Summer 2007

Examining the ability of the Halstead -Reitan Battery and the Wide Range Achievement Test to utilize severity in discriminating among Alzheimer's dementia patients

Gina Gibson-Beverly
Louisiana Tech University

EXAMINING THE ABILITY OF THE HALSTEAD-REITAN BATTERY AND THE WIDE RANGE ACHIEVEMENT TEST TO UTILIZE SEVERITY IN DISCRIMINATING AMONG ALZHEIMER'S DEMENTIA PATIENTS

by

Gina Gibson-Beverly, MS

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

COLLEGE OF EDUCATION
LOUISIANA TECH UNIVERSITY

August 2007

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author Ima Gibson-Beverly

Date 7-3-07

GS Form 14
(5/03)

LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

July 20, 2006
Date

We hereby recommend that the dissertation prepared under our supervision by Gina Gibson-Beverly

entitled EXAMINING THE ABILITY OF THE HALSTEAD-REITAN BATTERY AND THE WIDE RANGE ACHIEVEMENT TO UTILIZE SEVERITY IN DISCRIMINATING AMONG ALZHEIMER'S DEMENTIA PATIENTS

be accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Supervisor of Dissertation Research

Head of Department

Psychology and Behavioral Sciences
Department

Recommendation concurred in:

Advisory Committee

Approved:

Director of Graduate Studies

Dean of the College

Approved:

Dean of the Graduate School

GS Form 13a
(6/07)

Abstract

Measuring change in cognitive status is essential for the diagnosis, prognosis, and treatment of brain dysfunction. Psychological abilities are differentially affected by brain dysfunction severity, as some abilities are more vulnerable to brain dysfunction than others. Neuropsychological assessments can be viewed as a continuum of "hold" and "don't hold" tests. "Hold" tests assess abilities that remain stable in spite of brain dysfunction, while "don't hold" tests measure skills that are significantly compromised by brain impairment. The present study ranks the Halstead-Reitan Battery (HRB) and the Wide Range Achievement Test (WRAT) subtests based on their ability to discriminate between two levels of severity in an Alzheimer's Disease (AD) sample. Subtest rankings showed pronounced sex differences, suggesting that test performance is influenced by severity and sex. There appears to be a distinct neuropsychological profile associated with AD. Overall, results indicate that the WRAT is clearly a moderate "don't hold" test implying that the WRAT cannot be a good estimator of premorbid functioning, as it is moderately related to severity. Use of the WRAT to estimate previous abilities in AD patients would be unwise and would likely underestimate premorbid levels. The HRB is a combination of "hold" and "don't hold" subtests, which are directly related to the physiology of the disease process.

Table of Contents

List of Tables

Chapter 1: Introduction

Neuropsychology uses assessment as an objective method to elucidate brain-behavior relationships. Documenting disease progression yields information about etiology, prognosis, and treatment. Identifying deficits related to injury and disease emphasizes areas of decline, which informs forensic and rehabilitation decisions. Knowledge of impairments leads to better understanding of the general disease process and its effects on each individual. To examine the impact of disease on the individual, present psychological deficits must be monitored for changes.

Longitudinal measurement of cognition creates a dynamic picture of how physiological processes affect psychological functioning. Clinicians are often interested in determining whether current functioning represents a change from previous levels. Documenting when a patient's performance has changed is necessary for accurate diagnosis, tracking disease progression, and determining whether observed impairments are premorbid or post-injury. Change is the defining feature of some diseases such as dementia, which can only be diagnosed when present abilities represent a decline from earlier functioning, and this change is greater than would be expected with normal aging. Neuropsychological tests measure a broad range of cognitive abilities, each with a unique sensitivity to change. Certain

cognitive skills, such as reading and vocabulary, are believed to be relatively resistant to brain impairment (Babcock, 1930; Crawford, 1992; Yates, 1956).

Instruments that assess these abilities show similar performance despite changes in severity of brain impairment. This type of measure is considered to be a "hold" test, as the abilities it assesses appear relatively resistant to brain dysfunction. Other abilities provide indications of deficits in psychological functioning because as a disease progresses patients perform more poorly on tests that assess these skills. These measures are considered to be "don't hold" tests, in that the skills they assess are compromised at some point during the disease process (Wechsler, 1958).

One way to measure change is by focusing on illness severity. Despite the importance of examining severity as an indication of change, there is a lack of research available. The Halstead-Reitan Battery (HRB) is a series of neuropsychological tests that assesses a broad range of abilities (Reitan & Wolfson, 1993). These skills are differentially affected by brain dysfunction, with some abilities more resilient than others. Therefore, the HRB can be considered to be a combination of "hold" and "don't hold" tests.

Research has been conflicted regarding the point at which illness severity negatively impacts specific cognitive skills and test performance. In general, it appears that psychological abilities eventually succumb to the disease process, but some at lower levels of severity than others. Certain abilities, such as memory and reasoning, are significantly impacted early in the course of the disease (Barth & Macciocchi, 1986; Lezak, 1995). However, other abilities, such as reading or motor functioning, are not affected until later (Barth & Macciocchi, 1986; Bigler, Steinman, & Newton, 1981; Hom,

1992; Storrie & Doerr, 1980). Research is inconclusive regarding the point at which reading ability is significantly compromised after brain dysfunction (Smith-Seemiller, Franzen, Burgess, & Prieto, 1997). The skills affected and magnitude of their decline is related to the premorbid nature of the ability and the specific disease process, with each disease resulting in a distinct pattern of change.

