items are not ordered according to their level of difficulty and have poor scoring guidelines (Merten & Beal, 2000). Researchers have debated the best use of the test. In accordance with its original purpose, some believed the VOT to be a measure of general impairment and therefore a good screener to determine if individuals require more intense testing for neurological impairment. Additionally, the test was thought to differentiate
between those with psychopathology versus those with neurological impairment. It has also been proposed that the test merely measures some type of visual-spatial ability which may be unaffected by certain neurological impairments, therefore spoiling the utility of the VOT as a general screening tool. Despite its use since 1958, these questions have not been fully addressed or answered in the literature.

Item analysis has been completed on the VOT in two prior reported studies. A study by Verma et al. (1993) considered normal and psychiatric patients in an Indian population but did not consider those with brain injury. They concluded that most of the items performed satisfactorily in regards to item consistency and item discrimination values in this population with only a few items showing a degree of deviancy thought to be within tolerable limits. They did report that their analysis showed some of the items to be too easy and others too hard (at the beginning and end of the test respectively). Additionally, they found that the items were not arranged in the order of difficulty according to the percentage correct for each item shown in their sample. Merten and Beal (2000) considered neurologically impaired individuals in a German-speaking population without comparison to a control group. They suggested several changes in scoring, item inclusion, and item order.

The best use of the VOT has also been investigated. Initially, the test was developed as a measure of generalized brain injury (Hooper, 1983; Spreen & Strauss, 1998). Several investigators have tried to determine the efficacy of the VOT in differentiating between persons with brain injury, those with psychiatric illnesses, and those with no known impairment (normals). While the VOT is now generally accepted as
a measure of visual-spatial abilities rather than a general measure of impairment, the question remains as to its usefulness in differentiating between these three groups.

Finally, some question the degree to which the VOT measures visual-spatial abilities versus a more general ability, such as a spatial intelligence. The focus of neuropsychology as a whole has shifted from searching for measures of general impairment for the purpose of identifying brain injured persons to understanding which specific brain functions a test measures. The field of neuropsychology focused for some time on the “localization” of functions for the purpose of using neuropsychological measures to identify which areas of the brain may be injured in a certain patient. Researchers were therefore interested in correlating performance on certain measures with different areas of the brain. This task was attempted with some success regarding the VOT, however, the field has since been changed again by the advent of neuroimaging that is more capable of identifying injured brain tissue. The focus of neuropsychology has shifted somewhat from a primary goal of identifying and localizing injury to contributing to the rehabilitation of patients through identifying specific functional deficits and suggesting rehabilitation or coping strategies. Clarifying the specific aspects of visual-spatial abilities that are tapped by the VOT would make the test more useful in today’s practice of neuropsychology.

Purpose of the Study

This study consists of an analysis of archival data from a large database of test data on individuals with known neurological impairment, psychiatric diagnoses, and individuals with no known neurologic or psychiatric impairment for the purpose of
evaluating the VOT and adding to the current body of research on this measure. There are
three main areas that are investigated in this study. First, the ability of the VOT to
distinguish between normal, non-neurologically impaired persons from those who have a
known neurological impairment and those with psychiatric disorders will be studied.
Investigation of the VOT in these populations will provide additional reference data to
current VOT users and help determine its usefulness in differentiating between the three
groups.

Second, this study will include an item analysis, both qualitative and quantitative,
to provide information about the nature of the individual VOT items. This will provide
additional data should a revision of the test be undertaken. Specifically, previous item
analyses have been completed using non-English speaking populations. While language
functioning is not thought to be significantly involved in VOT performance, the role of
naming abilities has been questioned. This study will add data drawn from an English
speaking population to that from previous studies drawing on German and Indian
speaking populations.

Third, the association of VOT performance with other tests of visual-spatial
abilities will be examined through comparisons with scores on other measures thought to
tap these abilities. This will further clarify the cognitive functions that are tapped by the
VOT. This is specifically necessary as neuropsychological assessment continues to
develop. While initially, tests were designed to identify brain damage, they were
increasingly asked to measure specific functions that could be localized. With the rapid
development of imaging technology, localization of injury is less important, however,
clearly understanding functional deficits remains important as it relates to patient outcomes and rehabilitation strategies.

Research Questions

Research Question 1

Is there a significant difference in VOT scores between three groups of individuals consisting of individuals without neurological or psychiatric impairment, individuals with psychiatric impairment, and individuals with known neurological impairment? The first research question was investigated by examining the following three questions and their associated null hypotheses:

a. Is there a significant difference between the groups in mean VOT raw scores?

b. Is there a significant difference between the groups in mean VOT scores after age and education correction?

c. Is there a difference between the three groups that is evident based on qualitative scoring?

Null Hypothesis 1a

There is no significant difference in mean VOT raw scores between normal controls, psychiatrically impaired and neurologically impaired individuals.
**Null Hypothesis 1b**

There is no significant difference in mean age and education corrected VOT scores, between normal controls, psychiatrically impaired and neurologically impaired individuals.

**Null Hypothesis 1c:**

There is no difference between normal controls, psychiatrically impaired and neurologically impaired groups that is evident based on qualitative scoring.

**Research Question 2**

To what extent do individuals with diagnosed neurological impairment, psychiatric impairment and those with no impairment differ in their responses to individual items on the VOT?

Null hypothesis 2: Normal individuals and individuals diagnosed with neurological impairment or psychiatric disorder will not demonstrate significant differences in frequency of correct responses given to individual items of the Hooper VOT.

**Research Question 3**

To what extent does the VOT correlate with other neuropsychological measures including the Verbal and Performance IQ measures of the Wechsler Adult Intelligence Scale-Revised (WAIS-R), Raven’s Matrices Short Form (SF), and the SDMT? This final
research question was investigated by examining the following two questions and their associated null hypotheses:

a. Are there significant simple correlations between subject scores on the WAIS-R Performance and Verbal IQ’s, Raven’s Matrices Short Form, and the SDMT, and either uncorrected or age and education corrected VOT scores across the entire sample or within the individual groups?

b. Using hierarchical multiple regression analysis, after the effects of age and education are controlled for, are subject scores on the Performance and Verbal IQ’s, Raven’s Matrices SF, and the SDMT found to be significant predictors of VOT scores.

**Null hypothesis 3a**

There are no significant simple correlations between scores on the Performance IQ and Verbal IQ, Raven’s Matrices SF, and the SDMT and uncorrected or age and education corrected VOT scores.

**Null hypothesis 3b**

Scores on the WAIS-R Performance IQ, and Verbal IQ, Raven’s Matrices SF, and the SDMT will not be significant predictors of VOT scores after age and education are controlled for when examining the entire data set using hierarchical regression analysis.
Significance of the Study

At this point in time there remain several important questions regarding the VOT. Despite the move away from using neuropsychology to determine brain injury, there are still clinicians who are called to perform this task such as those in psychiatric care environments or in the case of litigation for mild to moderate brain injury. It remains unclear if the VOT is useful in differentiating normal persons from psychiatrically impaired or neurologically impaired persons. This study used a large data sample to examine the usefulness of the VOT in differentiating between these three groups.

The manner in which the test was developed has left questions about the item characteristics and their organization. These questions suggest a revision of the test with possible modifications to the included items and their order. This study will add to literature in this area as it is the first to include an item analysis using English speaking subjects that come from three categories of interest, normal, psychiatrically impaired, and neurologically impaired adults.

Finally, the association of VOT performance with other measures of visual-spatial abilities requires clarification specifically as the availability of advanced neuroimaging has caused a shift in neuropsychology from attempts to localize dysfunction to increasing understanding of specific cognitive abilities and the implications for an individual’s future functioning. This study will examine the correlations between the VOT and other neuropsychological measures and measures of visual-spatial ability and reasoning to add to knowledge of the specific abilities related to VOT performance.
Definition of terms

*Visual-spatial Abilities*

In this study, the term visual-spatial abilities is used as a general term referring to cognitive functions related to visual processing and manipulation of external stimuli. Stedman’s Medical Dictionary defines the word visuospatial as “the ability to comprehend and conceptualize visual representations and spatial relationships in learning and performing a task” (Pugh, 2000, p. 1974). Spreen and Strauss (1998) include the VOT in their compendium as a test of visual-perceptual skills. The word visual, as used in this study, refers to the various processes involved in normal perception through vision. This includes perception of color, depth, and form as well as object recognition (Rapp, 2001). These abilities are required, at least minimally, to complete the VOT. In the INS Dictionary of Neuropsychology, spatial orientation is described as “the ability to judge the orientation of information in either two or three dimensional space” (Loring, 1999, p. 148). The term “visual-spatial sketch pad” is a term used for one of the processes involved in working memory allowing the temporary storage of visual information (Loring, 1999, p. 106). This ability allows the manipulation of this information in space. The term visual-spatial ability refers to these visual and spatial abilities considered as a group.
Normal Individuals

Participants in the study without known or diagnosed neurological impairment or psychiatric impairment were labeled “normal.” Participants in the normal group were individuals who had completed the neuropsychological testing measures who had no history of psychiatric or neurological impairment. Participants in the normal group were primarily individuals who had been tested as a part of graduate student training in neuropsychological assessment.

Neurological Impairment

Neurological impairment refers to impairment of brain function due to disease, lesion, concussion injury, or other traumatic event such that brain functions are thought to be impaired. For the purpose of subject selection, individuals were considered neurologically impaired if information in their archival record indicated that they presented for testing due to an established diagnosis of a neurological impairment such as those listed above, e.g. a diagnosed brain tumor. Additionally, individuals were considered neurologically impaired if they presented for testing due to a condition known to lead to neurological impairment, e.g. a significant car accident with subsequent loss of consciousness.

Psychiatric Impairment

Psychiatric impairment was defined as having a psychiatric diagnosis and/or significant psychiatric symptomatology that was recorded in the history section of the
neuropsychological testing report in the subject’s archival record. Participants classified as having a psychiatric impairment had been assessed as part of an inpatient psychiatric hospitalization.

**VOT**

The term VOT refers to the Hooper Visual Organization Test (Hooper, 1983). The test was originally developed by H. Elston Hooper and published in 1958. Unless otherwise noted, in this text, the term VOT refers to the most recent revision published by Western Psychological Services in 1983. In the manual the VOT is described as:

a brief screening instrument designed to measure the ability of adolescents and adults to organize visual stimuli, a task that is particularly sensitive to neurological impairment. The test consists of 30 line drawings depicting simple objects which have been cut into pieces and re-arranged in a puzzle-like fashion. The respondent is asked to identify what each object would be if it were put back together correctly (p. 1).

**WAIS-R**

WAIS-R stands for the Wechsler Adult Intelligence Scale- Revised (Wechsler, 1981). According to the Manual, the WAIS-R is a revision and re-standardization of the original Wechsler Adult Intelligence Scale (WAIS) which was published in 1955. Spreen and Strauss (1998) describe the WAIS-R as a measure of general intelligence that is often considered the standard in intelligence testing. Furthermore, it is reported to give information about the overall intellectual function of an individual. The test consists of
both verbal and performance measures that yield a Full Scale Intelligence Quotient (FSIQ), a Verbal IQ, and a Performance IQ.

SDMT

The acronym SDMT stands for the Symbol Digit Modalities Test (Smith, 1982) which was originally developed by Aaron Smith in 1973 (Spreen & Strauss, 1998). The test was developed as a screening measure of cerebral dysfunction in children and adults. It requires the conversion of meaningless geometric designs into written and/or oral number responses (Smith, 1982). Examinees are allowed ninety seconds to perform the task and the score is the number of correct response in the allotted time. The test is similar in principle to the Digit Symbol subtest on the WAIS-R but requires that the examinee give a number rather than a symbol response, so it can be given orally as well as in a written form (Spreen & Strauss, 1998). This unique aspect of the test allows the differentiation between those who score low on the measure due to motor deficits that interfere with writing versus those who score low solely based on their inability to perform the mental and verbal aspects of the task. The manual suggests that the nature of the task requires a variety of mental abilities and substitution tasks of this kind have been shown to be highly sensitive to cerebral dysfunction (Smith, 1982).

Raven Matrices Short Form

The Raven Matrices referred to in this study is the short form of the Raven’s Progressive Matrices (RPM) test (Channell et al., 1997; Raven, 1996) or Raven Matrices SF. The RPM is described as a test of visual reasoning (Spreen & Strauss, 1998). Spreen
and Strauss describe the RPM as requiring the examinee to infer a rule relating to a collection of elements, and then use the rule to generate the next item in a series. The problems become progressively more difficult. The items are grouped into five sets with 12 items per set. The short form is 30 items selected from the total of 60 with 6 items selected from each set. This short form is used as part of the Michigan Neuropsychological Battery (MNB) (Smith & Berker, n.d.) and the norms used are provided in the MNB manual. The test is not timed.
CHAPTER 2

LITERATURE REVIEW

Factors Thought to Influence Performance on the VOT

The VOT manual (Hooper, 1983) identifies several factors that may affect an individual’s score on the VOT. In research, these factors or variables that affect scores on the VOT are dealt with in one of two ways. Confounding variables are identified so that their effects can be minimized while additional variables are identified so that the specificity of the test can be better understood. The VOT was originally intended to differentiate between different groups such as normals, psychiatrically impaired, and neurologically impaired individuals. Variables such as age, level of education, and pre-morbid intellectual functioning were considered to be possible confounding variables to this intended purpose. Corrections were established based on age and education and published in the manual (Hooper, 1983; Mason & Ganzler, 1964). Other variables such as pre-morbid visual-spatial ability, language skills and naming ability, and the location of the lesion or nature of the disease process, could have also confounded research findings if one was trying to show the efficacy of the VOT in distinguishing between impaired and unimpaired groups. As theories changed over time regarding the extent to which brain damage and neurological impairments were generalized versus localized, these latter variables have generally been studied with the intent of understanding the specificity of the VOT. For example, by separating individuals into groups with similar lesion or impairment characteristics, researchers have been able to suggest connections between
VOT scores and the location and type of impairment. Each of the identified factors will be reviewed in turn to discuss what is known about their effects on VOT scores.

**Age**

The original sample referenced in the VOT Manual (Hooper, 1983) was taken from three groups, 30 junior high school students, 166 college students, and 28 residents from a home for the aged. Hooper’s original manual (Hooper, 1958) claimed that VOT scores are unrelated to age, education, intelligence, and sex. The frequency distribution in the 1983 manual, however, showed that the first two groups, the junior high and college students have similar VOT score distributions. The mode for the older adults is much lower at 19.5 versus 27.5 for the student groups. A cut off score of 20 correctly classified 97% of the junior high school students and 94% of the college students but only 33% of the older adults. A cut off score of 25 yielded correct classification levels of 66%, 72%, and 21% respectively. It was later pointed out that Hooper’s own data contradicted the assumption that VOT scores are unaffected by age (Wentworth-Rohr et al., 1974).

Mason and Ganzler (1964) obtained a more continuous sample with respect to age from a Veterans Administration hospital. Patients, non-professional hospital staff, and hospital volunteers were administered the VOT. Participants were excluded if they reported histories suggesting mental illness, substance abuse, or brain damage. A sample size of 231 was reported. The authors recognized that the normal decline in cognitive abilities that can occur with normal aging could lead to misclassification of patients if corrections for age are not used in cognitive test scores. This assumption is also voiced by Boyd (1981) in justifying matching subjects based on age and other variables as part of...
his validity study of the VOT. In the Mason and Ganzler (1964) study, the authors used the Shipley Institute for Living Vocabulary score, age, and education in a multiple regression formula to predict an expected VOT score. They also developed a prediction equation without the Shipley Vocabulary score where age and education were used to predict an individual’s VOT score. These prediction equations are given in the VOT Manual (Hooper, 1983), and were used to develop age and education corrected scores provided in the appendix of the VOT Manual.

Wentworth-Rohr et al. (1974), citing the likely affect of confounding variables such as age, education, intelligence, and gender, studied their effects on VOT scores in four groups. They tested 200 psychiatric inpatients, 85 social worker nuns, 51 parochial school principals, and 61 undergraduate psychology students with the VOT using group administration procedures. They found a curvilinear relationship between age and the VOT. The slope of the function is said to be approximately zero from the teens to the mid thirties with a negative slope from the thirties onward. They used scores on intelligence measures to control for variations in intelligence in the different samples. The psychiatric group did show a slight but significant negative correlation between age and VOT scores between the ages of 14 and 36 years old. The older groups, showed much larger significant negative correlations varying from -.37 in the social worker nun group to -.69 for the principals. The authors report clear evidence that the VOT is affected by an individual’s age after intelligence level is controlled.

Tamkin and Jacobsen (1984) investigated the relationship of VOT scores to patient age, education, and IQ in a Veterans Administration (VA) hospital population. They noted that these variables were known to affect scores on various
neuropsychological tests. Their study provides age-based norms for the VOT. Their data was collected from 211 male psychiatric inpatient files, chosen from records of those who were administered the tests as part of routine procedures. Data was grouped into six age decades from the youngest decade of, age 20-29, to the oldest, age 70-79. Each age group consisted of 40 individuals with the exception of the oldest age group, for which only 11 individuals were available. They found that age was significantly correlated with VOT scores and explained a significant amount of the variance in the scores. A negative correlation between age and VOT scores was seen from the youngest age group with a mean VOT score of 23.7, through the oldest age group with a mean score of 14.09. The r for this correlation was -.50. Sterne (1973) reported -.42 for male veterans age 20-73 while Wentworth-Rohr et al. (1974) found -.52 in psychiatric inpatients age 36-77. Tamkin and Jacobsen (1984) found that multiple regression analysis showed that age accounted for 19.8 percent of the variance in VOT scores when education, age, and IQ together accounted for 46 percent. They concluded that age-base norms should be used to interpret VOT test scores in similar populations to their male inpatient sample.

While the previous studies have examined age effects on VOT scores in adults, Seidel (1994) investigated the validity of the VOT's use with children. In a normative sample of 211 children between the ages 5 and 11, he found a positive correlation between age and VOT scores. Children age 5 had a VOT score mean of 18.4 while children age 11 had a mean score of 24.1. This was consistent with the research cited by the author suggesting that the VOT measures abilities that vary with the developmental stages children undergo between the ages of 5 and 11 years old. The most significant differences in scores seen in this study are between children younger and older than age 8.
Smaller differences are seen within groups of children age 5 to 7 and within those ages 9 to 11.

On the other end of the age spectrum, two studies examined VOT performance in older adults. Walsh, Lichtenberg, and Rowe (1997) examined the performance of geriatric rehabilitation patients on the VOT due to the limited normative data available in this population. Data from three groups of older adults was examined. One hundred and forty-four patients who were referred for cognitive assessment were administered the VOT. Their ages ranged from 60 to 95 years old. Those who scored within normal limits on cognitive tasks, such as the Dementia Rating Scale, and showed no evidence or history of neurological impairment were placed in the cognitively intact group. Forty-six subjects were placed in the mildly impaired group based on their test scores. Sixty-six subjects were judged to have moderate to severe impairment and were grouped together as the third comparison group. One possible confounding variable was noted, the cognitively intact group had significantly more years of education (11.7 years) than the moderate to severely impaired group (9.3). The influence of education on VOT scores, however, has been found to be less than that of other variables, as will be discussed in the next section. The authors found a significant correlation between age and VOT scores ($r = -.34$). Using hierarchical multiple regression, age was found to be a significant predictor of VOT performance over and above the effect of education, which was also correlated with VOT performance. Even when the age of participants was restricted to older adults, as in this study, age was found to be a significant predictor of VOT performance suggesting the need for good normative information for this age group.
Nabors, Vangel, Lichtenberg, and Walsh (1997) studied 58 cognitively intact older adults and 59 cognitively impaired seniors. This study utilized an urban population of African-American and Caucasian-American older adults to examine the VOT’s utility in differentiating between cognitively intact and cognitively impaired groups. Additional normative data was also gathered regarding the performance of older adults on the VOT. The authors reported a significant difference between the comparison groups in their age and level of education. The cognitively intact group was found to be younger and better educated. No significant differences were found between the groups on the variables of gender or race. The mean age of the cognitively intact group was 76 years old. The mean level of education was 11 years. For the cognitively impaired group, the mean age was 79 years old with a mean education level of 8.8 years. A significant correlation with age was found for the total sample (r = -23). Age was found to be an independent predictor of VOT scores predicting 4.5 percent of the variance. This relationship between age and VOT scores was found in the cognitively impaired group and also when the entire sample was considered together, however, age was not found to be a significant predictor of VOT scores in the cognitively intact group. The authors suggested that larger sample sizes are needed to further delineate the effects of demographic variables, such as age, on VOT scores.

The sum of the research on the effect of age on VOT scores suggests a positive correlation from age five into the second to third decade of life. From age twenty or thirty the correlation reverses and there is a negative correlation between age and VOT scores that is consistent into the ninth decade. These correlations are significant and range from -.42 to -.52 in adults age 20 to 30 and beyond.
Hooper's original normative sample for the VOT included three distinct groups, two of which had predictable levels of education as they were groups of junior high and college students. While the original reported norms did not account for education, Mason and Ganzler (1964) noted that analysis of this original data suggested the need for norms based on age and education. As noted previously, their study developed a prediction equation for VOT scores based on these two variables. They found that in their new sample, these two predictor variables accounted for 21 percent of the variance in VOT scores. They noted, however, that when using the three variable prediction equation including the Shipley Institute of Living Scale Vocabulary Score varying the education across the entire range of scores from grade one to college only changed the predicted VOT score by one point.