Neuropsychological testing highlights patterns of change that result from disease specific physiological processes. Russell and Polakoff (1993) found that Alzheimer's Dementia (AD) patients were less impaired on motor tests than other types of dementia patients. Compared to vascular dementia patients, individuals with AD performed worse on tasks involving sequencing and cognitive flexibility, experienced more language dysfunction, and had difficulty with non-verbal memory (Baillon et al., 2003). Johnstone, Hogg, Schopp, Kapila, and Edwards (2002) found a specific deficit pattern for AD that is different from traumatic brain injury (TBI) and systemic lupus erythematosus, providing evidence that each disease yields a unique neuropsychological profile.

Research shows that the HRB assesses a range of cognitive skills that differ in their sensitivity to brain impairment, with some being compromised and others being relatively unaffected. Studies (Hom, 1992; Storrie & Doerr, 1980) have found that AD patients perform worse on the Category Test than on other HRB subtests. This suggests that abstraction and reasoning skills are highly related to deficits from the disease process. Therefore, the Category Test can be considered a "don't hold" test for this population, as performance of AD patients was significantly poorer than matched controls. Other skills, such as motor abilities, seem to be relatively preserved in AD until advanced disease stages (Barth & Macciocchi, 1986; Bigler et al., 1981; Hom, 1992;

Russell & Polakoff, 1993). This suggests that motor abilities are not significantly impaired in AD patients and motor tests can be considered "hold" tests, as performance does not significantly differ from controls.

Another test believed to be a strong "hold" test is the Wide Range Achievement Test (WRAT) – Reading subtest. As reading has been considered to be an ability that remains intact despite brain damage, it has been used to estimate premorbid functioning. An estimate is needed to determine whether current functioning represents a decline from previous levels when previous functioning is unknown (Wechsler, 1958). However, the research regarding the relationship between reading and severity calls into question its status as a "hold" test, particularly at severe levels of impairment.

There are several approaches to estimating premorbid functioning, including present ability measures, regression formulas, and use of historical information. The present ability approach (i.e. a "hold" tests) measures abilities that remain intact throughout the disease process. Regression methods and historical information use variables that are unrelated to brain impairment Barona, Reynolds, & Chastain, 1984). Each method provides an estimate of premorbid functioning that serves as a comparison to current test performance.

The present ability approach utilizes certain skills, such as reading, that appear to "hold" despite brain impairment. However, the resilience of reading ability has been questioned, and research suggests that even the most resistant skills are compromised at severe stages of impairment (Fromm, Holland, Nebes, & Oakley, 1991; Paolo, Troster, Ryan, & Koller, 1997; Stebbins, Gilley, Wilson, Bernard, & Fox 1990a; Stebbins, Wilson, Gilley, Bernard, & Fox, 1990b). Evidence conflicts regarding when abilities are

impacted by brain dysfunction and premorbid estimates are invalid. Additionally, reading approaches are influenced by external factors, including intellectual range, stage of recovery, and severity (Fromm et al., 1991; Griffin, Mindt, Rankin, Ritchie, & Scott, 2002; Johnstone, Callahan, Kapila, & Bouman, 1996; Johnstone & Wilhelm, 1996; Wiens, Bryan, & Crossen, 1993).

Methods for estimation must be applicable to both normal and clinical populations. However, approaches that are efficacious for normal, control groups may not be accurate for individuals with diseases of the brain. The WRAT has been found to yield accurate estimates of premorbid ability in both healthy and neurologically impaired populations (Griffin et al., 2002; Johnstone et al., 1996; Orme, Johnstone, Hanks, & Novack, 2004; Wiens et al., 1993). However, many studies examining the efficacy of reading as an estimate of premorbid intelligence have utilized the National Adult Reading Test (NART), which unlike the WRAT, consists of words with irregular pronunciation. It is unclear how differences in these tests affect the accuracy of estimation, particularly in clinical populations.

Some studies (Crawford, Parker, & Besson, 1988; Maddrey, Cullum, Weiner, & Filley, 1996; Nebes, Martin, & Horn, 1984; O'Carroll & Gilleard, 1986) have found that the NART is relatively insensitive to brain impairment even at increasing levels of severity, suggesting that reading ability is preserved despite advanced disease. Other research has identified problems using the NART at higher levels of severity (Cockburn, Keene, Hope, & Smith, 2000; Fromm et al., 1991; Stebbins, 1990a; Stebbins, 1990b). Overall, these studies suggest that reading is a "hold" test, but may not produce accurate

estimation at higher levels of severity because it is negatively affected by severity at some point during the disease process.

Statement of the Problem

Documenting change in neuropsychological functioning is necessary to understand the general disease process and determine its unique effect on each individual. Changes in functioning guide decisions regarding diagnosis, prognosis, and treatment. The identification of change facilitates the identification of strengths and weaknesses, which lead to the development of rehabilitation goals and effective coping strategies. Determining the impact of change on specific psychological abilities is especially important when assessing individuals with chronic brain illnesses that result in continual deterioration of psychological abilities. Despite the importance of monitoring changes in cognitive status, there are significant gaps in the literature. Cognitive abilities and the assessments that measure them are differentially affected by brain dysfunction. "Hold" tests measure abilities relatively unaffected by brain dysfunction where test performance is unrelated to illness severity. "Don't hold" tests assess abilities that are compromised during the disease process and have test scores highly correlated with severity.

The HRB is an excellent example of a well-researched neuropsychological battery that examines a range of skills (Reitan & Wolfson, 1993). It is a valuable instrument used to understand neuropsychological functioning and overall brain-behavior relationships. It can be considered a combination of both "hold" and "don't hold" tests. Given this, certain subtests will be correlated with measures of severity, while others will be unrelated to severity. However, there is a dearth of research examining how HRB subtests respond to

change in illness severity within a clinical population. This study will determine the sensitivity of HRB subtests to changes in severity within an AD sample.

The WRAT is a widely available, easily administered test that has a subtest examining reading ability. Previously, reading has been considered an ability that is relatively resistant to brain dysfunction and has been used to estimate premorbid functioning in clinical and normal populations (Babcock, 1930; Crawford, 1992; Yates, 1956). However, evidence conflicts regarding usefulness of the WRAT as a "hold" test as functioning changes, or severity increases. Some studies suggest that reading is not a "hold" test as it is compromised by the disease process, particularly at advanced levels of severity (Cockburn et al., 2000; Fromm et al., 1991; Stebbins et al., 1990a; Stebbins et al., 1990b). The present study will determine the sensitivity of the WRAT to changes in severity within a clinical AD sample.

Overall, this study will address gaps in the literature regarding the sensitivity of some commonly used neuropsychological assessments to changes in severity. The HRB and WRAT subtests will be examined to determine their relationship to a measure of severity. This will allow conclusions regarding the "hold/ don't hold" status of subtests. The subtests highly correlated with severity will be considered "don't hold" subtests, while those weakly correlated with severity will be considered "hold" tests. Performance of two severity groups will be analyzed to determine the relationship between each subtest and severity, which will provide an indirect measure of change.

Justification

Although it is difficult to determine changes in psychological functioning, it is necessary to identify deficits and track disease progression. To understand specific areas

of deterioration, an assessment of a range of cognitive abilities is necessary. The differential impact of brain dysfunction results in certain cognitive skills affected more than others. Therefore, the measures that assess these skills fall on a continuum of "hold" and "don't hold" tests. More research is needed to examine the relationship between neuropsychology subtests and measures of severity to determine whether skills change as severity increases. This has implications for their status as "hold" or "don't hold" tests, which likely depends on the specific illness or injury.

In the present study, change will be assessed indirectly by examining the relationship between WRAT and HRB subtests and Mini Mental State Exam (MMSE) score, a measure of severity. This will be accomplished by computing the correlation for each subtest and severity score. The correlation between WRAT and HRB subtests and MMSE will be used to rank each subtest with regard to its status as a "hold" or "don't hold" test. The WRAT was selected because previous research has labeled it a good "hold" test, but studies have found it is compromised at high levels of severity (Fromm et al., 1991; Paolo et al., 1997; Stebbins et al., 1990a; Stebbins et al., 1990b). Determining the relationship between the WRAT and MMSE will yield information regarding its status as a strong "hold" test in this clinical AD sample, which will allow conclusions to be drawn regarding its accurate estimation of premorbid functioning.

The HRB is one of the most widely researched neuropsychological batteries available. However, more investigation is needed to determine how each subtest responds to changes in severity within a clinical population. The HRB subtests examine a myriad of cognitive skills, some which are more impacted by brain dysfunction than others. The result is a battery consisting of a combination of "hold" and "don't hold" subtests, which

vary in sensitivity to brain impairment. Overall, this yields information about how physiological processes associated with AD affect psychological abilities.

Literature Review

Historical Methods of Quantifying Change

Early attempts at determining change in neuropsychological functioning largely grew out of the necessity of premorbid estimation. Change was monitored by comparing current functioning to previous functioning. However, previous abilities were often unknown and needed to be estimated. Methods of documenting change focused on ways that provided accurate comparisons of functioning or could distinguish brain damaged individuals from normal individuals.

Babcock (1930) made one of the first attempts to quantify change by observing that vocabulary measures appeared to be less affected by brain dysfunction than other cognitive abilities. This was one of the first applications using the idea of a "hold" test. Since then there has been a plethora of research investigating various methods to determine their utility to establish change in cognitive functioning. Early research advocated the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) as a valid measurement of premorbid function, while others proposed the best estimate to be an average of Vocabulary and Picture Completion, or the higher performance of the two (McFie, 1975; Yates, 1956; Yuspeh, Vanderploeg, & Kershaw, 1998). Babcock's research led to the development of an index of deterioration that included tests of memory, learning, and motor abilities. She combined them into a scale and compared its scores to vocabulary test performance, which was expected to be resistant to the effects

of brain damage. The result was a mental deterioration index using the difference score method.