Wentworth-Rohr et al. (1974) used college and psychiatric patient groups to examine VOT scores in participants who ranged in educational level from early high school to graduate school. When controlling for age and intellectual level, they found that education was unrelated to VOT scores in this combined group. Additionally, they visually compared their college group scores to a distribution of scores from a junior high group reported in the VOT manual as they did not have the necessary data for statistical analysis. This examination suggested complete overlap in the distributions with no difference that could be attributed to their educational level. They noted that although age and intelligence level were not controlled in the latter comparison, the expectation would be that these variables would be lower in the junior high group and would yield a
relatively lower distribution of scores, but this was not observed. They suggested that this was further evidence that there is not a significant relationship between VOT scores and educational level. This finding is consistent with work done by Boyd (1981). While the intent of Boyd’s study was to compare normal and brain damaged groups that were matched for age, education, and intellectual levels, he also found that VOT scores were not significantly correlated with educational level in either of his two groups.

In the study by Tamkin and Jacobsen (1984), it was found that among 211 male VA patients education was significantly correlated with VOT scores. Further examination provided that education was also correlated with age and when multiple regression analysis was completed to control for age, educational level did not account for a significant amount of variance in VOT scores. While the authors suggested the correction of VOT scores based on age and IQ, their study did not support a relationship between VOT scores and education.

In seeming contrast to previous findings, Nabors et al. (1997) found a significant correlation between education and VOT scores when examining a cognitively impaired group of older adults. They did not find a significant correlation in their cognitively intact group. They did find a significant correlation when the two groups were combined and reported that regression analysis suggested that education accounted for an independent 5.5 percent of the variance in VOT scores. It does not appear, however, that they accounted for the difference between groups in terms of their cognitive dysfunction. They reported that the two groups differed significantly in terms of educational level with the cognitively impaired group having a mean educational level of 8.8 years while the mean for the cognitively intact group was 11 years. It can be expected that since the cognitively
impaired group also did worse on the VOT that combining the groups could lead to a correlation between VOT and education which is an artifact of the a priori differences between the two groups.

Walsh et al. (1997) had a situation similar to that seen in the previous study with a differing result. They also examined data from older adults with varying degrees of cognitive impairment. Of their three groups, the cognitively intact group showed a significant difference in educational level from the moderately to severely impaired group. The intact group had a mean of 11.7 years of education, the mildly impaired group had a mean of 10.4 years, and the moderately to severely impaired group had a mean of 9.3 years. They found a significant correlation between education and VOT score in the entire sample, but when hierarchical regression analysis was performed using age and education scores, education was found not to be a significant predictor of VOT scores.

Despite the fact that there is little evidence that VOT scores are predicted by educational level, Walsh et al. suggested that the large degree of variation seen in the VOT scores within the cognitively intact and mildly impaired groups may be due to large variation in the educational level of subjects in the groups. Given the studies that have been completed and reviewed here, however, it seems more likely that variation in VOT scores reported to be associated with education may be related to individual differences in age, intelligence, and cognitive impairment. Both intelligence and cognitive impairment have been suggested to affect VOT performance and both variables will be discussed further.
The VOT manual offered some discussion of the connection between VOT scores and general intellect. The original manual reported correlations between the VOT and the Shipley Institute of Living Scale Vocabulary of .19 for junior high students and .34 for psychiatric inpatients. At very low levels of IQ, between 30 and 80, the VOT was found to have a correlation of .57 with the Wechsler-Bellevue Intelligence Scale. Examining the studies available at that time, the manual author noted a mild to moderate correlation between VOT and intelligence. Correlations between VOT scores and intelligence from .07 and .63 were reported. Additionally, the manual cautions that interpretation of the VOT should take into account that VOT scores have been shown to be related to variables such as age, educational level, and non-verbal intelligence. The manual also stated that in the extreme case of intellectual deficit, such as in people with “mental retardation,” the VOT should not be used (Hooper, 1983).

This latter point was made by Love in his study of New Zealand psychiatric patients (Love, 1970). Love saw 115 patients in two inpatient hospitals and a rehabilitation center and administered the VOT as part of a general assessment. He found that using a cut off score of 20 to determine the presence of neurological impairment yielded a false positive rate of 25 percent for the non-organic group (mentally ill) and a 30 percent false negative rate for the organic (neurologically impaired) group. Forty-eight percent of the 21 false positive results came from participants with IQ’s less than 75. It was suggested that the VOT not be used in patients whose IQ’s are know to be less than 75.
In the study by Wentworth-Rohr et al. (1974) the age controlled correlation between intelligence and VOT scores was measured in three groups, psychiatric patients, nuns, and college students. In this study, intelligence was measured by three different measures in the three groups. The Wide Range Vocabulary Test, Form B was used in the psychiatric group, the Otis-Lennon Intelligence Test was used in the group of nuns, and the Scholastic Aptitude Test in the College group. The correlations for the three groups were .50, .33, and .31 respectively. These findings were interpreted as confirming that from dull to very superior ranges, there is a significant positive relationship between intelligence and VOT scores.

Gerson (1974) investigated the validity of the VOT in his study of brain injured, functionally impaired, and normal groups. The brain injured group consisted of 16 individuals with chronic and acute brain injuries from toxic, degenerative, and traumatic etiologies. The functional group included 19 inpatients with schizophrenia. The normal group included medical inpatients without evidence of psychiatric symptoms and from hospital employees. The Quick Test was used to approximate IQ scores in all three groups. The IQ range for the subjects was from 63 to 116. Significant correlations between IQ and VOT scores were found in both the normal and the functionally impaired groups, .63 and .50 respectively. The .07 correlation found in the brain injured group was not significant. There was a significant mean difference in IQ between the normal group and the brain injured group but not between either of these two groups and the functionally impaired group. The author used IQ as a covariate in determining that the VOT scores were significantly different between groups even when the effects of IQ
were considered. This study confirmed the importance of considering IQ scores and their effect on VOT scores when studying the VOT.

One of the questions raised when comparing VOT scores to IQ is: What is an appropriate measure of IQ? Given that the VOT is said to be sensitive to nonverbal intelligence, should a measure such as the Full Scale IQ from the Wechsler Adult Intelligence Scale be used when the score includes measures of nonverbal intelligence? In several studies, such as Boyd (1981), investigators have used a measure of vocabulary or other verbal measure as an estimate of IQ or of premorbid IQ. In his sample of two inpatient groups, one with neurological impairment and one consisting of a mixture of psychiatrically impaired and medical inpatients, Boyd used the Peabody Picture Vocabulary Test. In this manner, he was able to obtain impaired and normal groups that were relatively equal in IQ as opposed to Gerson (1974) who’s groups showed significantly different IQ scores as measured by the Quick Test (Ammons & Ammons, 1962). Despite this difference in IQ measures, Boyd (1981) had results similar to Gerson’s when correlating IQ and VOT scores. Boyd found a significant correlation of .41 in the non-brain damaged group and a non-significant correlation of .25 in the brain damaged group.

Tamkin and Jacobsen (1984) studied the VOT using the Shipley Institute of Living Scale estimate of IQ. In examining six age groups, they found a significant correlation (.52) between VOT scores and intelligence. They reported a direct relationship between IQ and VOT scores with an r of .52. They concluded that VOT scores should be corrected for age and IQ in order to make appropriate interpretations.
A significant correlation between VOT scores and IQ appears to exist in children as well. Seidel (1994) found a correlation of .19 between a prorated Verbal IQ score and VOT scores in a normative sample of 207 children. The participants were randomly selected from volunteers at an elementary school. Participants were between the ages of 5 and 11 years and had prorated Verbal IQ's of >70. Intelligence scores were prorated from the Information, Similarities, Arithmetic, and Digit Span subtests of the WISC-R and from the WPPSI Information, Similarities, Arithmetic, and Sentences subtests. Despite the significant correlation, the authors note that Verbal IQ only accounted for 3 percent of the variance in VOT scores.

The most recent study comparing VOT scores to intelligence was completed by Greve et al. (2000). In this study, 98 patients who had suffered cerebral vascular accidents (CVA) were evaluated using several test measures. IQ was measured using the WAIS-R. They reported a bivariate correlation of .53 between WAIS-R full scale IQ scores and VOT scores. The correlation with the Verbal IQ scores was .38 while the correlation with Performance IQ scores was .65.

In general, the data suggests that VOT scores are significantly correlated with IQ scores, however, some studies show that this correlation is less strong or non-significant in groups with neurological impairment. This situation is likely due to the correlation between overall IQ scores and scores on tests of specific cognitive abilities. In other words, persons with high IQ scores, which include scores from measures of a variety of cognitive abilities, tend to perform well on other tests of individual cognitive abilities. It holds to reason that if the VOT measures nonverbal intelligence, the correlation between VOT scores and IQ would be less strong in neurologically impaired subjects whose
nonverbal intelligence is impaired because IQ scores often measure both nonverbal intelligence and verbal intelligence. The inclusion of verbal abilities, often thought to be more resistant to brain injury, in the IQ score may decrease the correlation with the VOT when nonverbal abilities are primarily impaired. This contrast might be expected to be more striking if the investigator uses a more purely verbal test to estimate participant IQ as this score would not necessarily be impaired when nonverbal ability is impaired.

**Naming Ability**

In examining the VOT, it may seem intuitive that good performance might require some type of visual-spatial ability. It might also come to mind that it requires at least basic verbal abilities. The manual noted that it is not a “pure measure of visual integration ability,” but rather, it requires “elementary vocabulary skills and the ability to label common objects” (Hooper, 1983). The ability to label common objects is often referred to as naming ability. Several of the more recent studies that have examined the VOT have attempted to determine the degree to which naming ability affects VOT performance. The obvious concern being that if the VOT requires a significant degree of naming ability, then it is less useful as a measure of visual processing or visual-spatial abilities.

The concern about the potentially confounding nature of naming ability in the use of the VOT as a test of visual-perceptual abilities is what motivated Seidel to control for naming as a factor in his study of the VOT in children though the effects of naming were not actually studied (Seidel, 1994). Additionally, he noted concern that naming ability is thought to improve with age, therefore any effects would be more important in children who may not have yet developed the prerequisite naming ability to perform the VOT. He
utilized a Modified VOT (MVOT). The MVOT consisted of the items from the VOT reassembled into their whole shapes. The children were administered the VOT and then the MVOT. The children were required to identify the items on the MVOT and their responses were scored according to the manual. MVOT scores were used along with prorated Verbal IQ scores from the WISC-R and WPPSI as covariates to control for variation in verbal and naming abilities. Unfortunately, MVOT scores were not compared to VOT scores to determine the relationship between the two and the possible contribution of the ability to name the specific items on the VOT to overall VOT performance.

The use of the reconstructed items from the VOT as a naming task was a novel idea with face validity when used as a covariate to control for possible effects of naming performance on VOT scores. In the study by Seidel, however, the scores on reconstructed items were not compared to VOT scores to determine if there was a relationship between the two measures. Ricker and Axelrod (1995) used a test of the reassembled VOT items to help determine the role of naming ability in VOT performance. They report two studies. In the first study, 50 male veterans, consecutively referred for neuropsychological evaluation were given the VOT, WAIS-R, and the Visual Naming Test (VNT) from the Multilingual Aphasia Exam (Benton & Hamsher, 1989). Hierarchical multiple regression was used to compare the contributions of WAIS-R and VNT performance in predicting age and education corrected VOT scores. The investigators found that regardless if the VNT scores were added first or after WAIS-R scores in the regression equation the VNT scores accounted for 11 percent of the variance in scores on the VOT. The investigators noted that this finding suggests that clinicians
should consider a patient’s general naming ability when utilizing the VOT. However, they also noted that their findings could not determine if the ability to name the specific items on the VOT was related to overall VOT performance. For this reason, they undertook a second study.

In the second study, Ricker and Axelrod (1995) attempted to determine if the ability to name the specific items from the VOT affects performance. They again tested 50 consecutively referred patients to the Psychology service at their VA center. They used a test of reassembled VOT items rendered by an artist. This new test measure was administered immediately following the VOT. When using VNT and VOT scores (essentially repeating the first study) regression analysis of this second sample yielded similar findings to the first. General naming ability was again found to explain a significant amount of the variation in VOT scores. However, when scores on the reconstructed version of the VOT were also entered into a regression equation, it was found that the subjects’ ability to name the specific items on the VOT was not significantly related to VOT scores. The authors concluded that the degree of object naming required on the VOT is not high enough to impair performance. They suggested that impaired general naming performance should not significantly impair VOT performance.

Paolo et al. (1996) replicated the initial study reported by Ricker and Axelrod (1995), but they substituted the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983) for the VNT. Paolo et al. (1996) studied 98 consecutively referred patients at an outpatient neurology clinic. They found that the BNT explained a significant amount of the variance in VOT scores (5%). They then constructed two
groups, one with impaired and one with intact naming ability. The two groups were matched based on perceptual organization abilities (measured by the Block Design and Picture Completion subtests of the WAIS-R), on the basis of age, and on the basis of educational level. They compared VOT scores between the two groups and found that while the group with impaired naming ability scored an average of two points lower than the group with intact naming abilities, this difference was not significant. They concluded that impaired naming performance as measured by the BNT does not significantly impair VOT performance.

Greve et al. (2000) reported results similar to the previous studies when examining the effects of confrontational naming ability on VOT performance. They examined 98 patients who suffered cerebrovascular accidents. The participants were given a neuropsychological test battery that included the VOT and the confrontational naming task of the Neurobehavioral Cognitive Status Exam (Cognistat) (Group, 1995). Using partial correlations between Cognistat naming and the VOT, the authors determined that naming uniquely accounted for 7 percent of the variance in VOT scores. The authors went on to divide the sample into three groups based on naming performance to further examine the connection between naming and the VOT. They found that the group that scored “impaired” on Cognistat naming was more globally impaired based on WAIS-R performance. When Perceptual Organization scores from the WAIS-R were used as a covariate to control for impaired perceptual abilities, it was found that there was no longer a significant effect of naming on VOT scores. Additionally, they found double dissociation of VOT and naming scores. This is to say that some patients had severely impaired naming but intact performance on the VOT. Other patients had the opposite
pattern of scores with intact naming ability but impaired performance on the VOT. This is said to suggest an independence of performance on the two tests. The presence of impaired VOT performance in patients with impaired naming is said to likely represent overall cognitive impairment versus a direct relationship between the two measures. The authors concurred with Paolo et al. (1996) in that they suggested that practically, the naming component of the VOT is simple enough that the small but significant effects of naming on VOT scores will have little clinically important impact on VOT scores.

One study failed to find a significant relationship between confrontational naming and VOT scores (Paul et al., 2001). Paul et al. examined 23 individuals in an archival data set. The data was from participants suffering due to vascular dementia. The study compared VOT scores with performance on the BNT. While there was a significant correlation between VOT performance and scores on the BNT, when stepwise regression analysis was utilized, the BNT did not significantly contribute to VOT scores. The failure of this study to find a significant contribution of naming to VOT performance may be due to the small relationship between the two variables suggested in other studies and the small number of data files examined.

When commonly used neuropsychological measures designed to assess confrontational naming are used, the results appear consistent. Naming is found to explain a small but significant amount of the variance in VOT scores. The percentage of variance ranges between 5 and 11 percent in the studies reviewed. The consensus, however, is that the VOT requires such a low degree of naming ability that the small relationship between VOT scores and naming performance has little clinical importance.
This conclusion is born out by the lack of a significant relationship between the ability to name the reconstructed items of the VOT and naming the VOT scores themselves.

*Visual-spatial Abilities*

Suggestions that the VOT is a measure of visual abilities of some form have certain face validity. It is obvious that the task requires the examinee to view the stimuli and mentally rotate or otherwise make sense of the objects in order to provide an answer. Initial studies of the VOT were thought to suggest that the VOT measured visual organizational ability because in individual cases persons with damage to brain regions thought to correlate with visual organization scored markedly worse on the VOT (Rathbun & Smith, 1982). Several studies noted that the VOT is accepted as a measure of visual perceptual abilities, however, it was not until the 1990’s that studies were presented that actually examined the correlations between VOT scores and other accepted measures of visual abilities.

Seidel (1994) compared VOT scores to scores on the various subtests of the WISC-R and its three derived factors. This study utilized a clinical group of 49 children referred for neuropsychological evaluation. The results showed a significant correlation between the Perceptual Organization Deviation Quotient and the VOT. Non-significant correlations were found between the VOT and both the Verbal Comprehension and Freedom from Distractibility Deviation Quotients. There were also strong correlations between several individual subtests of the WISC-R and the VOT. The VOT was significantly correlated with Block Design, Picture Arrangement, Object Assembly, and Picture Completion. Each correlation was said to account for more than 21 percent of the
variance in VOT scores. Principle Component Analysis showed that the VOT loaded on the same factor as WISC-R Performance IQ and the Developmental Test of Visual-Motor Integration. WISC-R Verbal IQ, the Peabody Picture Vocabulary Test-Revised, Arithmetic, and the Sentence Memory Test loaded on a second factor while spelling and reading tests loaded on a third factor. The author concluded that the VOT was significantly related to “visual-perceptual-constructional skills” (p. 67) and not as highly related to verbal skills.

In their study of 50 VA patients referred for neuropsychological evaluation, Ricker and Axelrod (1995) compared VOT scores to scores on the three factor scores from the WAIS-R. They used hierarchical multiple regression analysis on the VOT scores corrected for age and education. This analysis showed that the three WAIS-R factor scores taken together, Verbal Comprehension, Perceptual Organization, and Freedom from distractibility, accounted for 48 percent of the variance in VOT scores. The Perceptual Organization Factor score was found to be the only statistically significant factor. The authors concluded that the study lends support for the use of the VOT primarily as a measure of perceptual organization ability.

A study similar to that done by Ricker and Axelrod was completed by Paolo et al. (1996). The Paolo et al. study utilized 96 patients consecutively referred to a university medical center. The participants were given a short form of the WAIS-R which included the Block Design and Picture Completion performance tasks. These two subtests were used to generate a modified perceptual organization quotient. When entered first into a stepwise multiple regression equation, these two subtests accounted for 44 percent of the
variance in VOT scores. This finding is consistent with previous studies suggesting that
the VOT is significantly related to perceptual abilities.

The previous studies had begun to look at VOT more closely in terms of what it
actually measures. They began to move away from the assumption that the VOT was a
measure of impairment or generalized brain damage. Johnstone and Wilhelm (1997) give
a clear summary of the development of perspectives concerning the VOT and the
uncertainty regarding the specific construct measured by the VOT. They hypothesized
that if the VOT loaded on the same factor as the performance subtests of the WAIS-R as
in the experiment by Ricker and Axelrod (1995), then it should be considered a test of
global visual-spatial intelligence. The alternative would be that the VOT load on another
patients referred for neuropsychological assessment at a Midwestern hospital. The
patients had various disorders leading to concerns regarding their cognitive functions.
VOT scores were significantly correlated with all tests administered with the exception of
the Digit Span subtest of the WAIS-R. This included all of the other subtests of the
WAIS-R, the Rey Complex Figure Test (Rey), the Category Test, the Boston Naming
Test, the Controlled Oral Word Association Test, and the Wechsler Memory Scale-
Revised Visual Reproduction (WMS-R VR). Principle component factor analysis was
conducted. Three factors emerged with the VOT, WAIS-R Performance IQ subtest, Rey,
Category Test, and the WMS-R VR falling on the first factor. This was said to be the
visual-spatial factor. Language tasks composed the second factor with digit span and the
Controlled Oral Word Association Test making up the last factor. Johnstone and Wilhem
(1997) concluded that the VOT is best considered a measure of global visual-spatial intelligence rather than a measure of a distinct construct such as visual integration.

Based on the studies reported to date, there is no evidence that the VOT measures a distinct construct, but rather it is suggested to measure something similar to that measured by Performance subtests of the WAIS-R. While exact correlations show some variability between studies, the overall conclusion is that the VOT is highly correlated with other measures of visual-spatial and perceptual abilities. The VOT has been found to have smaller correlations with verbal tasks such as those on the WAIS-R or the BNT (Greve et al., 2000), and found to have the highest correlations with Block Design, Object Assembly, Picture Completion, and Picture Arrangement (Greve et al., 2000; Paul et al., 2001).

Validity

*Generalized Impairment*

Because the VOT was originally developed as a test of brain damage, it follows that much of the validity research completed to date has focused on the test’s efficacy in accomplishing this end. The effect of the location of a brain injury on VOT scores and the high correlation between VOT scores and visual-spatial abilities has called the VOT into question as a measure of generalized brain injury. More recent studies have attempted to determine the VOT’s efficacy in assessing visual-spatial deficits and de-emphasized its use as a screener for brain damage.
When the VOT was still "a new test for detecting brain damage," Walker (1956) attempted to validate its use in determining if patients had incurred brain damage. He categorized 38 patients into three groups: no cortical involvement, probable cortical involvement, and definite cortical involvement. The latter two groups were combined due to the small N and labeled as the probable cortical involvement or brain damage group. These groups were all "neuropsychiatric patients" and considered representative of routine admissions to the Neurological Service. The two remaining groups were compared using Chi Square analysis and no significant difference was found using VOT scores.

A second analysis was done using the qualitative scoring measures described by Hooper to Walker (Hooper, 1983) in a personal communication. Walker described the "concrete" responses to VOT items as those that focus on one aspect of the stimulus and fail to integrate the parts. In using this qualitative scoring method by computing the number of concrete responses, it was found that there was a significant difference between the two groups. The control or no cortical involvement group had one or less concrete responses with the suspected cortical involvement group having two or more concrete responses. Age and intelligence, as measured by Shipley Vocabulary scores, were not found to be related to the number of concrete responses. Walker concluded that however effective, the low cutoff point of one concrete response allowable made that the VOT vulnerable to administration and scoring errors. He suggested lengthening the test as a worthwhile revision.

By 1970, Love (1970) reported that few studies of the VOT were published with much of Hooper's work being in unpublished formats. Love used a New Zealand
Psychiatric Hospital population to study the validity of the VOT. He tested 115 patients in two psychiatric hospitals. He compared individuals with brain damage including sclerosis, meningitis, space occupying lesions, tumors, hemorrhage, Parkinson's disease, and Wernicke's, with non-organics including those with schizophrenia, mood disorders, neuroses, behavior disorders, mental retardation, anti-social behavior, and those who were court remanded for treatment. He also had a group of patients with alcoholism and epilepsy as a third comparison group. Using the cutoff score of 20 on the VOT, he obtained a false positive rate of 14.5 percent in his non-organic group when those with IQ's less than 75 were omitted. This is to say that 14.5 percent were falsely identified as brain damaged by their under 20 score on the VOT. A false negative rate of 30 percent was found in the organic group. Thirty percent scored above the cutoff of 20. The author concluded that the VOT would not be helpful in differentiating diagnoses within the organic group but may work well as a screening device as it is quick and easy to administer and yields a "satisfactory" identification rate.

Gerson (1974) had similar results in his testing but came to a different conclusion. Using 68 participants, he compared 16 patients with degenerative, traumatic or toxic etiological organic pathology, 19 patients with acute or chronic schizophrenia, and 33 "normals." The normal participants were those with medical diagnoses only and hospital employees. He found significant differences between the groups in terms of IQ scores with the organic group having a mean IQ of 87.6, the functional group had a mean of 89.9, and the normal group having a mean of 95.9. Above and beyond these IQ differences, he found that there was a significant difference in VOT scores between the normal group and the organic group and between the organic and functional groups.
There was a non-significant difference between the functional group and the normal group.

When considering VOT cutoff scores, Gerson used the scale given in the manual of 25-30 being unimpaired, 20-24.5 having mild impairment, 10-19.5 having moderate impairment, and <10 having severe impairment. Using a cutoff score of 24.5, the entire organic group was correctly identified. Nineteen percent of the functionally impaired group scored above the cutoff score and were considered to be false negatives. This finding addresses the issue of the use of the HVOT to discriminate between neurological impairment and psychiatric impairment. The perspective expressed in Gerson's article is in contrast to the study by Love (1970). Gerson's (1974) hypothesis is that psychiatrically impaired or functionally impaired patients will score below the cutoff of 24.5. In the study by Love (1970) the test was expected to identify non-organic psychiatrically disordered patients as unimpaired. This difference in language between studies points to the disagreement as to whether psychiatric patients are expected to score similar to normals, unimpaired on the VOT, or similar to those with neurological impairment, impaired on the VOT.

In Gerson's study (1974) within the normal group, 51 percent scored below the cutoff of 24 on the VOT. This was seen as an unacceptably high false positive rate, however, when a cutoff score of 20 is used as in the study by Love (1970), the false positive rate goes down to 25 percent. Gerson (1974) stated that the test should not be used as a screen due to the high false positive rate using the cutoff score of 24. However, the manual suggests that those in the 20-24.5 mild impairment range should be identified as those in need of further testing if the VOT is used as a screening measure. Gerson also
noted that the test did not identify functionally disordered patients from the normal group and is not useful in making this determination. This point is considered in the most recent Hooper Manual (1983) suggesting the test may not be useful in identifying functional disorders. Gerson (1974) suggested that the VOT should be used as part of a larger battery to discriminate normal from those with organic impairment and not as a stand alone screening tool due to its pattern of false positives.

Wang (1977) studied laterality differences as well as the validity of the VOT as a screen for brain damage. His study included 49 brain damaged subjects and 17 non-brain damaged subjects. The brain damaged subjects were divided into three groups, right hemisphere damage, left hemisphere damage, and bilateral damage. Control subjects were patients hospitalized for low back pain, osteoarthritis, peripheral neuropathy, or spinal cord injury. The groups were found to be statistically similar in terms of education and the interval between the onset of their brain damage and testing. The brain damaged group was significantly younger than the control group. Wang used a Kruska-Wallis one-way ANOVA and Chi Square analysis procedures. He found that the control group scored significantly better on the VOT than did the brain damaged groups. Using the cutoff score of 20, he obtained a 17 percent false negative rate and an overall correct diagnosis rate of 83 percent for the brain damaged groups. The false positive rate for the normal control group was 40 percent, however. The author noted that when the lowest score from the control group, 14, was used as a cut off point, a 57 percent correct diagnosis rate was achieved in the brain damaged group. Using 24, the highest score of the brain damaged group, as a cutoff score yielded no false negatives and a 40 percent correct diagnosis rate in the control subjects. He concluded that depending on the purpose
of the test and it use, cutoff scores of 14 or 24 could be used. The lower score could be
used to minimize false positives and the higher score could be used to minimize false
negatives.

One validity investigation sparked a debate in the literature that can be seen as the
beginning of a change in how the VOT was perceived. Boyd's (1981) study of the VOT's
use as a screen for brain damage may have led to the decline in its use as such and
increased interest in its functionality as a measure of visual-spatial abilities. To this point,
investigators had been attempting to validate the use of the VOT as a screening
instrument for brain injury and there was little discussion of the use of VOT to assess a
particular aspect of cognition such as visual-spatial ability.

Boyd (1981) reported his study of 80 patients (all but two were inpatients) at a
university hospital. He put together a sample of patients under age 55 and with more mild
forms of neurological impairment. The brain damaged group consisted of 40 patients
with closed head injury, seizure disorder, vascular malformations, normal pressure
hydrocephalus, multiple sclerosis, penetrating injuries, and cerebral vascular accidents.
The control group consisted of 80 percent functionally impaired patients with depression,
Bipolar Disorder, neurological disorders without brain involvement (exact diagnoses
were not given), personality disorder, medical problems, and adjustment problems. This
mix of patients in the control group was used to control for the mood disturbance and
anxiety that can be associated with hospitalization for any reason. The two groups were
found to be quite equal in terms of age, education, and IQ (as estimated by picture
vocabulary scores).
The mean VOT scores for Boyd's two groups were higher than those reported in previous studies suggesting that his sample did include less impaired patients. The mean VOT score for the brain damaged group was 23.2 and the mean for the control group was 26.5. Using a cutoff score of 25, marking the mild impairment range suggested in the manual, yielded an overall correct classification rate of 73.8 percent. This cutoff score correctly classified 67.5 percent of the brain injured group and 80 percent of the control group. Boyd concluded that in this population of more mildly impaired individuals, a higher cutoff score should be used. He noted that more massively impaired patients are often used in validity studies but their neurological status is likely to be obvious from their presentation. He also noted that had a cutoff score of 20 on the VOT been used, only 15 percent of the brain injured patients would have been correctly classified. He again suggested that the specific purpose of the administration be considered when choosing a cutoff score and that age, education, and intelligence should be factored in when interpreting test results.

One response to Boyd's research was by Woodward (1982). She criticized that testing the validity of the VOT should be accomplished by using consecutive referrals for evaluation and not by matching equal numbers of subjects. She stated that by matching subjects, the base rate for the acute care setting is artificially set at 50 percent. When the true base rate for a population is much lower, she stated that the diagnostic utility of the test will also be much lower. She also argued that individuals with psychotic processes or thought disorders should not be excluded from the analysis because these symptoms are often confused with brain damage. She suggested that the VOT has a high false negative
rate. Woodward refuted Boyd's (1981) conclusion that the VOT was a good screen for brain damage.

The concern regarding the false negative rate on the VOT was supported by Wetzel and Murphy (1991), who found that 72 cases out of their overall sample of 253 were classified as unimpaired on the VOT while classified as impaired on other neuropsychological measures. It should be noted, however, that in this study by Wetzel and Murphy, the cases were from mixed etiologies from multi-infarct dementia to psychiatric diagnoses. While all but four scored as impaired on other neuropsychological, medical, and imaging tests, it can be argued that the VOT should not be expected to diagnose the cases of depression, psychiatric disorder, and chronic alcohol dependence as impaired.

In his response to Woodward, Boyd (1982b) pointed out that every study attempts to control for threats to validity and variation that is not caused by changes in the independent variable. He described his study as having modest aims saying that not all studies can be consecutive series designs and he described the flaws of these types of designs. In the end, while Woodward’s criticisms are valid, the point made by Boyd was that every study makes compromises in either control or ability to generalize from the results and none are perfect in all respects.

A second criticism of Boyd’s (1981) article came from Rathbun and Smith (1982). It is this critical evaluation and the change in field of neuropsychology that began the shift in how the VOT was viewed and investigated. Rathbun and Smith pointed out that the concept of brain damage as a unitary concept was beginning to lose momentum. Specifically, all injuries or illnesses affecting the brain are not equal. Different locations
within the brain were being found to be involved in different cognitive functions. In relation to the VOT, Rathbun and Smith described it as a test of visual organizational functions. They suggested that it is a test of a specific deficit rather than a general deficit. Furthermore, they noted that some case studies have indicated that specific visual-spatial functions necessary for completion of the VOT may be related to the right posterior region in the brain. For this reason, they suspected that the utility of the VOT reported in validity studies will vary according to the proportion of the cases with damage to this right posterior area of the brain. They criticized Boyd (1981) and other researchers who only reported types of disorders and placed them into an overall brain damaged group, failing to account for the specific location of damage within the brain.

In his response to this criticism, Boyd (1982a) argued that the evidence that the VOT measured some specific deficit was not clear at that time. As reported in this review of the literature, more recent studies have supported the idea that the VOT correlates with other tests thought to measure visual-spatial functions. Additionally, while both Boyd's (1981) initial study and that by Wang (1977) did not show a correlation between VOT scores and either right or left hemisphere damage, more recent studies have suggested a relationship with the right posterior region of the brain (Fitz et al., 1992; Lewis et al., 1997; Nadler et al., 1996).

In the ten years that followed the 1981 article by Boyd, few studies can be found that addressed the validity of the VOT. McCaffrey et al. (1988) compared the VOT with two other measures in an inpatient substance abuse setting and found that the VOT was neither correlated with the Symbol Digit Modalities Test or the Trail Making Tests nor were VOT scores impaired in this population. Tamkin and Kunce (1985) found that the
VOT was able to predict brain damage in 66 adult veteran psychiatric patients at a rate significantly better than chance. Neither of these studies examined the validity of the VOT specifically, but rather made comparisons between different neuropsychological test measures.

Rathbun and Smith (1982) made an argument against using the VOT as a screen for generalized impairment based on their evidence that the VOT measures some specific cognitive ability that can be associated with specific areas of the brain that are not necessarily damaged in every injury or with every illness. As brain damage is not a unitary concept necessarily affecting all areas of the brain, the VOT or other tests of specific abilities would not make good measures of generalized neurological impairment. This perspective slowly began to permeate validity studies conducted using the VOT.

One of the first studies to examine the VOT acknowledging its use to indicate both generalized and specific impairments was that done by Seidel (1994). As previously described, this study examined the VOT’s validity for use with children and found a significant correlation between perceptual organization abilities as measured on the WISA-R and VOT scores. Seidel also found that the clinical and control groups were significantly different when VOT scores were compared. The author noted, however, that the latter finding should not be interpreted as evidence that the VOT should be used as a screen for brain damage. Seidel suggested that given the association of the VOT with non-verbal abilities, that studies with larger sample sizes are needed to examine VOT performance in more homogeneous groups. This would allow for closer scrutiny of VOT performance in individuals with deficits that are limited to discrete domains of
functioning. The implication is that the VOT may not be a good indicator of brain damage if impairment is isolated to abilities other than visual-spatial skills.

Despite the concerns that the VOT may not be a good measure of neurological impairment given its high correlation with non-verbal skills, research continues on its applicability as a general screening measure. Nabors et al. (1997) examined the VOT’s use in an urban medical inpatient setting. They found that the mean for cognitively intact older adults was two points lower for their sample. They suggested that in this population, specific norms should be used to account for this difference. They also found that a cutoff score of 15 provided the best classification rate of 81 percent. Linear discriminant function analysis was performed and showed that by using age, education, and VOT scores, 78 percent of subjects were correctly classified. They concluded that using an appropriate cutoff score can make the VOT reliable for identifying cognitively impaired individuals, but they add the caveat that this applies to individuals with visual-spatial-perceptual difficulties. They refrained from implying that the VOT is a screen for generalized impairment.

Like Nabors et al., Lewis et al. (1997) examined the use of the VOT in a specific population. They examined an African American sample with acute unilateral cerebral lesions. They used a software program to measure and divide the computer axial tomography images of the subjects’ brains into four quadrants, left anterior, right anterior, left posterior, and right posterior. They had a sample size of 153 cases. Thirty seven cases had brain damage to the left hemisphere. Forty one had right hemisphere damage. The control group was composed of inpatient and outpatient cases without evidence of brain injury. Brain injury was defined as a focal neurological deficit that was confined to one of
the quadrants of interest. Subjects were administered the Michigan Neuropsychological battery along with the VOT and other neuropsychological measures. They found that the brain injured cases scored significantly worse on the VOT than the controls. Cases with right anterior lesions also scored lower on the VOT than did the controls and those cases with lesions in any of the other three quadrants. In contrast, none of the cases with lesions in the other three quadrants scored significantly lower than each other or the control group. This suggested that the VOT was not likely a reliable measure of general impairment in cases with focal deficits.

Walsh et al. (1997) performed a study of the validity of the VOT in discriminating between unimpaired, mildly impaired, and moderately to severely impaired older adults. It was found that the mean scores of the three groups on the VOT were significantly different from each other. A cutoff score of 17 correct on the VOT was used and achieved 68.75 percent sensitivity and 71.88 percent specificity when the intact group was compared to the other two groups. No cutoff score was found that would differentiate between the two impaired groups. They noted significant overlap in scores between the intact group and the mildly impaired group. They concluded that the VOT does discriminate between impaired and unimpaired individuals, but may more properly be used to measure specific skills of executive functions and visual-perceptual integration.

The conclusion reached by Walsh et al. (1997) is consistent with the research to date on the validity of the VOT. It has some utility in distinguishing impaired from non-impaired individuals but no cut off score seems to provide adequate classification rates for it to be used alone to determine the existence of brain damage. The VOT appears
better used to measure visual-spatial abilities or as part of a slightly larger battery when screening for brain injury.

_Lesion Location and Laterality_

As early as 1977, the VOT was evaluated for use in determining the location of impairment within the two hemispheres of the brain (Wang, 1977). Wang's study had a relatively small number of subjects with 15 having “predominantly” left hemispheric dysfunction, 19 with right hemispheric dysfunction, and 15 with bilateral involvement. No significant differences between the three groups were found, however, the group with predominantly right hemispheric damage scored lower than those with left hemispheric damage. It was concluded that both hemispheres were involved with VOT performance, but a caveat was added that with a larger sample, a significant difference between left and right hemispheric dysfunction groups could be evident.

Similar findings were reported in the study by Boyd (1981). He too had a small sample size with 18 patients with right hemispheric impairment, 9 with left hemispheric impairment, and 13 with medial or diffuse damage. He failed to find significant differences between the group scores on the VOT. This finding was criticized for not differentiating between right frontal and right posterior lesion locations as it was suspected that the intralobe location of the lesions would affect VOT performance (Rathbun & Smith, 1982). Preliminary data was said to suggest a relationship between damage to the right posterior portion of the brain and performance on visual-spatial tests. This would suggest that damage to this region would also affect VOT performance. Failure to account for the specific lobe of the brain that was impaired within each
hemisphere was postulated to account for the failure to find evidence of laterality or region specificity related to lower VOT scores. In reviewing the findings of both Wang (1977) and Boyd (1981), however, Wetzel and Murphy (1991) concluded that the VOT was not effective in determining laterality of impairment. This conclusion was consistent with their own findings.

A study by Fitz et al. (1992) was the first to begin to answer the questions raised by Rathbun and Smith (1982) regarding the VOT’s validity in measuring lobe specific versus generalized neurological impairment. Fitz et al. (1992) attempted to examine the relationship between VOT scores and specific lesion sites in the brain. It was hypothesized that lesions in the right parietal lobe would lead to the lowest scores on the VOT. The authors chose 41 archival data files from an acute rehabilitation center that met the criteria of having an IQ of 80 or above, a Mini-Mental Status score of 18 or higher, a Randt Memory Test score greater than 10, and no history of psychiatric disorders. The participants were placed in one of three groups: left hemisphere lesion, right hemisphere lesion, and bilateral lesion. Only subjects with discrete lesions were included in the study. There were 13 subjects with left hemispheric lesions, 24 with right hemispheric lesions, and 4 with bilateral lesions. The groups were also divided into those having damage to the right parietal region (n of 11) and those without right parietal damage (n of 30). The data was examined using hierarchical analysis of covariance with age, education, and IQ as control variables. Preliminary tests showed that there were no significant differences between the groups in terms of these three variables; however, previous research had suggested they were related to VOT performance. There were no significant differences found between left, right, and bilateral hemisphere groups. The investigators did find that
those with right parietal damage scored significantly lower than those without right parietal damage when age and education were controlled and when age, education, and IQ were controlled. Interestingly, when only age and IQ were controlled, the difference between the groups was not significant. This may suggest that the variability in scores on the VOT explained by participants' level of education is significant enough to mask the difference between those with and without right parietal lesions.

Another analysis of laterality in VOT scores was completed by Nadler et al. (1996). Unlike the study by Fritz et al. (1992), the subjects in the Nadler et al. (1996) sample were homogeneous in the type of lesion present. All of the subjects had Cerebral Vascular Accidents (CVA) that were confined to one hemisphere. The sample included 44 subjects with right hemisphere CVA and 23 with left CVA. In addition to quantitatively examining VOT scores, the authors also qualitatively examined the types of errors made by subjects in each group. A revised qualitative scoring system was used for the study. Briefly summarized, part errors were defined as those that name one part of the card or naming the correct object but labeling it as "broken." Perseverative errors were those that were repetition of previous responses or those that were clearly perseverations in terms of categories of responses. The example given is answering with different tool names after the hammer item is identified. Language based errors were those that involved typical anomic or aphasic errors of naming a closely related item, giving a partial word response, giving a description of the item, or giving a grammatically incorrect response. Part/Language responses combined Part and Language errors in one response. Unformed or Unassociated response were those that did not relate at all to the item. A last category was for "Don't know" answers or when no response was given.
While the two groups did not differ in terms of age and education, significant differences were observed between left and right hemisphere groups. The right CVA group was found to score significantly poorer than the left CVA group on the VOT. Due to the non-parametric nature of the qualitative scoring data, Mann-Whitney tests were used in this analysis. It was found that the right CVA group made more Part and Unformed responses while the left CVA group made more language based errors. These results were said to be consistent with knowledge of brain functions in the left and right hemispheres.

The most recently reported study of the VOT in relationship to lesion laterality and site was completed by Lewis et al. (1997). As previously described, the researchers examined data from African American patients with lesions confined to one of four quadrants of the brain. They found that those with damage to the right anterior region scored significantly worse than those with damage in any other quadrant while scores from subjects falling in the other three quadrants did not significantly differ from each other. These findings added to the growing evidence of a relationship between VOT scores and damage to the right hemisphere and challenged earlier reports that VOT scores were related to right posterior or parietal damage. One of the strengths of this study is the equality of the number of subjects in each group with 37 subjects having left hemisphere damage and 41 having right hemisphere damage. The quadrant groups were also roughly equal with 10 patients with left anterior damage, 7 with left posterior damage, 11 with right anterior damage, and 9 with right posterior damage. Of all the studies examining laterality and lesion site, this included the largest sample size.
It appears that a relationship between right hemisphere damage and lower VOT scores is emerging in the literature as larger sample sizes are utilized. This is consistent with the understanding that the VOT measures visual-spatial skills that are associated with the right hemisphere. It remains unclear as to which portion of the right hemisphere may play the largest role in VOT performance, however, as studies alternately suggest right anterior, parietal, and posterior areas to be most involved. More research is needed utilizing well defined lesion location information and large sample sizes.

**Neurological versus Psychiatric Impairment**

The VOT manual lays out most of the studies that have examined the utility of the VOT in differentiating between normal individuals and those with psychiatric and neurological impairment (Hooper, 1983). Two initial studies by Hooper that are reported in the manual examined the use of the VOT in discriminating between psychiatric and neurological impaired individuals. The first study suggested that subjects with schizophrenia scored similarly to normals and those with neurotic disorders. These three groups scored significantly better than those with brain injury. The second study suggested that three groups scored in a tiered fashion with individuals with neurotic diagnoses and personality disorders scoring best, individuals with schizophrenia scoring slightly lower, and those with neurological impairment scoring the lowest. Combining the first two groups, the neurologically impaired group was found to score significantly lower than those with psychiatric impairment. This finding was thought to suggest the usefulness of the VOT in differentiating between patients on a hospital psychiatric ward with mental illness and those with brain injury or other neurological impairment.
One of the challenges in examining the studies that have used psychiatric populations is that they do not always clearly report the specific diagnoses that are included in the psychiatric groups. Gerson (1974) was the only study that compared neurologically impaired individuals with those with schizophrenia. All of the other studies involving psychiatric groups used individuals with mixed diagnoses from schizophrenia to mood disorders. Some studies did not explain the diagnoses of the individual's in their psychiatric group. This is important as Hooper's research suggested that individual's with schizophrenia scored differently than those with other diagnoses. With this caveat in mind, the following studies have reported the use of the VOT to differentiate between psychiatrically impaired and neurologically impaired groups.