The difference score approach relies on the differential impact of brain damage. The assumption is that injury and disease affect areas of the brain differently, with some abilities being more vulnerable than others. Determining premorbid neuropsychological functioning and comparing it to current test performance facilitates the identification of specific areas of brain impairment. This approach to estimation has led researchers to investigate tests that assess abilities that are resistant to brain damage (e.g., "hold" tests) and can be compared to current measures of functioning. However, problems arise when patients have advanced levels of severity. Given this, other approaches, such as those using multiple regression, may be better choices for prediction (Wittenborn, 1951).

There are many examples of using difference score methodology to identify brain dysfunction. Shipley (1940) compared scores on abstract thinking and vocabulary tests as a measure of mental deterioration under the assumption that vocabulary skills remain intact after brain damage and abstract reasoning abilities do not. Hewson (1949) developed ratios using the Wechsler-Bellevue Intelligence Scale (Wechsler, 1958) subtests and the Substitution Test (Kaufmann, 1968) to differentiate individuals with psychoneurosis and cerebral pathology from normal subjects. Hunt (1943) utilized tests of vocabulary; learning and retention; and speed and efficiency to assess for organic brain damage. These approaches had varying degrees of success and were all dependent on the assumption that vocabulary abilities are unaffected by changes in cognitive status.

The WAIS Deterioration Quotient is another application of the differential score approach and one of the first attempts at estimating premorbid functioning using a "hold/

don't hold" methodology. In 1944, Wechsler presented the quotient, which compared "hold" and "don't hold" subtests of the Wechsler-Bellevue Intelligence Scale (Wechsler, 1958). It was based on the belief that certain intelligence subtests are impacted by brain impairment, whereas others are not. To calculate the index, the age scale scores for the "don't hold" tests are subtracted from the sum of scaled scores for the "hold" tests. The "hold" tests were Vocabulary, Information, Object Assembly, and Picture Completion. The "don't hold" tests were Digit Symbol, Block Design, Similarities, and Digit Span (Franzen, Burgess, & Smith-Seemiller, 1997; Wechsler, 1958). In general, this method was believed to provide information concerning current and premorbid cognitive.

However, significant problems have been found with traditional WAIS "hold/ don't hold" approaches. "Hold/ don't hold" status is influenced by type of impairment, which results in inconsistencies with regard to brain dysfunction. It has also been shown that brain impairment can influence performance on all WAIS subtests, making the assumption of "hold/ don't hold" tests invalid (Russell, 1972). Overall, subtest performance changes depending on location, pervasiveness, and chronicity of brain injury (Klesges, Wilkening, & Golden, 1981). Researchers have been unable to consistently estimate premorbid abilities with the WAIS using this method and the index is no longer considered valid (Klesges & Troster, 1987). However, the general premise of "hold/ don't hold" tests appears useful and continues using other tests.

Another approach to documenting brain dysfunction is the specific hypothesis method, which identifies some ability that is lacking in those with brain impairment. Tests that use the specific hypothesis approach are simple screening procedures, which assess specific skills that are indicative of brain damage. When this identified ability is

not present, it suggests some type of cerebral dysfunction. This methodology is prevalent throughout neuropsychology and has been used primarily to identify individuals with aphasia or organic brain damage. Tests that incorporate this approach include Gallese (1956), who examined the negative aftereffect of the Spiral Test for evidence that it differentiates between normals and those with cortical involvement. The spiral test used two disks with black spirals, one that was clockwise while the other was counter clockwise. After being shown the first disk, the second disk was presented. The participants were then asked if the disks seem to change, alluding to the aftereffect of the second disk. Although he found support for the use of the test, he reports significant limitations that impact its usefulness. These limitations include the inability to differentiate between types of brain impairment and that it is insufficient for diagnosis of cortical involvement. Focusing on visual-motor abilities, Graham and Kendall (1946) used the Memory-For-Designs Test to capitalize on the lack of these abilities in individuals with brain impairment. They found that this skill is often absent in those with impairments, but is rarely deficient in normals.

Canter (1966) also utilized the assessment of visual-motor abilities to detect brain damage. He developed a technique called the Background Interference Procedure (BIP) used in conjunction with the Bender-Gestalt Test. It increased the test's ability to identify brain impairment by employing paper with confusing, intersecting lines. One advantage to this approach is that the subject's performance without BIP is used as the comparison instead of normative data, which enables the patient to serve as his/her own comparison. He found evidence that BIP has strong validity as a screening test for brain damage. Focusing on aphasia, DeRenzi and Vignolo (1962) devised the Token Test, which uses

basic commands to assess receptive language functioning and is simple enough not to be influenced by intellectual level. This test is still used as a measure of language and an individual's ability to follow directions.