Walker (1956) used data from 38 files from the neurology service to examine the use of the VOT in distinguishing between neurological and psychiatric impairments. He compared two groups, one that was not suspected of cortical involvement (neurological impairment) and a group in which neurological impairment was suspected. Walker sampled from “neuropsychiatric” patients and labeled those thought to have only psychiatric disorders as “no cortical involvement.” Specific diagnoses for the individuals in this group were not reported. The diagnoses for 90 percent of the individuals had been made based on all available information at the time of discharge with the exception of the psychological report. The remaining 10 percent were treatment cases with well established diagnoses. Comparisons between the two groups failed to find a significant difference using the quantitative scoring methods. At that time there was no formal qualitative scoring method but the author noted that Hooper suggested that qualitative evaluation of concrete responses, or those which referred to only a part of the VOT.
stimulus, could differentiate between the two groups. By examining the data in this qualitative manner, the author found that more than one concrete response suggested "cortical involvement" or brain injury.

The study by Love (1970) described earlier did find that the quantitative scoring of the VOT was useful in differentiating between psychiatric and neurologically impaired groups. The psychiatric group in this case included 15 individuals with schizophrenia, 13 with mania or depression, 15 with neuroses, 10 with behavior disorders, 6 with mental retardation, 13 with antisocial behavior, and 14 who were court remands. With a cutoff score of 20 a 14.5 percent false positive rate was seen in the psychiatric group and a 30 percent false negative rate was seen in the neurologically impaired group.

These previous studies did not examine whether the VOT could be used to distinguish between the three classes of cases, normal individuals, psychiatric patients and neurologically impaired patients. Gerson (1974) did examine this question using a homogeneous psychiatric group of individuals diagnosed with schizophrenia. It was found that VOT scores were significantly different between the normal group and the neurologically impaired group and between the psychiatrically impaired group and the neurologically impaired group. No significant difference was found between the normal group and the psychiatrically impaired group. The qualitative scoring categories were not found to provide further information given that they were rare in the organic group with no neologistic or bizarre responses noted in any group. This study suggested that psychiatrically impaired individuals scored more like normal individuals and the two groups could not be differentiated based on VOT scores. Findings indicated that the VOT
might be useful in identifying neurologically impaired individuals but neither quantitative nor qualitative scoring in this study helped identify those with psychiatric impairment.

Research to date is mixed on whether the VOT is useful in differentiating between groups with neurological impairments and those with psychiatric impairment or no impairment. While most would agree that the VOT is expected to differentiate between normal individuals and those with neurological impairment, some studies suggest that those with psychiatric impairment score more like normals while others suggest that they score as more impaired on the VOT. While the move away from the use of the VOT as a screen for unspecified neurological impairment suggests that one measure alone will not likely be used as a definitive test for brain injury, it remains clinically useful to have a relatively short measure that might suggest if a patient's clinical presentation is more likely due to a neurological impairment than a psychiatric disorder. To this end, the study by Gerson (1974) suggested that psychiatrically impaired persons (with schizophrenia) tended to score more like normals than those with neurological impairment. If there is no reason to suspect that psychiatric impairment would affect visual-spatial abilities, it remains possible that the VOT could provide evidence of the existence of neurological impairment. Caution should be taken, however, when trying to distinguish psychiatric impairment from neurological impairment as the VOT manual references initial studies that showed that psychiatric impairment was not a unitary concept and that individuals with schizophrenia scored lower than those with "neurotic disorders." The specific diagnoses included in a psychiatric sample are therefore important. A combined use of quantitative and qualitative scoring may be best given earlier findings, though evidence for the usefulness of qualitative scoring is mixed.
Item Analysis

No empirical data is provided in the VOT manual regarding the justification of the item order or the inclusion of certain items instead of other possible items (Hooper, 1983). Merten and Beal (2000) noted that:

the manual does not include an item analysis … the item order appears to have been established a priori, without empirical testing. Likewise, the manual does not indicate an empirical basis for which answers are to be scored as correct and which are to be given half credit (p. 522).

Verma et al. (1993) completed an item analysis of the VOT in Indian subjects to determine the test characteristics within that population. To date, these are the only two studies to examine the item characteristics of the VOT.

Merten and Beal (Merten & Beal, 2000) analyzed the VOT in German-speaking subjects in an attempt to obtain data on the psychometric characteristics of the test and suggested a revision of the item order. Merten and Beal analyzed previously collected data from 320 neurological patients. The group was of mixed pathology including traumatic brain injury, CVA, neoplasm, and a variety of other neurological impairments. Patients ranged in impairment from those said to have nearly or completely recovered to those with severe impairment or dementia. Analysis was completed using a credit/no credit system where half-credit items were treated as incorrect. They computed an F ratio for each item based on full credit responses versus wrong responses.

Based on their analysis, they made several suggested revisions and comments on the test. They determined that items 2, 3, and 18 are too easy though they admit that these
items may be useful as a warm-up phase. Item number 1 was said to be too difficult given that the instructions for the test were considered to be too brief. Subjects often gave two responses spontaneously, obviously not comprehending the task requirement of assembling the items into one object. Item number 6 was said to be too difficult with relatively less impaired individuals having difficulty with the item. Item number seven has two correct responses, sheep or dog. They found that the half-credit response of polar bear or bear was given just as often as sheep, but less often than dog suggesting that sheep is too weak of a response to be given full credit. Item number 11 is an apple with half credit given for fruit. The researchers found that fruit was an uncommon answer and should be treated as incorrect. Item number 12 was said to have too little discriminative power as incorrect answers were often given by those who had good overall scores. On item 14, snorkel was found to be as good a response as hockey stick, which is given full credit. Items 16 and 19 are a kettle and a teapot respectively. The researchers suggested that since the answers for the two items are often interchanged that it may be necessary to score either answer correct given trends in style and language use. Item 17 is a chair but the half credit response of sofa is given by nearly as many people and those who responded with sofa had higher overall scores. It was suggested that the two responses should be considered equivalent. The researchers suggested revision of item 21. Two incorrect answers were often given by those with high scores on other measures as well as high overall VOT scores. The incorrect response of iron to the shoe presented in item 27 was given by a relatively high percentage of respondents who also have high overall VOT scores. Finally, on item 30, paint brush was given nearly as often as the correct
answer of broom. Additionally, the answer of tube was given by those with higher than
average VOT scores.

Merten and Beal concluded that while the VOT is an often used test, it could
benefit from revision. They suggested that the item order is not consistent with the level
of difficulty of the items. The scoring rules were found to lack empirical base. The
authors suggested four major areas of revision. The introduction should be made clearer
with a graphic example demonstrating the test principle. The item order should be
changed to reflect the level of difficulty. This would allow for the implementation of a
discontinue rule suggested by Wetzel and Murphy (1991). Merten and Beal (2000) also
recommended omitting half credit responses from the scoring of the VOT. Lastly, they
recommended clear instructions on how the examiner may assist anomic patients so that
differentiation can be made between those with language problems and those who score
poorly on the VOT due to visuospatial difficulties. A more comprehensive revision could
examine the possibility of deleting items with unsatisfactory characteristics such as being
too difficult, too easy, or unclear, and considering new items that have been empirically
validated. Any differences based on the language, given that this study was completed
with a German-speaking sample, were thought to be minimal due to the visual nature of
the task. The researchers suggested a more profound generational effect due to the older
styles represented in the items such as a 1940’s era truck. At least minor revision of the
test was suggested to be worthwhile and accomplishable with little effort.

As noted above, Verma et al. (1993) studied the VOT in a sample of Indian
subjects. They used a large sample including 133 normals and 152 psychiatric patients.
The normal group consisted of family and friends of patients. They divided both groups
into three subgroups. The three subgroups were the high performers, midrange performers, and low performers on the VOT. Correlation of overall test performance with individual test items was said to progress satisfactorily across the groups from low to high range performers. Additionally, the results showed that most items differentiated well between the normal and psychiatric groups using Chi Square analyses. The analysis did suggest that the items were not in order of difficulty. As with the study by Merten and Beal (2000), they also found that the first few items are too easy and the last two are too difficult but this was said to be desirable. This study was not an evaluation of the VOT per se as the intent was to evaluate its use in the Indian population. They concluded that the VOT was suitable for use in the Indian population and made no recommendations regarding revision of the test.

Both studies presented in this section suggest that the items on the VOT are not in their correct order of difficulty based on the samples obtained. This evidence has significant relevance to the use of a discontinue rule suggested by Wetzel and Murphy (1991). Obviously, if relatively easier items occur late in the test, the examiner can have little faith that discontinuation of testing early will not impact the subject’s score. Despite the concern about the item order, in their study of 253 cases, Wetzel and Murphy found that scoring the VOT using items answered prior to five consecutive wrong answers made little difference in the category of impairment suggested. They used the standard categories of interpretation suggested in the manual ranging from low probability to very high probability of impairment. The change in raw scores that occurred using this discontinue criterion ranged from 0 to 5 points with a mean of .32. In only four cases did the score change enough to change the range of impairment when using the standard
classification system. Using the discontinue rule changed the classification of the four protocols from the moderate impairment to the severely impaired classifications. The authors suggested that the use of a discontinue rule is justified given these findings and the potential to save time and limit examinee frustration.

The criticisms raised by Merten and Beal (2000) suggest that a revision of the VOT could easily improve its psychometric properties. Clarification of instructions and modifying the standards for which answers receive full or half credit would lead to increased standardization. Changing the item order would allow for the addition of a discontinue rule and lower average administration times. The VOT is due for revision as it has changed very little since its development in the 1950’s.

Summary

It has been twenty years since the last revision of the VOT and new findings have changed how the VOT is used. Any future revision of the VOT test and manual will benefit from contemporary research data regarding factors that influence VOT scores, validity data, and item analysis data. Research on the VOT has provided valuable information in these areas.

Several variables have been demonstrated to affect VOT scores. Age, intelligence, naming ability, and visual-spatial abilities have all been shown to correlate with VOT scores. Age and VOT scores have been found to be positively correlated from childhood to early adulthood. After age 30, the relationship appears to reverse, with a negative correlation evident in the research data. VOT scores are positively related to intelligence. There appears to be a small positive correlation between naming ability and VOT scores.
VOT scores have also been shown to positively correlate with visual-spatial abilities. These variables need to be considered when interpreting or researching the VOT.

Education and gender do not appear to have a significant impact on VOT scores. Some studies found that educational level was correlated with VOT scores in some instances, however, when variables such as intelligence, age, and degree of impairment were controlled, the relationship between education and VOT scores was not significant. It may be important to further clarify the effect of education on VOT scores in future research. This study will add to what is known about the effects of demographic variables on the ability of the VOT to distinguish between different groups.

Some research has shown the utility of the VOT in screening for generalized impairment, but more recent studies suggest that the VOT should be used as a measure of visual-spatial abilities. The test has been shown to be less reliable in determining brain impairment when lesions are isolated in discrete portions of the brain that are not thought to be involved in visual-spatial processing. VOT scores appear to be associated with right hemisphere functions, though data conflicts regarding what portions of the right hemisphere are most involved in VOT performance. The association of visual-spatial skills with certain portions of the brain may contribute to the false negative rate in the 25 to 30 percent range which is typically reported when the test for generalized impairment. (If VOT performance is associated with a discrete area of the brain, then someone with a generalized impairment may not have significant injury to all parts of the brain equally and the portion of the brain associated with VOT scores may remain relatively intact. This study investigates the utility of VOT scores in distinguishing between those with generalized impairment, psychiatric impairment, and no neurological or psychiatric
impairment. Additionally, this study will examine the relationship between VOT scores and other measures of visual-spatial abilities.

The VOT manual (Hooper, 1983) suggests that the VOT can be used in an attempt to differentiate between three groups, those individuals without schizophrenia or neurological impairment, those with schizophrenia, and those with neurological impairment. Early studies suggested that these three groups scored in a three tiered fashion with the first group scoring highest, the group of individuals with schizophrenia scoring lower, and the neurologically impaired group scoring the lowest. More recent studies suggest that samples of individuals with schizophrenia or other psychiatric impairment score similar to those without impairment and significantly better than those with neurological impairment. It is unclear if the VOT has utility in differentiating between individuals with schizophrenia and other psychiatric impairments, neurologically impaired individuals, and those without impairment. This study examines the qualitative and quantitative scoring characteristics of normal, psychiatrically impaired, and neurologically impaired groups to help determine the utility of the VOT in differentiating between persons in these three populations in a clinical setting.

Any revision of the VOT will need to consider the literature regarding the item analysis of the test. The two reported item analyses of the VOT suggest important revisions to the item order and possible additions and subtractions of certain items. Additionally, some scoring procedures have been questioned when the item characteristics are considered. The two studies conducted have been completed on non-English speaking populations and did not include data from all three aforementioned groups leaving a gap in the literature that will be addressed in this study.
This study provides information from a large sample to address some of the aforementioned questions that remain regarding the appropriate use of the VOT.
CHAPTER 3

METHOD

Overview

This study consisted of an archival review of files from the practice of a neuropsychologist whose practice consists of both inpatient and outpatient neuropsychological assessment. The intent of the file review was to identify files from individuals who had been assessed using the VOT and other neuropsychological measures as part of routine evaluations in a neuropsychological practice. The archival files were reviewed according to the procedures outlined in this chapter to identify individual files that could be categorized into three distinct groups, those without psychiatric or neurological impairment (normal), those with psychiatric impairment, and those with neurological impairment. The clinical files were from the neuropsychologist’s practice in an inpatient psychiatric unit and an outpatient neuropsychology service. All of the evaluations consisted of assessment with a variety of neuropsychological measures as part of the Michigan Neuropsychological Battery (Smith & Berker, n.d.).

The first group consisted of adult subjects aged 18 and older without history of neurological impairment, stroke, heart attack, or seizure disorder. Individuals in this normal group had completed the neuropsychological testing measures and had no history of psychiatric or neurological impairment. These individuals were primarily volunteers who had been tested as a part of graduate student training in neuropsychological assessment.
The second group consisted of neurologically impaired individuals age 18 and older who suffered from any of the number of neurological impairments resulting from Alzheimer's disease, stroke, hydrocephalus, a motor vehicle accident, neoplasm, seizure disorder, or other diagnosed brain disorders.

The third group consisted of files from adults 18 years of age and above which showed evidence of established psychiatric diagnoses and/or significant psychiatric symptomatology, that was recorded in the history section of the neuropsychological testing report completed during their inpatient psychiatric hospitalization, and who had no history of neurological impairment. The archival file review procedures are outlined in the following sections.

Archival File Review

Individual files included data from a variety of neuropsychological test measures as part of the Michigan Neuropsychological Test Battery. This research project was conducted as archival research in accordance with HIPAA guidelines. The activities preparatory to research, such as the removal of Protected Health Information (PHI), were performed in accordance with HIPAA guidelines as well. The data collection took place at the record storage facility. A contract was signed between the neuropsychologist and the investigator, ensuring that the investigator maintained confidentiality of the data and followed HIPAA guidelines regarding PHI. Approximately 600 files were reviewed to identify cases which met the inclusion criteria and were distributed between the three comparison groups. When appropriate case files were identified, a random code number was assigned and basic demographic and test data was anonymously recorded on a
separate data sheet. None of the 18 elements that could conceivably identify an individual as defined by HIPAA were recorded as part of the data recording process to render the data set de-identified. These 18 items specifically not recorded included: names, all geographic subdivisions smaller than a state, all elements of dates except the year, telephone numbers, facsimile numbers, electronic mail addresses, Social Security numbers, medical record numbers, health plan beneficiary numbers, account numbers, certificate/license numbers, vehicle identifiers/license plate numbers, device identifiers/serial numbers, web universal resource locators, internet protocol address numbers, biometric identifiers (finger or voice prints), full-face photographic images, and any other unique identifying number, characteristic, or code. Each individual data set was assigned a random code number.

Data items that were coded included: age, gender, level of education, co-morbid medical and psychiatric diagnoses, and the location, nature, and type of neurological impairment. Scores on the VOT, the Symbol Digit Modalities Test (SDMT), and Raven Matrices short form were recorded. In addition, Verbal, Performance, and Full Scale IQ scores and the Subtest scores of the WAIS-R were recorded. Since the data set was archival and since imaging studies and other methods of identifying the location of lesions were not available for each archival file, the degree to which an injury can be localized in each case was limited. Neurological impairment was coded as either general or focal. Focal lesions were coded minimally by hemisphere left versus right. For files that provided more specific localization information, the quadrants were coded with right anterior, right posterior, left anterior, and left posterior being coded 1 to 4 respectively. Those files for which information was available on the specific lobe of the brain that was
injured were coded for right and left frontal, temporal, parietal, and occipital yielding eight coding categories. If damage encompassed more than one category at this level, such as damage to both the right frontal and right temporal areas, then the file was coded for both categories in addition to a broader category based on the larger anatomical structure which contains the damaged region such as right anterior and right hemisphere. In actuality, most of the cases did not include evidence of focal damage, but rather, had etiologies such as a motor vehicle accident suggesting generalized injury. The specific locations of damage were therefore not used in the analysis. The exact, verbatim responses on the VOT were recorded for the purpose of performing qualitative analysis of error types. The verbatim responses were later reviewed and coded by the primary investigator according to instructions/description in the VOT manual. No reliability data was obtained on the investigator's qualitative coding of the item errors.

Only those archival client files which included the demographic information necessary for coding such as age, gender, and level of education, as well as administrations of the VOT and the Subtests of the WAIS-R were included. Some of the files reviewed did not include the administration of the Raven Matrices Short Form or the SDMT. Analyses regarding these measures were limited to a smaller subset of the overall sample.

Participants

Archival data files were analyzed from individuals who had been seen in the course of a neuropsychology practice that included inpatient and outpatient evaluations as well as evaluations of normal adults tested as a part of graduate student training in
neuropsychological assessment. The file review yielded 338 files that met the inclusion criterion. The age range for all three groups described below was from 18 to 89 years. The mean age was 38.06 (SD=15.58) years. The gender distribution was 198 (58.6 percent) male and 140 (41.4 percent) female. The mean level of education was 12.93 years (SD = 3.03). The distribution of right and left hand dominance was 309 (91.4 percent) right hand dominant and 27 (8 percent) left hand dominant with 2 (.6 percent) unknown.

The file review yielded 146 files from normal adults that fit the inclusion criteria. The files came from individuals with no known psychiatric or neurological impairment. The individuals' files showed no evidence of psychiatric diagnoses at the time of testing, neurological impairment, stroke, or seizure disorder. One person in this group had a history of depression and one had undergone triple bypass heart surgery. The gender distribution for this group was 75 males (51.4 percent) and 71 females (48.6 percent). The mean level of education for this group was 14.32 years (SD= 3.02). The average age was 40.40 years (SD= 17.04). There were 135 individuals (92.5 percent) reported to be right hand dominant and 11 individuals (7.5 percent) reported to be left hand dominant.

The second group included 92 files with evidence of established psychiatric diagnoses and/or significant psychiatric symptomatology and no history of neurological impairment. These files were from individuals tested while in an inpatient psychiatric ward. No files that indicated a history of closed head injury with post-traumatic amnesia greater than thirty minutes were included in this group. Only six files in this group (6.6%) indicated a history of closed head injury. No history of neurological impairment for these individuals was recorded in the files. Many of the individuals in this group had more than
one Axis I diagnosis. The percentages listed are based on primary diagnoses and the number of participant files in the group. Diagnoses, as reported in the files, included: Adjustment Disorder with depression 1.1%, Adjustment Disorder with mixed disturbance of emotion and conduct 1.1%, Alcohol abuse 2.2%, Anxiety 1.1%, Attention Deficit Hyperactivity Disorder 1.1%, Bipolar Disorder 3.3%, Delusional Disorder 4.4%, Depression 3.3%, Dysthymia 1.1%, Developmental Disorder 2.2%, Major Depressive Disorder 7.7%, Major Depressive Disorder with psychoses 1.1%, Obsessive Compulsive disorder 1.1%, Polysubstance abuse 6.5%, Psychosis 10.9%, Schizoaffective Disorder 5.5%, and Schizophrenia 17.6%. Sixteen individuals had diagnosed personality disorders. In 12 cases (13%), the primary diagnosis was a personality disorder. Diagnosed personality disorders included Borderline, Antisocial, and Organic personality disorders. In sixteen cases (17.6%), clear symptomology suggesting psychiatric impairment was present, however, there was no definitive diagnosis in the record. The gender distribution for this group was 67 males (72.8 percent) and 25 females (27.2 percent). The mean level of education for this group was 11.29 years (SD = 2.09). The average age was 33.96 years (SD = 13.13). There were 83 individuals (90.2 percent) reported to be right hand dominant and 9 individuals (9.8 percent) reported to be left hand dominant.

The third group consisted of files from 100 neurologically impaired individuals. These files were from individuals tested in an outpatient neuropsychology practice. The majority of the files included evidence of diffuse, generalized injury with no objective evidence to suggest a focal injury, e.g., generalized injury due to a motor vehicle accident rather than focal injury from neoplasm or stroke. The files including a history of generalized injury due to motor vehicle accident or other trauma to the head were
categorized as having a closed head injury. Some tolerance of psychiatric diagnoses was made as psychiatric problems can be secondary to neurological impairment. The following conditions or diagnoses were evident in participant files: alcohol abuse 2%, depression 3%, poly-substance abuse 1%, seizure disorder 3%, bi-frontal atrophy 1%, right parasagittal benign tumor removal 1%, open head injury 2%, closed head injury 87%, cerebral vascular accident 2%, sub-dural hematoma 3%, anoxia 1%, aneurysm 1%, cerebral vascular disease 1%, and transient ischemic attacks 1%. The gender distribution for this group was 56 males (56%) and 44 females (44%). The mean level of education for this group was 12.43 years (SD = 2.89). The average age was 38.40 years (SD = 14.78). There were 91 individuals (91.0%) identified as right hand dominant, 7 individuals (7.0%) identified as left hand dominant, and 2 individuals (2.0%) identified as ambidextrous.