An alternative method of identifying change in cognitive status is the best performance method. Based on the deficit measurement model, this approach advocates that an individual's highest test score, non-scoreable behavior, or premorbid achievement is the best estimate of previous ability (Franzen et al., 1997; Lezak, 1995). Significant discrepancies between best performance and other cognitive functions are suggestive of disease or impairment (Lezak, 1995). This method has a number of important underlying assumptions that should be noted. First, an individual's overall cognitive skills can be captured in one performance score. That is, one test or behavior can accurately represent a person's general cognitive development. A second assumption is that behavioral observations and historical information can be useful when estimating premorbid ability (Lezak, 1995). Limitations of this approach include the use of different tests that have divergent psychometric properties, as well as general problems regarding assessing the elderly since some neuropsychological abilities show decline even in normal aging when no disease is present (Franzen et al., 1997; Lezak, 1995). Overall, the best performance method has significant problems and has been shown to consistently overestimate premorbid functioning (Vanderploeg, Schinka, & Axelrod, 1996).

The pathognomonic sign approach is another method to differentiate brain-damaged subjects from normals. The premise is that specific signs point to the existence of brain dysfunction because they are rarely seen in normal populations. Given this, performances are judged on the presence of the signs instead of the speed of performance

or accuracy of task execution (Reitan & Wolfson, 1993). The HRB utilizes the pathognomonic sign approach in the Aphasia Screening Test to document brain impairment.

Piotrowski (1937) used ten signs on the Rorschach inkblot test to detect organic brain damage. These signs were thought to be indicative of changes in higher functioning, which are associated with brain impairment. Although no single sign suggests dysfunction, the cumulative effect of several signs indicates abnormality. Specifically, the presence of 5 out of 10 signs points to an organic disease process. These signs successfully differentiated between individuals with organic brain dysfunction, conversion hysteria, and a non-organic control group. An advantage of this method is that the presence of a sign reliably indicates cerebral involvement. However, it is limited by the ability of the sign to detect brain damage. Disadvantages are false-negative responses, where individuals with brain damage perform the sign task normally and their dysfunction remains undetected because they perform similar to normals (Reitan & Wolfson, 1993).

In contrast to traditional psychological evaluation, neuropsychological assessment operates under a deficit measurement paradigm. Using this approach, brain dysfunction is determined by comparing patients' test scores to individuals from normal populations, which facilitates the identification of deficits and preserved abilities (Johnstone et al., 1996; Lezak, 1995). However, there are significant problems with this approach. It is problematic to assume that an individual's performance outside of normal ranges indicates brain impairment. Also, normative tables do not consider life history, which omits helpful information that greatly adds to predictability (Snyder & Nussbuam, 1998).

Additionally, the only skills appropriate for this type of measurement are those that are well within the capability of all adults and independent of other variables such as age, gender, and education (Lezak, 1995). An individual's level of previous functioning must be considered. For example, if someone who was previously functioning at a superior level is now scoring in the average range, this is a significant deficit for them even though they are still within normal limits (Lezak, 1995). These limitations emphasize the need for more accurate measures of premorbid functioning that do not rely on comparison with normative data as the sole indicator of brain dysfunction. The HRB addresses this limitation and includes normative data for those with and without brain dysfunction.

In situations where information about previous cognitive functioning is needed, premorbid direct measurement of abilities is ideal. Test scores from earlier psychological evaluations or academic endeavors provide an excellent way to document previous cognitive abilities. However, this approach is limited by the accuracy of the data, accessibility, and knowledge of the patient's past (Lezak, 1976). History based methods that use school records or vocational information as estimates are sound methodology, but are limited by time and accessibility. If pre-illness test scores are not available, clinicians are asked to estimate previous abilities using indirect measures of functioning. The challenge is to find approaches of estimation that use indirect measurements and are accurate in both clinical and normal populations.

In the absence of direct measures, the use of clinical judgment is a common method to estimate previous abilities. Here, a clinician uses the patient's personal history and interview information to establish levels of premorbid functioning. Many

neuropsychologists use clinical interviews or historical data to estimate premorbid abilities despite available research on more objective methods, such as "hold" tests and regression equations (Smith-Seemiller et al., 1997). Limitations of this approach include low levels of inter-rater reliability and problems of validity associated with subjective methodology (Barona et al., 1984).

Kareken and Williams (1994) found that clinicians believe that the correlation between an individual's intelligence quotient (IQ) and demographic variables are stronger than research indicates. They also tend to consider primarily one variable, education, as the basis for their estimates and place more confidence in their estimates than is warranted. Although there are concerns about the importance of educational level in estimation, it has been found to correlate with intelligence in the .5-.7 range and is the best single predictor of intellectual performance. These results suggest that human limitations impact the ability to estimate premorbid functioning using only clinical judgment (Matarazzo, 1972; Snyder & Nussbaum, 1998). Due to these problems, "hold" tests or regression equations that are unaffected by brain dysfunction are a more objective, and often a more accurate alternative.

"Hold" tests, are used to determine neuropsychological change by assessing a current ability believed to remain intact despite brain dysfunction. In order to be an effective estimator, a "hold" test must have satisfactory reliability, be correlated with IQ, and be resistant to the effect of biological and psychiatric disorders. In general, present ability measures assume that performance on one measure of cognition enables estimation of performance on another, in that individuals function at similar levels in all

areas of brain behavior (Bright, Jaldow, & Kopelman, 2002; Crawford, 1992; Frazen et al., 1997; Snyder & Nussbaum, 1998).