Instruments

*Hooper Visual Organization Test (VOT)*

The Hooper Visual Organization Test (Hooper, 1983) was originally developed by H. Elston Hooper and published in 1958. In the manual, the VOT is described as:

a brief screening instrument designed to measure the ability of adolescents and adults to organize visual stimuli, a task that is particularly sensitive to neurological impairment. The test consists of 30 line drawings depicting simple objects which have been cut into pieces and re-arranged in a puzzle-like fashion.
The respondent is asked to identify what each object would be if it were put back together correctly (p. 1).

Correct answers are scored as one point with some items having alternate answers which are worth half a point. The score for the test is the total number of points earned. The manual provides tables to correct for an examinee’s age and level of education. Hooper suggested a cutoff score of 24 out of 30 with raw scores lower than 24 suggesting neurological impairment. The raw scores can be used to determine a range of impairment with 25-30 suggesting no impairment, 20-24 suggesting mild impairment, 10-19 suggesting moderate impairment, and 0-9 suggesting severe impairment. Interpretation through the use of cutoff scores and range of impairment can similarly be used with age and education corrected scores though it is noted in the manual that these corrections limit the highest score possible for some combinations of age and level of education.

The reliability of the VOT has been found to be quite high. The original reliability studies completed by Hooper and quoted in the manual show a split-half correlation coefficient of .82 in a population of college students and of .78 in a clinical population (Hooper, 1983). Gerson (1974) completed additional reliability studies and found a split-half correlation coefficient of .80 in a more diverse clinical population. The test-retest reliability of the VOT was not reported in the manual, however, Lezak (1982) reported a coefficient of concordance of .86 after both 6 and 12 months. Levin, Llabre, and Reisman (1991) found the one year test-retest reliability to be .68 in elderly subjects. Sawrie, Chelune, Naugle, and Luders (1996) found the 8 month test-retest reliability to be .75 in intractable epileptics.
The validity of the VOT as a measure of neurological impairment is the subject of this study and is addressed in depth in Chapter 2. Briefly, the validity of the VOT has been investigated in two aspects, its validity as a screening tool for neurological impairment and its validity as a measure of visual-spatial skills. While the validity of the test as a screening tool for brain damage has been contested, recent studies have generally supported the validity of the VOT as a measure of perceptual organization abilities (Spreen & Strauss, 1998). With the changing understanding of brain injury and the knowledge that neurological impairment can be limited to certain abilities with other abilities remaining intact, the association of the VOT with certain perceptual abilities has provided arguments against its use in isolation as a screen for brain injury.

*Wechsler Adult Intelligence Scale-Revised (WAIS-R)*

The WAIS-R (Wechsler, 1981) is the 1981 revision of the Wechsler series of intelligence tests that began in 1939 with the Wechsler Bellevue Intelligence Scale. The original WAIS was published in 1955. According to the WAIS-R manual, the WAIS-R is designed as a test of general intelligence and is an attempt to quantify an individual’s capacity for intelligent behavior (Wechsler, 1981). The manual further explains that the WAIS-R uses sets of standardized questions and tasks to assess an individual’s potential for purposeful and useful behavior. The actual measure is comprised of 11 subtests that are divided into two scales. The Verbal scale is made up of six subtests: Information, Digit Span, Vocabulary, Arithmetic, Comprehension, and Similarities. An individual’s raw scores on each of the individual subtests are converted into Scale Scores by comparing the raw score to the normative reference group for the individual’s age group.
These Scale Scores that are based on an individual's age can now be added together and converted to a Verbal IQ (Intelligence Quotient) score. The WAIS-R includes 5 Performance subtests: Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol. The age corrected Performance subtest Scale Scores can be added to obtain a Performance IQ score. The total Scale Scores for all of the subtests can be added together and converted to a Full Scale IQ score.

The WAIS-R manual reports the reliability of the individual subtests as well as the Verbal, Performance, and Full Scale IQ scores. The reliability coefficients for the subtests are reported in the manual in table format as they were computed separately for each age group on each subtest. Most coefficients are from split-half reliability computations while those for Digit Symbol and Digit Span are test-retest coefficients. The manual reports the range of reliability coefficients to be from .52 to .96 for all the subtests and age ranges. Only six coefficients are said to fall below .70 with the coefficients for the Verbal subtest being slightly higher on average than the average for the Performance subtests. Spreen and Strauss (1998) reported a split half reliability across the non-speeded subtests at above .88. The reliability coefficients for the Verbal, Performance, and Full Scale IQ's across all nine age groups used are .97, .93, and .97, respectively (Wechsler, 1981).

In reporting the validity of the WAIS-R, the manual references studies of the correlation between IQ scores and educational performance, factor analysis research, and the correlation between the WAIS-R and other intelligence measures. A correlation of .50 between Full Scale IQ and performance in school is reported. Factor analysis is said to support the concepts of separate verbal and performance scales as three basic factors are
identified: verbal comprehension, perceptual organization, and memory. The correlation between the WAIS-R Full Scale IQ and the Stanford-Binet IQ was reported to be .85. Spreen and Strauss (1998) cite various studies which suggest correlations between the Wechsler IQ score and other measures of intelligence that are between .50 and .80. They also note that Verbal and Full Scale IQ’s are often better correlated with scholastic achievement while Performance IQ is more often found to be affected by neurological impairment.

*Symbol Digit Modalities Test (SDMT)*

According to the SDMT manual (*Smith, 1991*), the SDMT is a measure of cerebral dysfunction that requires the conversion of meaningless geometric designs into written or oral number responses. The test is one page in length. At the top of the page is a key that pairs nine symbols with the numbers 1 through 9. The remainder of the page contains 10 practice items and 110 test items. All of the items are similar in that they consist of symbols with empty boxes below them for the examinee to complete with the matching number from the key. The examinee is allowed ninety seconds to complete the task for both written and oral presentations. The format of the task, with the examinee being required to provide number responses, allows for oral administration of the test in cases in which writing ability is impaired. In addition to providing instructions for administering either the written or oral versions of the task, the manual provides instructions for administering both versions to the same individual allowing comparison of performance between the two tasks (*Smith, 1991*).
Interpretation of SDMT scores is facilitated by normative information provided in the manual. Adult norms are provided, adjusted for age and education. Normative information is provided based on two normative samples. Centrofanti (1975) provided data on 420 normal adults and a second sample of 887 normal adults was taken by the test authors to add additional normative data. Normative information is provided in table format and divided into two sections, those with 12 years of education or less and those with 13 years of education or more. Six age groups are separated from ages 18-24 up to the oldest group comprised of age 65 and above. Mean scores and scores for +/- .5 to 3.0 standard deviations are reported for each age group so that classification of clinical impairment can be made based on clinical judgment.

The manual reports reliability data on the SDMT from both test-retest data and comparisons between the written and oral forms of the measure. Using data from 80 normal adults, the test-retest reliability was found to be .80 for the written form and .76 for the oral format. The interval was 29.4 days. In data from untreated aphasic patients with an interval of 22.6 months, the mean scores at the initial testing and then upon retesting were shown to be within 1 point for both written and oral versions. Spreen and Strauss (1998) report results from Uchiyama et al. (1994) finding correlations of greater than .72 between administrations given yearly over two years. Regarding comparisons between written and oral forms, it is stated that there is a tendency for individuals to score higher on the oral form versus the written form. The correlation between scores on written and oral forms is reported as .82 at initial testing and .84 upon retesting in the study of 80 normal adults. In a study of 887 normal adults, the correlation was found to be .78 between the two forms after only one administration. A correlation of .88 between
written and oral forms was found in head injured patients as reported by Ponsford and Kinsella (as cited in Spreen & Strauss 1998).

The validity of the SDMT as a measure of neurological impairment is reported in the manual, numerous studies are cited suggesting that SDMT scores are lower in individuals with diagnosed neurological impairment than normals or are impaired when compared to age adjusted norms. Evidence of impairment based on normative data is reported in studies of commissurotomy patients (split brain), acute cerebrovascular disease, chronic aphasia following stroke, Huntington’s syndrome, Cushing’s syndrome, and those with chronic brain lesions. Spreen and Strauss (1998) reported that impaired performance on the SDMT has been associated with several conditions such as epilepsy, organic solvent exposure, Parkinson’s disease, lack of exercise in older adults, aging, general fitness, substance abuse, and closed head injury.

Of interest in this study, some studies reported in the manual examine the validity of the SDMT in differentiating “organics” or those with neurological impairment from those with “functional” or psychiatric disorders. The manual references a personal communication from C. G. Watson suggesting higher scores in those with neurological impairment than those with psychiatric impairment when examining SDMT scores from those with brain dysfunction and those with chronic schizophrenia. The manual authors suggest that this may be explained by evidence that chronic schizophrenia represents an underlying organic process and is incorrectly viewed as a functional disorder. When considering several groups of individuals including those with brain damage, never married schizophrenics, married or previously married schizophrenics, affective psychotics, alcoholics, and those with combined neurotic personality disorders, Watson,
Gasser, Schaefer, Buranen, and Wold (1981) found that the SDMT correctly classified 83 percent of those with brain damage and 81 percent of the remaining groups. Watson, Davis, and Gasser (1978) reported correct classification of 88 percent of the organics and 79 percent of those with depressive disorders.

*Raven Progressive Matrices Short Form*

Raven’s Progressive Matrices (*Channell et al., 1997; Raven, 1996*) is considered a test of inductive reasoning (Spreen & Strauss, 1998). The test items require the examinee to use a given pattern or series of visual patterns to infer a rule and then use the rule to identify the next item in a series or the missing part of a pattern. The items become progressively more difficult. Raven’s Standard Progressive Matrices test consists of 60 items grouped into five sets, A-E. Each set involves a different principle (Spreen & Strauss, 1998). Spreen and Strauss reported the test-retest reliability of the measure to be above .8. Reliability estimates are said to be about .7. Spreen and Strauss reported that some consider the Raven Matrices a test of Spearman’s g, a measure of general intellectual ability. Correlations of around .7 were reported with the Wechsler and Stanford-Binet intelligence tests. Berker et al. (1979) report correlations with WAIS Performance, Verbal, and Full Scale IQ scores between .61 and .66 for both the Standard Progressive Matrices and a 30-item short form of the test. In the current study, the short form of Raven’s Standard Progressive Matrices is used. A 30-item Short Form was developed and standardized by Channell et al. (1997). The Raven Matrices Short Form (SF) was created through a step-wise total correlation procedure using data from the 60-item version used with 115 clinical subjects. Reliability was confirmed by comparing the
predictive value in 115 separate clinical cases. Norms are available for ages 8-90 based on 466 normal subjects. They are included in the Michigan Neuropsychological Battery Manual (Smith & Berker, n.d.)

Data Analysis

The research design was a between groups design with three primary groups of participants. The participants in the first group, or normal group, were individuals with no established neurological or psychiatric impairment. The psychiatrically impaired group consisted of individuals who had been tested during their inpatient psychiatric treatment and had no diagnosis of neurological impairment. The third group consisted of neurologically impaired individuals who had been evaluated in an outpatient setting and had an established diagnosis of neurological impairment or who had confirmed trauma known to cause neurological impairment such as a closed head injury in a vehicle accident. The criterion variables included raw as well as age and education corrected scores on the VOT. The Verbal and Performance IQ scores from the WAIS-R, raw scores from SDMT written test, and Raven Matrices SF scores were used as predictor variables.

Basic correlations were computed between age, gender, and years of education and VOT scores to determine if the relationships between these variables and VOT scores seen in previous studies were found in the archival data set. The ability of the VOT to distinguish between normal, non brain injured subjects from neurologically impaired and psychiatrically impaired groups was analyzed using the Welch statistic (Welch, 1951) utilizing age and level of education corrected scores since the homogeneity of variance...
assumptions for an analysis of variance were not met. Additionally, the Welch statistic was used to determine differences in group performance based on qualitative scoring methods. The individual item responses were also qualitatively evaluated to determine if there were any obvious differences between the three groups in terms of the specific incorrect responses given. The percentage of full-credit responses were analyzed using Chi Square analyses to determine if the three groups showed differences in the frequency of full-credit responses to individual items. This study includes an item analysis to provide information about the nature of the individual VOT items. Finally, the association of VOT performance with other neuropsychological tests and tests of visual-spatial abilities was examined through comparisons to scores on other measures. Hierarchical regression analysis was used to determine the amount of variance in VOT scores that is explained by the other measures being employed, including the Raven Matrices SF, SDMT, and WAIS-R Performance IQ Scores.
CHAPTER 4

RESULTS AND ANALYSES

The archival file review yielded 338 data files for analysis. Outlier analysis was completed using two methods. VOT raw scores were compared with group means. Any data file with a VOT raw score of greater than 3.5 standard deviations from the group mean was considered an outlier. No outliers were identified with this method. Additional analysis was completed using the Mahalanobis distance (Tabachnick & Fidell, 2001). No outliers were identified with this method. The following analyses were completed with this entire data set unless otherwise specified.

Descriptive Statistics

The entire sample yielded a mean VOT score of 25.75 with a standard deviation of 3.75. These scores were corrected for age and education level using the table provided in the VOT manual (Hooper, 1983). The mean corrected score was 24.44 with a standard deviation of 3.46. The mean Full Scale IQ score of the sample was 100.22 (SD = 20.03). Means for Verbal and Performance IQ’s were 100.01 (SD = 19.36) and 99.55 (SD = 19.18) respectively. These WAIS-R scores are consistent with what would be expected given the estimated population mean for the three IQ measures is 100 (SD= 15).

Descriptive data is reported on all measures in Table 1. The n for the SDMT and Raven Matrices SF measures is lower due to the absence of these scores in some of the archival files.
### Table 1

**Entire Sample Descriptive Statistics**

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOT</td>
<td>338</td>
<td>10</td>
<td>30</td>
<td>25.52</td>
<td>3.75</td>
</tr>
<tr>
<td>Corr. VOT</td>
<td>338</td>
<td>11</td>
<td>30</td>
<td>24.44</td>
<td>3.46</td>
</tr>
<tr>
<td>SDMT written</td>
<td>335</td>
<td>0</td>
<td>105</td>
<td>44.16</td>
<td>16.34</td>
</tr>
<tr>
<td>SDMT oral</td>
<td>335</td>
<td>0</td>
<td>110</td>
<td>51.22</td>
<td>18.82</td>
</tr>
<tr>
<td>Raven SF</td>
<td>302</td>
<td>2</td>
<td>30</td>
<td>20.30</td>
<td>6.67</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>338</td>
<td>71</td>
<td>150</td>
<td>100.22</td>
<td>20.03</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>338</td>
<td>68</td>
<td>150</td>
<td>100.01</td>
<td>19.36</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>338</td>
<td>64</td>
<td>150</td>
<td>99.55</td>
<td>19.18</td>
</tr>
</tbody>
</table>

*Note.* Age and education corrected VOT scores (Corr. VOT); Full Scale, Verbal, and Performance IQ scores are from the WAIS-R

Group mean scores as discussed here are reported in Table 2. The mean scores for the normal group are higher than for the other two groups on all variables measured. This finding is consistent even in mean age and education corrected VOT scores.

### Table 2

**Group Mean Scores on Neuropsychological Measures**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>VOT</td>
<td>27.15</td>
<td>2.16</td>
<td>23.73</td>
</tr>
<tr>
<td>Corr. VOT</td>
<td>25.84</td>
<td>2.25</td>
<td>22.86</td>
</tr>
<tr>
<td>SDMT written</td>
<td>53.98</td>
<td>13.92</td>
<td>34.02</td>
</tr>
<tr>
<td>SDMT oral</td>
<td>62.24</td>
<td>16.54</td>
<td>40.33</td>
</tr>
<tr>
<td>Raven SF</td>
<td>24.35</td>
<td>3.69</td>
<td>15.62</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>115.60</td>
<td>15.36</td>
<td>83.16</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>115.02</td>
<td>14.60</td>
<td>84.13</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>112.34</td>
<td>15.40</td>
<td>83.89</td>
</tr>
</tbody>
</table>

*Note.* age and education corrected (Corr.), neurologically impaired (Neuro.), and psychiatrically impaired (Psych.)
Correlations

Pearson two-tailed correlations for the entire sample between demographic and predictor variables are reported in Table 3. As with previous studies, VOT scores were not significantly correlated with gender. There was a slight correlation between gender and an individual’s number of years of education. In the entire sample, female participants had slightly more education than the males. Significant correlations were found between age and VOT scores and between education and VOT scores (-.24 and .28 respectfully). The correlations between age and education and VOT scores are consistent with what has been found previously with older adults having lower scores and those with higher levels of education scoring better. No significant correlation was found between age and education corrected VOT scores and individuals’ age or years of education. This suggests that the use of the correction table for age and education provided in the manual effectively minimized the relationship between the new corrected VOT scores and these variables in this sample.

Correlations between demographic variables and criterion variables were computed for each of the three groups. These correlations are also reported in Table 3. Within the three groups, as with the sample as a whole, gender was not found to significantly correlate with raw or age and education corrected VOT scores. The significant correlation between gender and the number of years of education which was observed in the whole sample was also found to be significant in the psychiatrically impaired group while it was not significant in the other two groups. Females in the psychiatrically impaired group tended to have more education than the males in that group.
Table 3

Two-way Pearson Correlations Between Demographic and Criterion Variables

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire sample</td>
<td>1. VOT</td>
<td>-</td>
<td><strong>.913</strong></td>
<td><strong>-.241</strong></td>
<td><strong>.278</strong></td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>2. Corr. VOT</td>
<td>-</td>
<td>.077</td>
<td>.080</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Age</td>
<td>-</td>
<td>.051</td>
<td>-</td>
<td>.085</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Years of Ed</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>.131</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1. VOT</td>
<td>-</td>
<td><strong>.711</strong></td>
<td><strong>-.291</strong></td>
<td>.104</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td>2. Corr. VOT</td>
<td>-</td>
<td><strong>.342</strong></td>
<td><strong>-.258</strong></td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Age</td>
<td>-</td>
<td>-.056</td>
<td>-.043</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Years of Ed</td>
<td>-</td>
<td>.076</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Psychiatrically Impaired</td>
<td>1. VOT</td>
<td>-</td>
<td><strong>.965</strong></td>
<td><strong>-.480</strong></td>
<td>.154</td>
<td>-.056</td>
</tr>
<tr>
<td></td>
<td>2. Corr. VOT</td>
<td>-</td>
<td><strong>.316</strong></td>
<td><strong>-.015</strong></td>
<td>.009</td>
<td>-.065</td>
</tr>
<tr>
<td></td>
<td>3. Age</td>
<td>-</td>
<td>.053</td>
<td>.150</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Years of Ed</td>
<td>-</td>
<td>.243</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Neurologically Impaired</td>
<td>1. VOT</td>
<td>-</td>
<td><strong>.922</strong></td>
<td><strong>-.320</strong></td>
<td>.179</td>
<td>-.070</td>
</tr>
<tr>
<td></td>
<td>2. Corr. VOT</td>
<td>-</td>
<td>.015</td>
<td>.018</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Age</td>
<td>-</td>
<td>.007</td>
<td>.154</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Years of Ed</td>
<td>-</td>
<td>-.041</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Note. education (Ed), age and education corrected VOT scores (Corr. VOT)

* indicates significance at the .05 level ** indicates significance at the .01 level

Age was found to significantly correlate with VOT raw scores. The correlations are slightly higher than that seen in the sample as a whole. The correlations range from -.29 in the normal group to -.48 in the psychiatrically impaired group. VOT raw scores did not significantly correlate with years of education within each of the three groups.

While age and education did not correlate significantly with the corrected VOT scores in the entire sample, significant correlations were seen in the normal group (r = -.34 for age
and $r = -0.26$ for education. In the psychiatrically impaired group, there was a significant correlation between age and the age and education corrected VOT scores ($r = -0.32$).

Primary Analyses

Primary analyses reported in this section are organized based on the study research questions and the respective null hypotheses.

Research Question 1

Is there a significant difference in VOT scores between three groups of individuals consisting of individuals without neurological or psychiatric impairment, individuals with psychiatric impairment, and individuals with known neurological impairment?

Null Hypothesis 1a:

There is no significant difference in mean VOT raw scores, between normal controls, psychiatrically impaired and neurologically impaired individuals.

Group mean differences were tested for the three groups. A one-way, between-groups Analysis of Variance (ANOVA) on VOT raw scores was attempted. The test of the homogeneity of variances was performed using the Levene Statistic (Tabachnick & Fidell, 2001). This test yielded a significant result suggesting that the three groups have unequal variances. The Welch statistic (Welch, 1951) was used as an alternative as it is a more robust test of the equality of means for use with groups that have unequal variances.
The difference between the means was found to be significant (Welch statistic = 33.58 (2, 162.11), p < .001). Post hoc tests were performed using the Dunnett C (Dunnett, 1982) procedure as it does not assume equal variances for the three groups. Post hoc testing revealed that the psychiatrically impaired group mean VOT score was significantly lower than the normal group mean VOT score. The neurologically impaired group mean was also significantly lower than the normal group mean VOT score. These differences were significant at the .01 level. The difference between the means of the psychiatrically impaired and the neurologically impaired groups was not significant. Null hypothesis 1a was rejected because there was a significant difference in raw VOT score means between the normal group and the neurologically impaired group and between the normal group and the psychiatrically impaired group.

Null Hypothesis 1b:

There is no significant difference in mean age and education corrected VOT scores, between normal controls, psychiatrically impaired and neurologically impaired individuals.

Group mean differences were tested for the three groups. A one-way, between-groups ANOVA was attempted on the age and education corrected VOT scores. The test of the homogeneity of variances was performed using the Levene Statistic. This test yielded a significant result. The Welch statistic was again used to determine the significance of the differences between group means. This difference between the group means was found to be significant (Welch statistic = 28.28 (2, 170.25), P < .001). Post
hoc tests using the Dunnet C procedure revealed that the mean score for the psychiatrically impaired group was significantly lower than that of the normal group. The neurologically impaired group mean was significantly lower than the mean age and education corrected VOT score for the normal group. These differences were significant at the .01 level. The difference between the means of the psychiatrically impaired group and the neurologically impaired was not significant. Null hypothesis 1b was rejected. There was a significant difference in the age and education corrected VOT score means between the normal group and both the neurologically and psychiatrically impaired groups.

Null Hypothesis 1c:

There is no difference between normal controls, psychiatrically impaired and neurologically impaired individuals that is evident based on qualitative scoring.

Mean scores for the qualitative scoring of Isolate, Perseverative, Bizarre, and Neologistic responses are consistent with previous studies suggesting that these types of responses are rare in general with no Neologistic responses recorded in this data set. There were only three perseverative responses in the entire data set with two in the normal group and one in the psychiatrically impaired group. The difference between the three groups based on qualitative scoring was investigated by examining the number of isolate and bizarre responses observed in each group. First, the mean number of isolate and bizarre responses were examined to determine if there were obvious differences between the three groups based on inspection. Secondly, group mean differences were
tested on both isolate and bizarre responses to help determine if the number of these two types of errors was useful in distinguishing between the three groups.

Table 4 shows the frequency of isolate responses for the sample across the three groups. When assessing the usefulness of qualitative scoring it would be helpful if there were an obvious pattern that would suggest that psychiatrically or neurologically impaired individuals were more likely to make errors that fall into one of the four qualitative scoring categories. It is clear from the data that having one or more isolate responses made it more likely that the participant was from either the psychiatrically or neurologically impaired group. Less than 11 percent of the normal group had one or more isolate responses compared to almost 49 percent of the psychiatric group and 33 percent of the neurologically impaired group. Having more than one isolate response was infrequent across all the groups with just over 1 percent of the normal group, 15 percent of the psychiatrically impaired group, and 10 percent of the neurologically impaired group having more than one isolate response.

Table 5 shows the frequency of bizarre responses for the sample across the three groups. Bizarre responses occurred less frequently than isolate responses in this sample. Just over 8 percent of the normal group, 27 percent of the psychiatrically impaired group, and 16 percent of the neurologically impaired group had one or more bizarre responses. Percentages of individuals with more than one bizarre response were similar across the three groups: 2 percent for the normal group and approximately 5 percent for both the psychiatrically and neurologically impaired groups.
### Table 4

**Isolate Responses by Group**

<table>
<thead>
<tr>
<th>Group</th>
<th># of errors</th>
<th>Frequency</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>131</td>
<td>89.7</td>
<td>89.7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>13</td>
<td>8.9</td>
<td>98.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.7</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0.7</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>146</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
<tr>
<td>Psychiatrically Impaired</td>
<td>0</td>
<td>47</td>
<td>51.1</td>
<td>51.1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>31</td>
<td>33.7</td>
<td>84.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>9.8</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>4.3</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1</td>
<td>1.1</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>92</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
<tr>
<td>Neurologically Impaired</td>
<td>0</td>
<td>67</td>
<td>67.0</td>
<td>67.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>23</td>
<td>23.0</td>
<td>90.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10</td>
<td>10.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5

**Bizarre Responses by Group**

<table>
<thead>
<tr>
<th>Group</th>
<th># of errors</th>
<th>Frequency</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>134</td>
<td>91.8</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>9</td>
<td>6.2</td>
<td>97.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>1.4</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.7</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>146</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
<tr>
<td>Psychiatrically Impaired</td>
<td>0</td>
<td>67</td>
<td>72.8</td>
<td>72.8</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>20</td>
<td>21.7</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>3.3</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>1.1</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.1</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>92</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
<tr>
<td>Neurologically Impaired</td>
<td>0</td>
<td>84</td>
<td>84.0</td>
<td>84.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>11</td>
<td>11.0</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>3.0</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.0</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1</td>
<td>1.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The mean numbers of isolate and bizarre responses for each group are reported in Table 6. The mean number of isolate responses for the three groups was .12, .74, and .43 for the normal, psychiatric, and neurologically impaired groups respectively. The test of the homogeneity of variances was performed using the Levene Statistic. This test yielded a significant result ($p < .001$) suggesting that the three groups have unequal variances. The Welch statistic was again used to test the differences between group means. This difference between the means was found to be significant (Welch statistic = 20.07 (2, 160), $p < .001$). Post hoc tests using the Dunnett C procedure revealed that the mean number of isolate responses for the psychiatrically impaired group was significantly higher than that of the normal group ($p < .01$). The mean number of isolate responses in the neurologically impaired group was also significantly higher than the mean number of isolate responses for the normal group ($p < .01$). The difference between the means of the psychiatrically impaired group and the neurologically impaired group was also significant. This difference was significant at the .01 level.

Table 6

Isolate and Bizarre Response Means and Standard Deviations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Isolate</td>
<td>.38</td>
<td>.75</td>
<td>.12</td>
<td>.40</td>
</tr>
<tr>
<td>Bizarre</td>
<td>.24</td>
<td>.75</td>
<td>.12</td>
<td>.46</td>
</tr>
</tbody>
</table>

Note. neurologically impaired (Neuro.) and psychiatrically impaired (Psych.)
The mean number of bizarre responses for the three groups was .12, .36, and .30 for the normal, psychiatric, and neurologically impaired groups respectively. The test of the homogeneity of variances was performed using the Levene Statistic. This test yielded a significant result ($p < .001$) suggesting that the three groups have unequal variances. The Welch statistic was used to analyze the difference between the means. A significant difference between group means was found (Welch statistic = 4.96 (2, 166.48), $p = .008$). Post hoc tests using the Dunnett C test revealed that the mean number of bizarre responses for the psychiatrically impaired group were significantly higher than the mean number of bizarre responses for the normal group ($p < .05$). Neither the difference between the means of the normal group and the neurologically impaired group nor the difference between the psychiatrically impaired group and neurologically impaired group were significant.

Null hypothesis 1c was rejected as the normal group mean number of isolate responses was significantly lower than the means of the other two groups. Additionally, the neurologically impaired group mean number of isolate responses was significantly lower than the mean of the psychiatrically impaired group. For bizarre responses, the normal group mean was significantly lower than the mean of the psychiatrically impaired group.

**Research Question 2**

To what extent do individuals with diagnosed neurological impairment, psychiatric impairment and normals differ in their individual item responses on the Hooper Test of Visual Organization?
Null hypothesis 2:

Normal individuals and individuals diagnosed with neurological impairment or psychiatric disorder will not demonstrate significant differences in frequency of correct responses given to individual items of the Hooper VOT.

To investigate null hypothesis 2, a frequency count of full credit and non-full credit answers provided for individual items of the VOT was completed for the entire sample and also for the individual groups. The frequencies of full credit responses were inspected to identify any patterns of responses in the three groups. Chi Square analyses were performed to compare the frequency distributions of correct and incorrect responses for each item across the three groups. Additionally, the specific incorrect responses were examined to determine if the three groups displayed differences in the frequencies of specific incorrect responses.

Table 7 shows the percentages of full credit responses for each item for the whole sample and delineated by group. For many items, the percentages were relatively equal across the three groups. There were, however, three general patterns that were observed. The most common pattern, as might be expected by looking at mean scores for the three groups, was to see a higher percentage of full-credit responses to an item in the normal group, with the neurologically impaired group slightly lower, and the psychiatrically impaired group slightly lower yet. The items in which the frequency differences were found to be significant (p < or = .01) using Chi Square analysis that reflected this pattern included items 14 (cane, hockey stick), 17 (chair), 19 (teapot), 21 (flower), 22 (mouse),
Table 7

Percent Full-credit Responses for Individual Items by Group

<table>
<thead>
<tr>
<th>Item number</th>
<th>Combined</th>
<th>Normal</th>
<th>Psych.</th>
<th>Neuro.</th>
<th>$\chi^2$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.4</td>
<td>97.3</td>
<td>88.0</td>
<td>96.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>96.2</td>
<td>98.6</td>
<td>93.5</td>
<td>95.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>99.4</td>
<td>99.3</td>
<td>98.9</td>
<td>100.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>96.7</td>
<td>98.6</td>
<td>97.8</td>
<td>93.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>96.2</td>
<td>98.6</td>
<td>95.7</td>
<td>93.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>96.5</td>
<td>98.6</td>
<td>93.5</td>
<td>96.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>97.6</td>
<td>99.3</td>
<td>96.7</td>
<td>95.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>88.8</td>
<td>91.8</td>
<td>83.7</td>
<td>89.0</td>
<td>.157</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>95.3</td>
<td>97.9</td>
<td>89.1</td>
<td>97.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>94.4</td>
<td>97.9</td>
<td>89.1</td>
<td>94.0</td>
<td>*.016</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>98.5</td>
<td>100.0</td>
<td>97.8</td>
<td>97.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>92.0</td>
<td>95.9</td>
<td>84.8</td>
<td>93.0</td>
<td>**.008</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>95.6</td>
<td>99.3</td>
<td>92.4</td>
<td>93.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>66.3</td>
<td>78.8</td>
<td>47.8</td>
<td>65.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>82.5</td>
<td>91.1</td>
<td>76.1</td>
<td>76.0</td>
<td>**.003</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>78.1</td>
<td>83.6</td>
<td>66.3</td>
<td>81.0</td>
<td>**.005</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>49.7</td>
<td>61.6</td>
<td>35.9</td>
<td>45.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>96.7</td>
<td>99.3</td>
<td>93.5</td>
<td>96.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>77.2</td>
<td>89.0</td>
<td>62.0</td>
<td>72.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>94.4</td>
<td>98.6</td>
<td>91.3</td>
<td>91.0</td>
<td>*.012</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>79.0</td>
<td>90.4</td>
<td>65.2</td>
<td>75.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>89.4</td>
<td>96.6</td>
<td>80.4</td>
<td>87.0</td>
<td>**.002</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>92.6</td>
<td>97.3</td>
<td>87.0</td>
<td>91.0</td>
<td>**.010</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>86.1</td>
<td>95.9</td>
<td>76.1</td>
<td>81.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>76.9</td>
<td>86.3</td>
<td>65.2</td>
<td>74.0</td>
<td>**.001</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>75.4</td>
<td>88.4</td>
<td>62.0</td>
<td>69.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>38.8</td>
<td>41.8</td>
<td>41.3</td>
<td>32.0</td>
<td>.334</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>63.9</td>
<td>76.7</td>
<td>50.0</td>
<td>58.0</td>
<td>**&lt;.001</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>50.0</td>
<td>61.6</td>
<td>43.5</td>
<td>39.0</td>
<td>**.001</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>61.5</td>
<td>67.8</td>
<td>58.7</td>
<td>55.0</td>
<td>.103</td>
<td></td>
</tr>
</tbody>
</table>

Note. N=338, normal n = 146, psychiatrically impaired (Psych.) n = 92, and neurologically impaired (Neuro.) n = 100; Correct responses are those receiving full credit. A dash indicates that there was insufficient cell count in more than 30 percent of the cells and Chi Square analyses results were not reported.

*Indicates significance at the .05 level, ** indicates significance at the .01 level.
and 25 (block). For example, for item 14, 79 percent of the normal group gave full-credit responses while only 65 percent of the neurologically impaired group and 48 percent of the psychiatrically impaired group gave full-credit responses (Chi square = 24.28, p < .001). For item 17, 62 percent of the normal group gave full-credit responses. Only 45 percent of the neurologically impaired and 36 percent of the psychiatrically impaired groups gave full-credit responses (Chi Square = 16.25, p < .001).

Another pattern observed was the tendency, on some items, for the psychiatrically and neurologically impaired groups to score similarly and obviously lower than the normal group. This is evident on items 15 (sailboat), 23 (book), 26 (lighthouse), 28 (key), and 29 (ring). On item 15, 91 percent of the normal group gave a full credit response while only 76 percent of the other two groups gave full-credit responses (Chi Square = 12.88, p = .01). Similarly, on item 26, 88 percent of the normal group gave full-credit responses compared to 69 and 62 percent in the neurologically impaired and psychiatrically impaired groups, respectively (Chi Square = 24.41, p < .001).

Two items showed the pattern of the normal and neurologically impaired groups scoring similarly and the psychiatrically impaired group scoring lower. This pattern was seen in items 16 (teakettle) and 12 (basket). For item 16, 66 percent gave full-credit responses in the psychiatrically impaired group, compared to 84 percent in the normal and 81 percent in the neurologically impaired group (Chi Square = 10.52, p = .005). The frequency difference was also significant for item 12 (Chi Square = 9.66, p = .008).

In looking at individual items, the frequencies of specific incorrect responses were sometimes notably different between the groups. For item 17 (Chair), of the incorrect responses, 93 percent of the normal group gave the common, half-credit response of sofa,
or a synonym like couch or davenport. Only 81 and 80 percent of the psychiatrically impaired or neurologically impaired individuals who gave incorrect answers answered this way.

Similar to the above example, for item 19 (teapot, cream pitcher), 11 percent of the normal group gave non full-credit answers and nearly 70 percent of those responded with some variation of the common incorrect response of “tea kettle”, “kettle”, “coffee pourer”, etc. In the psychiatrically and neurologically impaired groups only 43 and 39 percent respectively provided a non full-credit response of this type. Item 16 is a teakettle. For this item, the common incorrect response is teapot, kettle, coffee pot, or some variation. In the normal group, 96 percent of incorrect responses were of this type, compared to 70 and 84 percent in the psychiatrically impaired and neurologically impaired groups.

Item 22 (mouse) is interesting for the pattern of incorrect responses observed in this sample. A relatively common isolate response was “pipe.” This item may provide some qualitative evidence for impairment as none of the normal sample gave this isolate response while “pipe” accounted for 50 percent of the incorrect responses in the psychiatric group and 23 percent in the neurologically impaired group.

Null hypothesis 2 was rejected. Normal individuals and individuals diagnosed with neurological impairment or psychiatric disorder demonstrated differences in the frequency of correct responses given to individual items of the Hooper VOT.

In addition to the group differences reported above, several other observations were made regarding the scale. When the percentages of full-credit responses are considered (Table 7), it is obvious that the items are not in the order of difficulty
observed in this sample. Item 3 (table, bench) was answered correctly by 100 percent of the neurologically impaired group and 99 percent of the other two groups or 99 percent of the entire sample. Items 1 and 2 were answered correctly by 94 and 96 percent of the entire sample, respectively. It could be argued that this represents initial confusion as to the expectations for the test. This would not be the case for later items. Item 11 (apple, peach, etc) was answered correctly by everyone in the normal group and 98 and 97 percent of the psychiatrically impaired and neurologically impaired groups or 99 percent of the entire sample. Items 4 through 6 were answered correctly by just 96 percent of the total sample and item 8 was answered correctly by only 89 percent of the entire sample. Item 18 (candle) was answered correctly by 97 percent of the sample. In contrast, item 17 was answered correctly by only 49 percent of the sample.

Classical item analysis procedures yielded a Cronbach's Alpha of .80 in this sample. Deleting any one item yielded Cronbach's Alphas ranging from .78 to .81. The scale mean was 25.05 (SD = 4.02). The scale mean with any one item deleted ranged from 24.06 to 24.66. Inter-item correlations ranged from -.07 to .51 with a mean of .14. Corrected item-total correlations are reported in Table 8. Corrected item-total correlations ranged from .12 to .56.

Research Question 3

To what extent does the YOT correlate with other neuropsychological measures including the Verbal and Performance IQ measures from the WAIS-R, Raven's Matrices SF, and the SDMT?
Table 8

*Correlations Between Items and Total Scale Score with Item Removed*

<table>
<thead>
<tr>
<th>Item number</th>
<th>Item-total</th>
<th>Item number</th>
<th>Item-total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.17</td>
<td>16</td>
<td>.17</td>
</tr>
<tr>
<td>2</td>
<td>.21</td>
<td>17</td>
<td>.22</td>
</tr>
<tr>
<td>3</td>
<td>.12</td>
<td>18</td>
<td>.33</td>
</tr>
<tr>
<td>4</td>
<td>.17</td>
<td>19</td>
<td>.33</td>
</tr>
<tr>
<td>5</td>
<td>.22</td>
<td>20</td>
<td>.50</td>
</tr>
<tr>
<td>6</td>
<td>.32</td>
<td>21</td>
<td>.48</td>
</tr>
<tr>
<td>7</td>
<td>.26</td>
<td>22</td>
<td>.52</td>
</tr>
<tr>
<td>8</td>
<td>.35</td>
<td>23</td>
<td>.45</td>
</tr>
<tr>
<td>9</td>
<td>.30</td>
<td>24</td>
<td>.56</td>
</tr>
<tr>
<td>10</td>
<td>.43</td>
<td>25</td>
<td>.53</td>
</tr>
<tr>
<td>11</td>
<td>.28</td>
<td>26</td>
<td>.50</td>
</tr>
<tr>
<td>12</td>
<td>.24</td>
<td>27</td>
<td>.16</td>
</tr>
<tr>
<td>13</td>
<td>.44</td>
<td>28</td>
<td>.46</td>
</tr>
<tr>
<td>14</td>
<td>.35</td>
<td>29</td>
<td>.37</td>
</tr>
<tr>
<td>15</td>
<td>.18</td>
<td>30</td>
<td>.29</td>
</tr>
</tbody>
</table>

*Null Hypothesis 3a*

There are no significant simple correlations between scores on the WAIS-R Performance IQ and Verbal IQ, Raven’s Matrices SF, and the SDMT and uncorrected, or age and education corrected, VOT scores.

Table 9 shows correlations between the criterion and predictor variables for the entire sample. VOT scores were significantly correlated with all other measures recorded. The correlations range from .47 with the VIQ to .61 with the Raven Matrices SF. The correlations are generally ordered as would be expected from the literature. VOT scores correlated higher with measures of visual abilities such as the Raven Matrices SF and PIQ. Correlations were slightly less for VIQ and Full Scale IQ.
Table 9

*Whole Sample Two-tailed Pearson Correlations Between Criterion and Predictor Variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VOT</td>
<td>-</td>
<td><strong>.913</strong></td>
<td><strong>.604</strong></td>
<td><strong>.578</strong></td>
<td><strong>.607</strong></td>
<td><strong>.521</strong></td>
<td><strong>.471</strong></td>
<td><strong>.551</strong></td>
</tr>
<tr>
<td>2. Corr. VOT</td>
<td>-</td>
<td><strong>.426</strong></td>
<td><strong>.423</strong></td>
<td><strong>.471</strong></td>
<td><strong>.486</strong></td>
<td><strong>.434</strong></td>
<td><strong>.522</strong></td>
<td></td>
</tr>
<tr>
<td>3. SDMT W</td>
<td>-</td>
<td><strong>.930</strong></td>
<td><strong>.655</strong></td>
<td><strong>.661</strong></td>
<td><strong>.597</strong></td>
<td><strong>.679</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SDMT O</td>
<td>-</td>
<td><strong>.613</strong></td>
<td><strong>.652</strong></td>
<td><strong>.588</strong></td>
<td><strong>.673</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Raven</td>
<td>-</td>
<td><strong>.697</strong></td>
<td><strong>.673</strong></td>
<td><strong>.674</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. FSIQ</td>
<td>-</td>
<td><strong>.963</strong></td>
<td><strong>.931</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. VIQ</td>
<td>-</td>
<td><strong>.807</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. PIQ</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Age and education corrected VOT scores (Corr. VOT), SDMT written (SDMT W), SDMT oral (SDMT O), Raven Matrices Short Form (Raven), WAIS-R Full Scale IQ score (FSIQ), WAIS-R Verbal IQ score (VIQ), and WAIS-R Performance IQ score (PIQ)

* indicates significance at the .05 level ** indicates significance at the .01 level

Correlations between VOT scores and the other neuropsychological measures in the study within each of the three groups are displayed in Table 10. Table 10 shows that the raw VOT scores were significantly correlated with all of the neuropsychological measures in the study across the three groups.