The relationship between reading ability and IQ provides evidence that reading ability is related to general intelligence and can be maintained despite cognitive dysfunction in other areas (Nelson & McKenna, 1975; Nelson & O'Connell, 1978). Single word reading is believed to be a cognitive ability that is one of the most resilient to brain damage (Blair & Spreen, 1989; Crawford, 1992). Additionally, reading aloud appears to be unaffected even at very severe impairment levels, suggesting it is an appropriate "hold" test (Cummings, Houlihan, & Hill, 1986). Assessment instruments, such as the WRAT and NART, have demonstrated efficacy in estimating previous cognitive functioning and have been shown to be an accurate estimator of premorbid abilities in both clinical and normal populations (Griffin et al., 2002; Johnstone et al., 1996; Orme et al., 2004; Wiens et al., 1993). However, there is conflicting evidence as to whether it remains unaffected as severity increases.

Reading as a "Hold" Ability

Although reading has previously been considered to hold despite brain impairment, the research is conflicted regarding its relationship to severity. Researchers have found that reading ability is affected at later AD disease stages, suggesting that reading measures are inaccurate premorbid estimates, especially at higher levels of severity (Fromm et al., 1991; O'Carroll et al., 1995; Paolo et al., 1997; Stebbins et al., 1990a; Stebbins et al., 1990b). However, other investigators have performed comparison studies that show no differences between demented patients and matched controls on reading performance, suggesting that reading is unrelated to severity (Crawford et al.,

1988; Nebes et al., 1984; O'Carroll & Gilleard, 1986). Overall, the results are largely inconclusive but suggest that disease processes eventually impact reading ability. Reading measures are likely sensitive to the effects of dementia, particularly at higher levels of impairment.

Much of the research on reading ability has been conducted with the NART, which was developed as a measure of premorbid functioning. It is similar to the WRAT-Reading test, but uses single words with irregular pronunciation to assess ability. This requires the subject to be familiar with the word in order to get it correct. It is postulated that this makes it more sensitive to previous abilities than tests using regular pronunciation of words (Nelson & O'Connell, 1978).

Paolo et al. (1997) used an AD sample, and patients with greater severity had lower IQ estimates on the NART. This suggests that the NART is sensitive to dementia severity. Maddrey et al. (1996) utilized the NART across several levels of dementia and found that although reading is relatively stable against cognitive decline, it does show signs of impairment at advanced levels of severity. Using a sample of AD patients, Cockburn et al. (2000) determined that performance on the NART is not as resilient to the effects of the disease as was initially thought. The results indicate that scores declined over time as a function of Mini Mental State Exam (MMSE) scores, independently of age, education, and dementia onset. Their conclusion is that the NART is not as effective at later stages of disease and is an unreliable estimator in individuals with severe impairment. The results suggest the cautious use of the NART to estimate premorbid functioning, particularly at more advanced levels of severity.

Stebbins et al. (1990b) compared AD patients to controls and also concluded that the test is not insensitive to dementia. They found that the NART underestimated IQ in those with mild AD, and it grossly underestimated IQ in those with moderate to severe dementia. However, education seems to moderate the impact of dementia. Those with mild dementia and some college education had more accurate estimates than those without such education, suggesting that the NART should be used in combination with a demographic equation, especially with higher levels of severity and with individuals without a college education. Overall, research suggests that the NART is more accurate at mild to moderate severity ranges and becomes less accurate at more severe levels, but can still yield important clinical information.

The American version of the NART (AMNART) is correlated with semantic memory, which is compromised early in individuals with AD (Storandt, Stone, & LaBarge, 1995). Therefore, using the AMNART with these patients may be problematic because of their prominent memory problems. Stebbins et al. (1990a) used a dementia sample and found problems using the NART in those with language disturbances, in that IQ estimates were lower in this population as compared to those without such disturbances. This is especially relevant to AD patients as language is significantly affected early in the disease. Additionally, Fromm et al. (1991) determined that in their AD sample, correct pronunciation of words decreased over time, while matched elderly controls did not show this difficulty.

Additionally, the accuracy of reading "hold" tests is influenced by external factors, such as intellectual range and stage of recovery (Fromm et al., 1991; Griffin et al., 2002; Johnstone et al., 1996; Johnstone & Wilhelm, 1996). In both healthy and

neurologically impaired subjects, the efficacy of premorbid estimation using the WRAT is dependent on the range of IQ. While the estimate is accurate for those in the lower or average IQ ranges, it underestimates intellectual ability for those with higher IQ scores. However, it is more accurate than the National Adult Reading Test-Revised (NART-R), which had problems outside the average IQ range at both ends of the IQ spectrum (Johnstone et al., 1996; Wiens et al., 1993). Other studies report that stage of recovery affects the ability of the WRAT to estimate previous functioning, with those who were improving in their recovery having less accurate estimates (Fromm et al., 1991; Johnstone & Wilhelm, 1996). These caveats have particular relevance for clinical patients, as recovery is associated with a dynamic disease process that impacts the ability of measures to monitor change.