VOT score correlations with Verbal IQ are similar across the three groups ranging from .31 in the normal group to .34 in the neurologically impaired group. Of the correlations between VOT raw scores and neuropsychological measures in the neurologically impaired group, Verbal IQ demonstrated the lowest correlation. This can be expected if verbal abilities are thought to be more resistant to neurological impairment and therefore a more reliable estimate of pre-morbid functioning. Across the groups there
Table 10

Pearson Correlations Between VOT Scores and Predictor Variables by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Predictor variable</th>
<th>SDMT written</th>
<th>SDMT oral</th>
<th>Raven</th>
<th>FSIQ</th>
<th>VIQ</th>
<th>PIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>VOT</td>
<td><strong>.414</strong></td>
<td><strong>.415</strong></td>
<td><strong>.415</strong></td>
<td><strong>.356</strong></td>
<td><strong>.305</strong></td>
<td><strong>.339</strong></td>
</tr>
<tr>
<td></td>
<td>Corr. VOT</td>
<td>-.034</td>
<td>.008</td>
<td>.066</td>
<td><strong>.219</strong></td>
<td>.167</td>
<td><strong>.247</strong></td>
</tr>
<tr>
<td>Psychiatrically Impaired</td>
<td>VOT</td>
<td><strong>.470</strong></td>
<td><strong>.429</strong></td>
<td><strong>.417</strong></td>
<td><strong>.335</strong></td>
<td><strong>.313</strong></td>
<td><strong>.327</strong></td>
</tr>
<tr>
<td></td>
<td>Corr. VOT</td>
<td><strong>.363</strong></td>
<td><strong>.352</strong></td>
<td><strong>.326</strong></td>
<td><strong>.354</strong></td>
<td><strong>.316</strong></td>
<td><strong>.351</strong></td>
</tr>
<tr>
<td>Neurologically Impaired</td>
<td>VOT</td>
<td><strong>.673</strong></td>
<td><strong>.638</strong></td>
<td><strong>.622</strong></td>
<td><strong>.502</strong></td>
<td><strong>.341</strong></td>
<td><strong>.641</strong></td>
</tr>
<tr>
<td></td>
<td>Corr. VOT</td>
<td><strong>.542</strong></td>
<td><strong>.534</strong></td>
<td><strong>.486</strong></td>
<td><strong>.501</strong></td>
<td><strong>.352</strong></td>
<td><strong>.615</strong></td>
</tr>
</tbody>
</table>

Note. Age and education corrected VOT scores (Corr. VOT), Raven Matrices Short Form (Raven), WAIS-R Full Scale IQ score (FSIQ), WAIS-R Verbal IQ score (VIQ), and WAIS-R Performance IQ score (PIQ)

*indicates significance at the .05 level ** indicates significance at the .01 level

is a pattern of higher correlations between the VOT raw scores and Performance IQ scores than with Verbal IQ scores. This pattern was reported by Greve et al. (2000) as well. This pattern is most striking in the neurologically impaired group. The correlations between VOT scores and both PIQ and VIQ scores for the normal and psychiatrically impaired groups range from .30 to .34. In the neurologically impaired group, the correlation between VIQ and VOT scores remains .34, similar to the correlation seen in the other two groups, while the correlation between PIQ and VOT scores is .64. VOT raw scores from the neurologically impaired group were highly correlated with the SDMT scores (.67 Written and .64 Oral), Raven Matrices SF (.62), and Performance IQ scores (.64). This finding is consistent with the theory that the SDMT is sensitive to generalized impairment and impairment of visual-spatial skills. As the individuals in the neurologically impaired group primarily had generalized impairment, it would be
expected that visual-spatial skills would be impaired as well and therefore VOT scores and SDMT scores would correlate. Raven Matrices and Performance IQ require visual-spatial abilities that are thought to be measured by the VOT. The relatively high correlations (above .60) suggest that impairment of visual-spatial abilities was reflected in scores on these two measures and the VOT.

Correlations between VOT raw scores and the other variables in the other two groups were somewhat lower, ranging from .31 for Verbal IQ in the normal group to .47 for the written form of the SDMT in the psychiatrically impaired group. Interestingly, correction for age and education resulted in lower correlations between VOT scores and the other tests across all of the groups, but with the most significant effect within the normal group. Within the normal group, after age and education correction, VOT scores were not significantly correlated with SDMT scores or Raven Matrices SF scores. The correlations with Full Scale, Verbal, and Performance IQ's were much lower.

Null hypothesis 3a was rejected as there were many significant correlations between scores on the neuropsychological test measures and both raw and age and education corrected VOT scores.

*Null Hypothesis 3b*

Scores on the Performance and Verbal IQ’s, Raven’s Matrices Short Form, and the SDMT written tests will not be significant predictors of VOT scores after age and education are controlled for when examining the entire data set using hierarchical regression analysis.
Hierarchical multiple regression analysis was performed on the raw VOT scores for the entire sample to determine if SDMT written test, WAIS-R Verbal and Performance IQ, and Raven Matrices SF scores were significant predictors of VOT performance when age and years of education were controlled. SDMT written scores were used alone due to the high correlation between written and oral SDMT scores. The variables were chosen based on assumptions regarding what they are thought to measure. Verbal IQ scores were used as an estimate of pre-morbid or overall IQ due to their correlation with Full Scale IQ scores and their tendency to be resistant to generalized neurological impairment. The SDMT is suggested to indicate generalized impairment (Spreen & Strauss, 1998; Watson et al., 1978; Watson et al., 1981). Beyond the amount of variance that might be explained by generalized impairment and intelligence, WAIS-R Performance IQ scores were used to determine the amount of variance explained by general performance abilities including some visual-spatial measures such as Block Design. Additionally, the Raven Matrices SF was added to determine the amount of variance explained by a measure of visual reasoning.

Age and years of education were entered simultaneously in the first model of the regression to control for their reported effects on VOT scores. The SDMT written scores, WAIS-R Verbal and Performance IQ, and Raven Matrices SF were added in the second model. Model 1 was found to account for 14 percent of the variance in VOT scores with $R = .38$, $F (2, 296) = 24.58$, $p < .001$. Age and education were both significant predictors of VOT scores in the first model ($p < .001$). Model 2 accounted for 48 percent of the variance in VOT scores with $R = .69$, $F (6, 292) = 44.54$, $p < .001$. Age ($p = .001$), SDMT ($p = .003$), Performance IQ ($p = .002$), and Raven Matrices SF ($p < .001$) scores.
were all found to be significant predictors of VOT scores in the second model. Verbal IQ 
(p = .647) and years of education (p = .056) were not found to be significant predictors of 
VOT scores. In this initial analysis, however, with one condition index > 30, there 
appeared to be a problem with multicollinearity. A decision was made to drop Verbal IQ 
from the second model given the high correlation between Verbal IQ and Performance IQ 
(r = .807). This is consistent with the suggestion that correlations between predictors of r > .80 are problematic in performing multiple regression (Grimm & Yarnold, 2000). The 
Regression analysis for the second model was then completed without Verbal IQ in the 
model. After deleting Verbal IQ from the second model, multicollinearity was not a 
problem and no condition index was > 30. The unstandardized and standardized 
coefficients are reported in Table 11. A summary of the two models is provided in Table 
12.

The first model, (Table 12) including only age and years of education, accounted 
for 14 percent of the variance in VOT performance, with R = .38, F (2, 296) = 24.58, p < .001. Both age and education were found to be significant predictors of VOT scores in 
this model (Beta = -.25 for age and .29 for years of education, p <.001 for both).

The second model accounted for 48 percent of the variance in VOT scores, with R 
= .69, F (5, 293) = 53.55, p < .001. Age, SDMT written scores, Performance IQ scores, 
and Raven Matrices SF scores were found to be significant predictors of VOT 
performance. The number of years of education was not found to be a significant 
predictor of raw VOT scores in the second model.

Null Hypothesis 3b was rejected as the predictor variables SDMT written scores, 
WAIS-R Performance IQ scores, and Raven Matrices SF scores were found to be
significant predictors of VOT scores after an individual’s age and number of years of education were controlled.

Table 11

*Performance IQ scores are from the WAIS-R

Table 12

*Note. SE is the standard error of the estimate. The models are as follows:
1. Years of education and age
2. Years of education, age, SDMT written, WAIS-R Performance IQ, Raven Matrices SF

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Discussion of the study will address findings related to the three main research questions. Additionally, the analysis of individual item characteristics will be discussed. Finally, limitations of this study will be identified and directions for future research will be suggested.

Before considering the research questions, the relationship between VOT scores and the demographic variables is worthy of discussion. Consistent with previous studies, there was no significant correlation between gender and VOT scores. Age was found to significantly correlate with VOT raw scores in the entire sample and within each of the three groups. Additionally, age was found to be a significant predictor of VOT raw scores in the hierarchical regression analysis with all of the predictors included. This finding is consistent with the previous studies that have found a significant negative correlation between age and VOT scores (Mason & Ganzler, 1964; Tamkin & Jacobsen, 1984; Wenthworth-Rohr et al., 1974).

VOT raw scores did not significantly correlate with years of education within each of the three groups despite a significant correlation between years of education and raw VOT scores across the entire sample. This is consistent with earlier findings by Nabors et al. (1997). In the study by Nabors et al., a significant correlation was found in the whole sample but not within the cognitively intact group. The authors also reported a significant difference between the groups in terms of the number of years of education with the cognitively intact group having more education. The finding of a significant
correlation between VOT raw scores and years of education in the whole sample in the present study and not within the three groups may be, in part, due to the difference in mean education levels between the three groups. In other words, the normal group has the highest mean VOT scores and the highest mean number of years of education (14.32). The neurologically impaired group is next and the psychiatrically impaired group has the lowest mean VOT scores and lowest mean number of years of education (11.29). The lack of correlation within the groups between education and VOT scores may also be due to more restricted ranges of variance in VOT scores within groups in contrast to across all the groups combined. For example, within the normal group the mean was 27.15 with an SD = 2.16 while for the overall sample the mean was 25.52 with an SD = 3.75.

In the hierarchical regression analysis age and education were both significant predictors of VOT scores when entered together in the first model with greater age associated with lower VOT scores and more education associated with higher VOT scores. However, in the full model, with all the other predictor variables entered and controlled for in the model, education was not a significant predictor. Thus, an individual’s level of education is not significantly related to VOT scores in the full model when all of the other predictor variables are entered and controlled for in the analysis. This finding is consistent with the results of regression analysis in other samples as well (Tamkin & Jacobsen, 1984; Walsh et al., 1997). This seems to suggest some overlap and shared variance in the relationships between education and VOT scores and the relationships between other variables in the model and VOT scores. Also another factor in the absence of a significant relationship between education and VOT scores in the full regression model is that the relationship between education and VOT is relatively small.
and this relationship overlaps with the relationship between other predictor variables and VOT scores when entered into the full model. In the study by Nabors et al. (1997), education accounted for only 5.5 percent of the variance in VOT scores when age was controlled. The majority of the evidence, however, seems to suggest that the relationship between education and VOT scores is minimal, specifically when age and intelligence level are controlled (Tamkin & Jacobsen, 1984; Walsh et al., 1997; Wenthworth-Rohr et al., 1974).

The correction for the number of years of education in the present study may have resulted in an over correction in VOT scores in the normal group, at the higher end of the education spectrum, resulting in the significant negative correlation between age and education corrected VOT scores and the number of years of education observed in this study. Given that the number of years of education and WAIS-R Full Scale IQ are significantly correlated ($r = .60, p < .01$) in the current study, (Spreen and Strauss (1998) report correlations between $r = .5$ to $.8$ between Wechsler IQ scores and measures of intelligence and academic achievement) the current correction for educational level in the manual may indirectly correct for differences in intelligence level which have been shown to be significant predictors of VOT scores (Gerson, 1974; Greve et al., 2000; Tamkin & Jacobsen, 1984; Wenthworth-Rohr et al., 1974). As intelligence level is important to consider in VOT score interpretation but may not always be quickly measured, the current correction for education level may be helpful in partially correcting for an individual’s intelligence level.

Future research with the VOT might suggest if the education correction is necessary given that several studies have failed to find an individual’s number of years of
education to be a significant predictor of VOT scores (Tamkin & Jacobsen, 1984; Walsh et al., 1997; Wenthworth-Rohr et al., 1974). If a future version of the VOT could be created with a more normal distribution, without the ceiling effect, a correction for age resulting in standard scores might be used. This would allow interpretation through comparison with a normative mean and a normal distribution that is often used in neuropsychological testing rather than through cutoff scores. This would fit better with the current understanding of brain injury as specific to different cognitive functions rather than working from an impairment/no impairment model of brain injury. This would also allow comparison of standard scores to an individual’s scores on other measures such as those of general intelligence to determine if the individual’s VOT scores are consistent with their measured level of intellectual ability.

Research Questions

Research Question 1

Is there a significant difference in VOT scores between three groups of individuals consisting of individuals without neurological or psychiatric impairment, individuals with psychiatric impairment, and individuals with known neurological impairment?

Initially, the VOT was thought to be useful in differentially diagnosing brain injury in a psychiatric population. The results of the comparisons of VOT scores of normal individuals, psychiatric patients, and those with neurological impairment demonstrate that differentiation between neurologically impaired and psychiatrically
impaired groups on the basis of the VOT alone is difficult. The only other study to attempt differentiating the three groups involved a much smaller sample and was able to discriminate normal and psychiatrically impaired groups from the neurologically impaired group but not normal and psychiatrically impaired groups from each other (Gerson, 1974). The manual (Hooper, 1983) suggests that the difficulty lies in differentiating psychiatrically impaired from neurologically impaired groups due to the overlap in frequent scores that makes using a cutoff score difficult. In this sample, the normal group scored significantly higher on the VOT than did the psychiatrically impaired and the neurologically impaired groups. The psychiatrically and neurologically impaired groups did not score significantly different from each other. The expectation at the time of the test's inception was that neurologically impaired individuals would score lower than psychiatrically impaired individuals. This study found that in the current sample, both psychiatrically impaired and neurologically impaired individuals scored lower than did the normal group and that the difference between the neurologically impaired and psychiatrically impaired groups was not different at a statistically significant level.

The most that can be said is that, in the current study, the two groups of neurologically impaired and psychiatrically impaired individuals could not be differentiated based on raw or age and education corrected VOT scores. The scores for both neurologically impaired and psychiatrically impaired groups were significantly lower than those in the normal group suggesting that the VOT is measuring some type of impairment that may be common to both neurologically and psychiatrically impaired individuals. It may be that the lower scores on the VOT for these two groups when
compared to the normal group is due to visual-spatial impairment that is common to the
two groups. However, it may also be that the lower scores in each group are due to two
different factors or a combination of different factors. For example, impairment of visual-
spatial skills may lead to lower scores in the neurologically impaired group and some
other factor such as anxiety or poor attention or a combination of anxiety, poor attention
and/or visual-spatial impairment may lead to lower VOT scores in the psychiatrically
impaired group.

Findings from this study do not support the supposition that those with psychiatric
impairment score similar to normals. This was the finding in the study by Gerson (1974).
In fact, the mean scores for the psychiatric group in the current study are lowest of the
three groups on all measures. This finding is similar to that reported in the Manual
(Hooper, 1983) in initial reports by Watson in personal communications with Hooper.
Hooper explained this phenomenon by suggesting that schizophrenia in particular may
represent an organic process that is more similar to neurological impairment than to a
functional disorder. The sample for the current study included 16 cases (17.6 percent) in
which schizophrenia was the primary diagnosis. Another 5 cases (5.5 percent) had
diagnoses of schizoaffective disorder and 10 cases (10.9 percent) had a generic diagnosis
of psychosis. If these groups are thought to be similar to schizophrenia or possible
diagnoses of schizophrenia in the case of the psychosis diagnoses, then this represents 34
percent of the sample. If schizophrenia is more similar to an organic disorder, then this
may lead to the similarity between the mean scores on the VOT in the psychiatric and
neurologically impaired group. Future research may address whether attention or anxiety
contribute to lower scores on the VOT in heterogeneous psychiatric samples or use more
homogeneous psychiatric samples to better understand what factors might lead to lower VOT scores in psychiatrically impaired individuals.

Another possible explanation for the ordering of the group mean VOT scores may be the difference in Full Scale IQ scores between the groups. The psychiatric group had the lowest mean scores for Full Scale, Verbal, and Performance IQ. These measures of general intellectual functioning have been shown to correlate with VOT performance (Boyd, 1981; Gerson, 1974; Greve et al., 2000; Hooper, 1983; Tamkin & Saucer, 1985; Wenthworth-Rohr et al., 1974). In this study, the correlation between Full Scale IQ and VOT raw scores was .52 (p < .01).

One of the weaknesses of the present study is that the three groups were drawn from the practice of a neuropsychologist across three different settings. Most importantly, the neurologically impaired individuals were from an outpatient setting and the psychiatrically impaired individuals were from an inpatient setting. Given the need for inpatient treatment, it is reasonable to assume that the functional level of the psychiatric group was generally lower than that of the other two groups who did not require inpatient care. Drawing samples of the three populations from similar settings, such as outpatient care facilities, may yield different results.

Despite the limitations of the study with respect to the sampling, the results do suggest that both those with psychiatric impairment and with generalized neurological impairment would be expected to score lower on the VOT than individuals without either type of impairment.

In terms of qualitative scoring, despite statistically significant differences between the groups on the mean number of bizarre and isolate responses, the VOT was not clearly
able to discriminate between the groups in a way that can be used in a clinical setting. As was the case in previous studies of the qualitative scoring of the VOT, the significant differences in the number of qualitative errors made does not appear to translate into a discrete cutoff score for use with qualitative scoring on the VOT. Walker (1956) stated that more than one concrete (isolate) response suggested "cortical involvement" or brain injury but admitted that the low cutoff score made the test vulnerable to administration and scoring errors.

In the current study, the psychiatric group was found to have the highest mean number of both bizarre and isolate responses with a mean of .74 isolate responses and .36 bizarre responses. As the mean number of isolate and bizarre responses for all three groups was less than one, qualitative scoring may not provide a practical means of differentiating between groups. Having more than one isolate response, as was suggested as a cutoff score by Walker, was infrequent across all the groups in this study with just over 1 percent of the normal group, 15 percent of the psychiatrically impaired group, and 10 percent of the neurologically impaired group having more than one isolate response. Almost half of the psychiatric group and 33 percent of the neurologically impaired group had at least one isolate response compared to less than 11 percent of the normal group so even one isolate response might suggest impairment as evidenced in this sample. While it would be difficult to differentiate between neurological and psychiatric impairment given that no cutoff score would provide clear differentiation between the two groups, the presence of one or more isolate response in a Hooper VOT protocol does appear to suggest the possibility of psychiatric or neurological impairment and additional assessment and evaluation may be beneficial.
Just over 8 percent of the normal group, 27 percent of the psychiatrically impaired group, and 16 percent of the neurologically impaired group had one or more bizarre responses. However, percentages of individuals with more than one bizarre response were rare across the three groups: 2 percent for the normal group and approximately 5 percent for both the psychiatrically and neurologically impaired groups. Similar to the findings for isolate responses this study suggests that no cutoff score would provide clear differentiation between the psychiatrically and neurologically impaired groups.

Although having one or more bizarre or isolate responses appears to increase the likelihood of impairment, qualitative scoring did not provide a means of differentiating between psychiatrically and neurologically impaired groups. The presence of one or more bizarre responses in a Hooper VOT protocol appears to suggest the possibility of psychiatric or neurological impairment, however, additional assessment and evaluation is necessary to determine the nature of the possible impairment.

Research Question 2

To what extent do individuals with diagnosed neurological impairment, psychiatric impairment and normals differ in their individual item responses on the Hooper Test of Visual Organization?

Chi Square analyses showed that the three groups differed significantly in the frequency of full-credit responses given to 16 of the 30 items in the scale. Most frequently, the three groups scored in the same order as would be expected given the mean group scores for the overall scale. For the 9 items which displayed this pattern, the
psychiatrically impaired group had the lowest percentage of full-credit responses, the
neurologically impaired group had a slightly higher percentage, and the normal group had
the highest percentage of full-credit responses. Two items showed the pattern of the
normal and neurologically impaired groups scoring similarly and the psychiatrically
impaired group scoring lower. On five items, the psychiatrically and neurologically
impaired groups scored similarly and obviously lower than the normal group.

Verma et al. (1993) reported a significant difference (Chi Square) between their
normal and psychiatrically impaired groups in 25 out of 30 items on the VOT. As they
did not report group mean scores, it is difficult to know if this finding is secondary to
larger differences in overall VOT score between the two groups than is found in the
present study. Additionally, they used a larger sample of psychiatrically impaired
individuals (n = 152) and only a slightly smaller normal group (n = 133) with only two
groups in the analysis. Differences in the number of significant differences between the
percentage of full-credit responses between the two studies are not likely solely due to
increased statistical power, however, as the study by Verma et al. generally found large
differences between the two groups on several items that are not seen in the present
sample. For example, for item 4 Verma et al. reported a nearly 30 percentage point
difference in percentage of full-credit responses between the two groups that is not
evident in the present sample (difference = .8 percent).

Differences between the three groups in the frequency of full-credit responses
may be useful in a future revision of the VOT as it may lead to a choice of items based on
their discriminative power. More research would be needed to determine what factors
influence the lower scores in the two impaired groups to determine if items could be
chosen to discriminate between these two groups. If current findings are cross-validated it is useful to know that many of the items in the first half of the test do not display statistically significant differences between these groups in the present sample. In designing a test, items with these properties would be limited to an initial warm-up phase with items in the majority of the test displaying between group differences (Anastasi & Urbina, 1997).

Another possible outcome of the analysis of individual item responses for between group differences is to potentially discover particular responses that are given by one group and less frequently by another group. Item 22 (mouse) provides a good example. The isolate response of “pipe” was absent in the normal sample while “pipe” accounted for 50 percent of the incorrect responses in the psychiatric group and 23 percent in the neurologically impaired group. However, this was the only item in this sample that appeared to generate a specific isolate response that was not observed in the normal group.