Research suggests that the addition of other variables may increase the accuracy of prediction. The NART has been used in conjunction with demographic variables. Crawford, Stewart, Parker, Besson, and Cochrane (1989) found the combination formula accounted for 73% variance in FSIQ, which was significantly more accurate than either method alone. These findings are supported by Crawford, Cochrane, Besson, Parker, and Stewart (1990a) who found that demographic variables mediated the relationship between NART and IQ. Factor analysis shows high construct validity and suggests that the NART/ demographic equation should be the method of choice for estimation (Crawford, Nelson, Blackmore, Cochrane, & Allen 1990b). Other studies have found similar results using combination formulas (Grober & Sliwinski, 1991; Willshire, Kinsella, & Prior, 1991).

In a study utilizing the WRAT with healthy subjects, Kareken, Gur, and Saykin (1995) found that adding ethnicity and parental education increased the test's ability to estimate premorbid functioning. They acknowledge that the NART accounts for more WAIS variance than the WRAT, which may be due to its development as an estimator of premorbid ability. Overall, results suggest that the combination of present ability measures and demographic variables is a viable alternative for estimating premorbid ability, accounting for more WAIS variance than either approach alone. However, there is a lack of research using the WRAT, especially with clinical populations. Further research is needed to determine the utility of the WRAT to estimate premorbid functioning.

In general, it is difficult to draw conclusions about the efficacy of reading measures as "hold" tests, especially in those with moderate to severe dementia. Initially, it may appear that the NART has more problems than the WRAT when estimating premorbid abilities at advanced disease stages, but this may be attributed to more research being conducted with the NART. Available research using the WRAT with clinical populations is also largely inconclusive, pointing to a need for more research in this area.

Wide Range Achievement Test

Compared to the NART, the WRAT has been found to yield more accurate estimates of premorbid ability in both healthy and neurologically impaired populations (Griffin et al., 2002; Johnstone et al., 1996; Orme et al., 2004; Wiens et al., 1993). The WRAT has been used as a functional baseline to identify other neuropsychological deficits in several clinical populations (Johnstone, Hexum, & Ashkanazi, 1995; Johnstone

et al., 2002). However, many studies examining the efficacy of reading as an estimate of premorbid functioning have utilized the NART, which unlike the WRAT consists of words with irregular pronunciation. It is unclear how differences in these tests affect estimation, particularly in clinical populations.

Estimates are affected by a variety of variables that are largely unknown, such as extent of reading impairment and actual relationship between previous reading level and premorbid abilities (Snyder & Nussbaum, 1998). Research has shown that reading tests are affected by various factors that influence their validity as premorbid measures. The accuracy of the measure is dependent on the IQ range of the individual (Griffin et al., 2002; Johnstone et al., 1996). Wiens et al. (1993) used healthy subjects and found that The North American Adult Reading Test (NAART) correctly estimated premorbid IQ in those who fell in the average intellectual range. However, it overestimated IQ for those in the lower ranges and underestimated it for individuals who scored in the higher IQ range. In contrast, the WRAT demonstrated accuracy for those in both the lower and average IQ ranges while it underestimated intellectual ability for those with higher IQ's. Overall, results suggest that range restriction problems arise in estimating premorbid functioning in healthy subjects when their IQ falls out of the average range.

Griffin et al. (2002) used the WRAT with chronic pain patients to determine its ability to estimate premorbid functioning across IQ ranges. They found that it most accurately classified individuals in the below average and average area, but the higher IQ range was more problematic. The majority of approaches perform best within average IQ ranges, which makes it difficult to estimate premorbid functioning for individuals outside the average intellectual range (Griffin et al., 2002; Maddrey et al., 1996). Because clinical

populations likely differ from normal ones, others have replicated these findings with neurologically impaired individuals (Johnstone et al., 1996). In general, their results corroborate those of Wiens et al. (1993).

Other factors have been shown to affect the ability of the WRAT to estimate premorbid functioning. Johnstone and Wilhelm (1996) studied a neurologically impaired sample over two testing sessions and found the ability of the WRAT to estimate previous abilities depended on current recovery. For groups that were either stable or declining, the WRAT was a good measure of premorbid intelligence. However, for groups that were improving the WRAT was accurate the first testing, but overestimated IQ the second testing. This implies that if an individual is expected to improve in their recovery, the WRAT is not a good measure of premorbid functioning. It is most accurate when individuals are cognitively stable or declining in their intellectual ability.

Research has demonstrated the utility of the WRAT to estimate premorbid functioning in normal populations, in that it accounts for significant variance in Wechsler Verbal IQ, Performance IQ, and Full Scale IQ (Kareken et al., 1995). WRAT scores of individuals with dementia have a high correlation, ranging from .58 to .81, with the WAIS-R at ranges of moderate to severe impairment (Margolis, Greenlief, & Taylor, 1985). Other researchers (Orme et al., 2004) have examined the validity of the WRAT as a measure of premorbid ability in clinical populations and found that the WRAT is a good "hold" test, in that scores are more consistent over time than other neuropsychological measures (Johnstone & Wilhelm, 1996). Using two test administrations a year apart they found that reading scores did not change over time. With regard to severity of injury, there was a non-significant trend that those with greater

TBI severity experienced a greater score change over time. This is expected considering that patients continue to improve in cognitive functioning for a significant amount of time post injury, and the accuracy of the WRAT is compromised when individuals are improving in their recovery (Johnstone & Wilhelm, 1996).