Other items did display a tendency to illicit certain types of responses that might suggest group differences. For item 17 (chair), nearly all of the normal individuals who answered with anything other than “chair” answered with “sofa” or some synonym. When a certain incorrect response is this common, it might be notable clinically when an individual gives some other incorrect response. In this case, in the normal group, there were only two individuals who gave other specific incorrect responses, “seat” and “grill” (two responded with “I don’t know”). When an individual responds with something other than a synonym of chair or couch, it may be worthy of note as this seems to be rare in a sample of normal individuals and more common in these two impaired samples. This
situation is true for items 16 (teakettle) and 19 (teapot, cream pitcher) where individuals often answer item 16 with teapot and item 19 with teakettle. Any incorrect answer for these two items that is not some variation of teapot or teakettle may suggest a greater likelihood of impairment.

**Item Analysis**

In addition to examining individual item response differences between the three groups, the data collected for this study yielded interesting findings regarding the individual item characteristics. A Cronbach’s alpha of .80 was observed in this sample. Deleting any one item yielded Cronbach’s alphas ranging from .78 to .81. Merten and Beal (2000) reported a Cronbach’s alpha of .89 in their neurologically impaired sample from a German-speaking population. Seidel (1994) reported a Cronbach’s alpha of .72 in a sample of children age 5 to 11. He compared this measure of internal consistency to the .82 and .78 for college students and adult psychiatric patients reported in the VOT manual and stated that they were not significantly different. The internal consistency for the VOT found in the present study is consistent with what has been reported in other samples.

The corrected item-total correlations for the present study range from .12 to .56. The study by Merten and Beal (2000) reported item-total correlations ranging from .18 to .62. The item-total correlations from the Merten and Beal study were slightly higher for each item than in the present study.

Merten and Beal (2000) provide a good format for an item analysis of the VOT, however, their study had some limitations that are not present in the study reported here.
Specifically, the study by Merten and Beal utilized a neurologically impaired population. In the present study, a normal group, a psychiatrically impaired group and a neurologically impaired group were included in the analysis. When considering item characteristics, it may not be best to use information solely from an impaired population as difficulties with certain items may reflect impairment rather than a problem with specific items.

Several items had interesting characteristics that suggest some revision to the scoring directions. Many of the questions about scoring would likely be addressed by creating clear administration guidelines with examples of common 1-point and ½-point responses. Clearer directions would likely improve the standardization of administration and reliability of scoring between administrators. One example is that for several items, the ½-point response is "animal" when the item is a specific animal such as a dog/sheep. For this item, item 7, the example is given in the manual where the response "animal" is scored ½-point. In the study by Merten and Beal (2000) they report "bear" and polar bear as common ½-point responses, but none of the examples in the manual are analogous to scoring a specific animal that is not a dog or a sheep as ½-point. The manual does say that other responses can be given full credit at the discretion of the administrator if they are synonyms or closely resemble the correct answer. Having a standard answer as to how to score animal responses is important as four items have "animal" as a ½-point response. With item 24 (rabbit), 83 percent of the normal group who gave incorrect answers answered with some type of animal as did 73 and 53 percent of the psychiatrically and neurologically impaired groups. This is an example of how
more clarity in the directions could make scoring more standardized in a case that is likely to be frequently encountered in a clinical environment.

Merten and Beal (2000) suggested that a warm-up or teaching phase be added to the beginning of the test to avoid certain types of incorrect answers to the first item. They theorized that responses noting two things, such as “two fish” were evidence that the directions were not fully comprehended. Interestingly, none of the normal individuals in this current sample gave a response of this type, for example, describing item 1 as two things (e.g. a “bird and a fish” or “two fish”). One individual in the neurologically impaired group and three in the psychiatrically impaired group responded this way. While it might be a good idea to have a teaching item for the VOT, answers of this type may also suggest impairment. If a teaching phase was utilized for the VOT, items 3, 7, and 11 might be used or at least moved closer to the beginning of the test as over 97 percent of the entire sample gave full-credit responses.

It should be noted that percentages of correct responses may vary a great deal between samples (Anastasi & Urbina, 1997). This is evident when comparing the percentages correct reported in this study to those reported by Verma et al. (1993) and Merten and Beal (2000). The percentages correct for some items are quite different. For example, item 3 is among the three easiest in both of these studies as well as in the current study. Item 11 (apple, peach) however, was answered correctly by only 75 percent of the normal and 49 percent of the psychiatric group in the Verma et al. (1993) study. It was answered correctly by 89 percent of the neurologically impaired group in the Merten and Beal (2000) study. In the present study it was answered correctly by 98 percent of the entire sample (100 percent for the normal, 98 for the psychiatrically...
impaired, and 97 for the neurologically impaired groups. It is unclear if the differences noted are related to the cultural differences in the samples or differences in the type and severity of impairment found in each sample.

Merten and Beal (2000) had suggested that item 6 (hammer) was too difficult to be so early in the test with only 26 percent of their sample receiving full-credit. This may suggest a cultural difference from their German sample or be reflective of their impaired sample. In the present study, across the three groups over 96 percent received full-credit. Additionally, Merten and Beal had noted a high percentage of answers related to other tools such as axe (32 percent in their sample). Only 1.2 percent of the current sample answered axe and 2.7 percent answered with any type of tool other than hammer across all three groups.

On item 12 (basket), a common incorrect response is “net.” Merten and Beal (2000) had suggested that this item had poor discriminative power as individuals with high overall VOT scores often provided this incorrect answer. In the current sample, only 3 individuals (2.1 percent) in the normal group and 1 person (1 percent) in the neurologically impaired group responded with any type of net. In the psychiatrically impaired group, however, 8.7 percent responded this way.

Items 16 (teakettle) and 19 (teapot, cream pitcher) are often answered incorrectly with “teapot” for 16 and “teakettle” or “kettle” for 19. Merten and Beal (2000) argue that that teapot for item 16 should get at least half-credit and imply that the two terms may be interchangeable based on changing styles, however, for both items, there are no half-credit responses listed and there is no discussion in the manual regarding how to score these common responses. The assumption that can be made is that with both items
included in the test, the individual is expected to distinguish the difference between a
tea pot and a teakettle, therefore, giving full-credit to item 16 (teakettle) for an answer of
“teapot” and vice versa for item 19 would not seem appropriate. The inclusion of both
item 16 (teakettle) and item 19 (teapot, cream pitcher) in the VOT seems problematic.
Obviously, a revision of the VOT should make some change to these items either with
their deletion or with improved scoring directions.

In this study, 96 percent of the incorrect responses for item 16 (teakettle) were
some variation of “teapot” in the normal group. This percentage was lower in the
psychiatrically and neurologically impaired groups but still high (70 percent and 84
percent respectively). Fewer individuals in the normal group misidentified item 19
(teapot, cream pitcher), 11 percent versus 16 percent who misidentified item 16. Of those
who answered item 19 incorrectly, 69 percent answered with a variation of kettle. Item 19
did display group differences that may make it a useful item, in that only 62 percent of
the psychiatric group and 72 percent of the neurologically impaired group received full-
credit. Additionally, smaller percentages of the impaired groups answered with a form of
“teakettle” (43 percent for the psychiatrically impaired and 39 percent for the
neurologically impaired group). Item 19 had a higher corrected item-total correlation as
well (.33 compared to .17), suggesting that it may be more consistent with what other
items in the test measure. Given the change in the use of the VOT from measuring
generalized impairment to a visual-spatial component it would be logical to have the
items be more consistent. Unless there is a valid reason to retain both items it might be
best to retain item 19 and award half or full credit for the response of “teakettle.”
Item 17 (chair) is interesting in that the common ½ credit response is “sofa.” Merten and Beal (2000) found that in their impaired sample, the average overall VOT score of individuals who answered “sofa” was higher than those that answered “chair.” They suggested that “sofa” was as good a response as “chair.” In the current study, while 36 percent of the normal sample answered with “sofa” or a synonym, 62 percent answered “chair.” In the psychiatrically impaired sample, 53 percent answered “sofa” compared to 36 percent who answered correctly with “chair.” The percentages were nearly even in the neurologically impaired group with 45 percent answering “chair” and 44 percent answering “sofa.”

Similar to the example of item 17, item 21 (flower) was suggested for revision by Merten and Beal (2000) based on their findings that persons who answered “island” or “clover” obtained comparatively high total VOT scores. In the current sample, while the common incorrect answers of palm tree, island, and tree were given by nearly all of the normal individuals that did not receive full-credit, over 90 percent of the normal group received full credit compared to 75 and 65 percent of the neurologically and psychiatrically impaired groups. Items 17 and 21 may be moved to later in the test due to their difficulty, but generally, incorrect answers appear to suggest impairment.

For item 27 (shoe), a common incorrect response is “iron.” Merten and Beal (2000) reported that this response is given by individuals with high mean VOT scores. In the current sample 22 percent of the normal group answered “iron” compared to 42 percent who answered “shoe.” This compares to 15 and 16 percent of the psychiatrically and neurologically impaired groups who answered “iron.” In a future revision, iron might be considered for a ½ point response. It is also notable that the corrected item-total
correlation for this item is low (.16) compared to adjacent items in the scale (.50 for item 26 and .46 for item 28). This may suggest that it is not consistent with the other items in what it is measuring.

In comparing the present study item analysis to other studies there are some interesting differences that may suggest significant differences in the samples. Both the samples used by Verma et al. (1993) and Merten and Beal (2000) were taken from non-English speaking populations (Indian and German respectively). The study by Merten and Beal utilized a neurologically impaired sample while Verma et al. (1993) had both normal and psychiatrically impaired groups. As the present study has samples from all three populations there are some comparisons that can be made to both studies. These comparisons raise the question as to whether differences are due to cultural characteristics or other characteristics of the samples.

When compared to the Verma et al. (1993) study, the percentage correct for the items are generally higher overall in the current study for both the normal and psychiatrically impaired groups. Verma et al. did not report group mean VOT scores or scores on intelligence measures so it is hard to say how the groups compare overall in their VOT performance. However, based on the individual items percentages of correct responses there appear to be significant differences between the samples in the Verma et al. study and the current study. For example, based on the item percentage correct only 5 items had over 90 percent correct in the normal group in the Verma et al. study. In the current study, 20 items have over 90 percent correct in the normal group. In the psychiatrically impaired groups Verma et al. reported 4 items with over 90 percent correct compared to 10 in this study. In the Verma et al. study, on 13 items, less than 30
percent of the psychiatrically impaired group received full credit. In the present study the psychiatrically impaired group score demonstrated less than 50 percent full-credit responses on only 4 items and lowest percent correct in the psychiatrically impaired sample in the present study was 36 percent. Also, visual inspection of the data reported by Verma et al. suggests that both of their sample groups scored lower on the VOT than in the present study. Using a comparison of the normal groups as an example, for 15 items, the percentage correct in the present study was 30 percentage points higher than the percent correct in the study by Verma et al. Interestingly, the first three items showed similar percent correct scores across the two studies. Items 10 (hand) and 13 (scissors) showed the closest comparison between groups outside of the first three items. Specifically, with item 10, the fact that in both studies 98 percent of the sample gave a correct response while for the item before and after there was nearly a 25 percentage point difference in the percentage correct in the Verma et al. study is interesting. Given that item 10 is a hand, and is the most ubiquitous of all the items on the test across cultures this contrast with the neighboring items may be evidence that some of the items display cultural differences in answers.

In the Merten and Beal (2000) study, percentage correct scores are lower for all of the items than they are in the present study. The neurologically impaired sample used by Merten and Beal has a mean VOT score of 18.8 (SD = 6.2) compared to 24.8 (3.92) in the neurologically impaired group in the present study. By way of comparison, 15 items in the neurologically impaired group in the present study had percentage correct scores of 90 percent or above. This compares to 3 items in the study by Merten and Beal. The Merten and Beal sample was older (mean of 51 versus 38 years) and more educated (13.4
versus 12.4 years) than the neurologically impaired sample for this study. These comparisons suggest that the sample in the Merten and Beal study was more impaired than the sample in the current study. While it cannot be clear if the differences in item percentage correct scores between the present study and that by Merten and Beal can be attributed to differences in the level of impairment or in culture, scores on certain items do seem to stand out. Item 5 (baseball or other round ball), for example, was answered correctly by only 50 percent of the German-speaking sample compared to 93 percent of the English-speaking sample in this study. Likewise, item 6 (hammer) was reported to be problematic in the Merten and Beal study as only 26 percent gave correct scores. This compares to 96 percent in the current study. Items 9 (cup), 12 (basket), 25 (block), 28 (key), 29 (ring), and 30 (Broom) all show 30 percentage points fewer percent correct scores in the German-speaking sample. While this 30 percentage point number is arbitrary, the large difference on certain items might suggest cultural differences that are worthy of further study.

Research Question 3

To what extent does the VOT correlate with other neuropsychological measures including the Verbal and Performance IQ measures from the WAIS-R, Raven's Matrices SF, and the SDMT?

The relationship between VOT scores and the scores from the other neuropsychological measures in the study were examined through correlation and hierarchical regression analysis. VOT raw scores were found to be significantly
correlated with all of the neuropsychological measures in the study both in the entire sample and in the three groups considered individually. VOT scores that were corrected for age and number of years of education (corrected VOT scores) were also significantly correlated with all of the neuropsychological measures in the study in the entire sample and in the three groups with the exception of SDMT and Raven Matrices SF scores in the normal group.

Overall, the correlations found in the present study are generally ordered as would be expected from the literature. VOT scores correlated higher with measures of visual abilities such as the Raven Matrices SF and PIQ. Correlations were slightly less for VIQ and Full Scale IQ. Greve et al. (2000) studied 98 patients who had suffered cerebral vascular accidents. IQ was measured using the WAIS-R. They reported a bivariate correlation of .53 between WAIS-R full scale IQ scores and VOT scores. In the Greve et al. study the correlation with the Verbal IQ scores was .38 while the correlation with Performance IQ scores was .65. The correlations reported in Greve et al. are nearly identical to those found in the present study for the neurologically impaired group. In the present study, the correlation between VOT scores and Full Scale IQ was .50 in the neurologically impaired group. Verbal and Performance IQ correlated .34 and .64 with VOT scores in the neurologically impaired group. The correlations with SDMT are relatively high, specifically in the neurologically impaired group. This may suggest that both VOT and SDMT scores are affected by generalized neurological impairment.

Hierarchical regression analysis showed that when the entire sample was considered, the full model with age, education, SDMT written scores, Performance IQ,
and Raven Matrices explained 47 percent of the variance in VOT scores. Only education was not found to be a significant predictor of VOT scores in this full model.

The finding that SDMT is a significant predictor of VOT scores when the other predictor variables are in the model may suggest that the VOT is sensitive to generalized impairment of neurological functioning. Spreen and Strauss (1998) reported that the SDMT is sensitive to a variety of neurological deficits and is suggested to measure processing speed and the scanning and tracking aspects of attention. The SDMT is also said to be related to "real world functioning." Given the relationship between the SDMT and the VOT and the association between the SDMT and real world functioning, it may be no surprise that the VOT was originally thought to be a good screen for neurological impairment. This sensitivity to general impairment and aspects of attention is a possible explanation for why both psychiatrically and neurologically impaired individuals scored more poorly on the VOT. This may be particularly relevant in the present study in understanding the ordering of the groups with both psychiatrically and neurologically impaired groups scoring significantly lower than the normal group and the psychiatrically impaired group scoring slightly lower than the neurologically impaired group (not statistically significant). The relatively similar and impaired performance of psychiatrically impaired and neurologically impaired samples may also be partially explained by the difference in setting as it might be assumed that the psychiatric inpatients were more significantly functionally impaired, given their inpatient status, than the outpatient neurologically impaired individuals but the neurologically impaired individuals might be expected to show more impairment of visual-spatial abilities.
Additional research and analysis is needed to determine the factors that influence the lower VOT scores in these two groups.

The significance of WAIS-R Performance IQ scores in the final model is consistent with several studies that show that Perceptual Organization, one of the sub-factors of Performance IQ, is a significant predictor of VOT scores (Greve et al., 2000; Paolo et al., 1996; Ricker & Axelrod, 1995). The finding that the Raven Matrices SF contributes significantly to the prediction of VOT scores with Performance IQ in the model may be an important addition to the understanding of VOT performance. Alderton and Larson (as cited in Spreen & Strauss 1998) as describing Raven’s Matrices as the “quintessential test of inductive reasoning” (p. 83). It is also thought to be a measure of non-verbal intelligence or fluid intelligence. Spreen and Strauss state that Raven’s Matrices scores are related to the Block Design Subtest of the WAIS-R most strongly compared to the other subtests. The predictive value of the Raven Matrices SF scores in the present study suggests that over and above the general performance abilities and visual-spatial skills that are related to VOT scores there may be a reasoning or non-verbal intelligence component. Additional research would be helpful in determining the relationship between the subtests of the WAIS-R Performance IQ and Raven Matrices SF scores in relation to VOT performance.

Limitations

As with any study, there are several limitations that can be identified in the present study that need to be taken into account when considering the findings and in conducting future research with the VOT. This study was archival so the procedures
regarding screening of the participants and the assignment to the three groups was based on the information available in the archival records. This information may not have been complete and important information regarding impairments or other factors may have been unavailable.

Another limitation of this study is that the three groups were drawn from different settings. The differences in settings and their implied differences in functionality at the time of testing may have also had an effect on group differences that is more than a function of the different diagnoses. The different settings may suggest group differences in overall functional level that can affect test scores and this is a factor that might be controlled for in future studies. The difference in setting also led to another limitation in the interpretation of the data from this study. Since protocols for the psychiatrically impaired group were obtained from psychiatric inpatients, it can be assumed that some of the individuals were taking psychotropic medication that may have changed their mental status. It is possible that the individuals in this group may have scored differently on the VOT if they were not taking medication. Additionally, the present study did not control for IQ when assessing group differences. Given the significant relationship between intelligence and VOT scores and group differences in mean IQ scores, it may be important in future studies to attempt to control for intelligence in assessing group differences in VOT scores. There was some tolerance of psychiatric diagnoses in the neurologically impaired group as neurological impairments can often lead to psychiatric impairments. The limited occurrence of psychiatric impairment in the neurologically impaired group lessens the ability to assume that scores in the neurologically impaired group were due solely to their neurological impairment as some variance in their scores.
may be attributed to the effects of their psychiatric issues. Future studies might select individuals without psychiatric impairment for the neurologically impaired group.

When assessing the utility of qualitative scoring, this study is limited in that there was only one qualitative scorer, so there was no check on the reliability of the qualitative scoring. This type of qualitative scoring is by nature subjective and the guidelines in the manual are understandably limited. When assessing the qualitative scoring of the VOT it would be helpful to have more than one qualitative scorer with more detailed guidelines for consistent scoring and to also consider assessing the inter-rater reliability.

Directions for Future Research

One of the most important areas for future research and use of the VOT is a comprehensive re-organization of the test. The VOT has been a quick and useful instrument for assessing visual-spatial organization and reasoning, but its utility is somewhat limited by inadequate scoring instructions and some items that may be outdated or confusing. At very least, deleting either item 16 (teakettle) or 19 (teapot) and updating item 8 (truck) with a more modern picture increase the validity of the test. New research that would include a large normative sample might add further support to the reordering of items based on percentage correct and possibly evaluate new items that could be added or used to replace other less discriminating items would be valuable. By reordering the items and potentially replacing some of the easier items with more difficult ones, a more normal distribution of scores might be obtained that would allow for interpretation through standard scores, which would be more consistent with how neuropsychological tests are typically used today. This would also be helpful in the use of
the VOT in assessing a particular cognitive domain versus generalized cognitive impairment since the VOT could then be more easily compared to other neuropsychological test performances.

Comparison of the item analysis in the current study to those completed using samples from German and Indian populations suggests that there may be cultural differences related to individual items and possibly the test as a whole. Future research comparing samples from different cultures may suggest if certain items are sensitive to cultural differences or if different norms and cutoff scores are needed for individual cultures.

Furthermore, research directed toward a better understanding of the facets of cognition that are tapped by VOT scores with additional analysis of visual-spatial components and visual reasoning in relation to the VOT would be helpful. Using regression analysis with the measures that have currently been shown to predict VOT scores and including other measures of visual-spatial skills and reasoning would help further define the cognitive abilities that are related to VOT scores.

Further understanding of VOT performance as it relates to specific brain functions and lateralization might also be obtained through the use of modern functional imaging. The small size of the stimulus booklet and the ability to answer verbally may make the VOT a useful tool for use with functional Magnetic Resonance Imaging protocols. This may provide greater understanding of the VOT in addition to providing researchers and clinicians with a measure to be utilized with new technologies to better understand impairment of certain cognitive functions.
Association of neuropsychological test score in general and VOT scores in particular with functional impairment and rehabilitation is an important area of research in neuropsychology. With the advent of improved technology for the diagnosis of neurological impairment it is important that neuropsychological measures are able to address issues of functionality and rehabilitation rather than diagnosis alone. While the initial use of the VOT was to determine neurological impairment, the VOT's future utility may be more related to understanding specific deficits in cognitive functioning and appropriate adaptation of rehabilitation strategies.
REFERENCES


Hooper Visual Organization Test in an inpatient substance abuse population.

*Archives of Clinical Neuropsychology, 3(2), 181-187.*


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Appendix

Human Subjects Institutional Review Board Approval Letter
Date: November 14, 2003

To: Patrick Munley, Principal Investigator
    Ennis Berker, Co-Principal Investigator
    Michael DeVries, Student Investigator for dissertation

From: Mary Lagerwey, Ph.D., Chair

Re: HSIRB Project Number: 03-10-01

This letter will serve as confirmation that your research project entitled “Evaluating the Hooper Visual Organization Test in Neurologically Impaired, Psychiatrically Impaired, and Normal Controls” has been approved under the expedited category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: November 14, 2004