The vulnerability of reading to brain dysfunction is unclear and complicated by conflicting research. Studies suggest that although reading may be unaffected during early disease stages, it appears to be negatively impacted later. This influences the accuracy of both the WRAT and the NART. However, some research indicates that the WRAT may have fewer range restrictions than the NART (Griffin et al., 2002; Johnstone et al., 1996). Although not developed specifically for this purpose, the simplicity and availability of the WRAT makes it an excellent alternative to the NART for premorbid estimation. However, the impact of increasing severity and the WRAT's efficacy in detecting changes in functioning is largely unknown. Due to the lack of research on the WRAT's ability to estimate premorbid functioning, the present study further investigates this area.

The Halstead-Reitan Battery (HRB)

The HRB is designed to examine brain-behavior relationships. Its objective is to define these relationships and determine how they connect the biological and behavioral aspects of brain functioning. This is accomplished through assessment of a variety of functions and measurement of a range of behavioral manifestations of the brain (Reitan & Wolfson, 1993). Tests differ in their ability to detect changes in neuropsychological functioning, with some measures assessing skills that are more impacted by various brain

impairments. This study seeks to further examine the sensitivity of tests to changes in cognitive status.

There are three components necessary to establish brain-behavior relationships according to Reitan & Wolfson (1993). First, the test must adequately measure the psychological and behavioral functions of the brain. Second, the assessment must permit application to individuals. Finally, it must be validated by formal research studies with consideration of both clinical and application aspects. The ability of the HRB to meet these requirements lends credibility to its ability to detect brain impairment as well as categorize overall cognitive functioning.

The HRB has been utilized with a variety of populations and ages. One of the most difficult types of change to diagnose occurs in the elderly. Differentiating age related change from organic change is particularly problematic. Since the normal aging process results in the decline of cognitive skills, it is important to understand how deficits are reflected on HRB performance over time (Reed & Reitan, 1963a; Reed & Reitan, 1963b).

In a comprehensive longitudinal study of non-neurological participants, Elias, Robbins, and Elias (1996) examined change using HRB subtests. On certain subtests performance did not decline over time: the Category Test; Trails A and B; and Tactual Performance Test (TPT) - total, TPT - memory, and TPT - Location scores. However, tests that relied on executive functioning, the ability to use information related to planning and modifying during new situations, did show a non-significant association between age and change. Scores on finger tapping decreased over time indicating a decline in psychomotor speed. Ratcliff, Dodge, Rirzescu, and Ganguli (2003) gave

elderly normals a battery of cognitive tests, including Trails A and B and found that this test was the most impaired over time. Trails B declined more than Trails A, most likely due to its reliance on executive functioning. Results suggest that the tests may be more sensitive to age-related effects, particularly because of their emphasis on processing speed, which has been shown to decrease with age (Schludermann, Schludermann, Merryman, & Brown, 1983). These studies highlight age-related differences on subtest performance, in that several HRB subtest scores decline over time in the absence of secondary pathological processes, which makes it difficult to assess change that is due to aging versus change due to disease.

Many studies have compared normal and clinical samples on HRB subtests to determine which tests correctly differentiate the groups. Reitan (1955a) found that brain injured subjects performed progressively worse on the Halstead Impairment Index (HII), Category Test, and TPT –Location, respectively. Speech Sounds Perception Test (SSPT), Seashore Rhythm Test, TPT- memory, and the Finger Tapping Test showed less impairment. The HII is a severity index based on the HRB that calculates the proportion of seven specific subtests that fall in the impaired range. The result is a number between 0 and 1 that describes severity of impairment. Additionally, the Category Test is one of the more complex problem-solving tasks that would be expected to be problematic for those with brain dysfunction. Researchers have found it to be one of the most reliable subtests to identify brain impairment (Reitan & Wolfson, 1993). Several studies have found motor tests more likely to be less impaired than tests relying on problem solving and abstraction abilities (Bak & Greene, 1980; Butters, Goldstein, Allen, & Shemansky, 1998; Reed & Reitan, 1963a; Reed & Reitan, 1963b).

The similarities of brain dysfunction and the normal aging process compound the difficulties of diagnosis (Bigler et al., 1981; Mack & Carlson, 1978; Reitan & Wolfson, 1986b). Research has shown that elderly and brain damaged populations score similarly on the HRB. Mack and Carlson (1978) found that elderly normal subjects performed as poorly as brain-damaged patients on the Category Test. The pattern of deficits for the two groups was similar in that Subtests III and IV of the Category Test were particularly problematic. This is most likely due to greater task complexity and the alternation between retention of information and responding. Reitan (1955c) divided groups into age-based intervals, comparing brain damaged and normal individuals. While age was not a factor in classifying the brain-damaged group, it was a factor in the older age group. Additionally, this group performed similarly to the brain damaged group on the HII. Reitan (1962) found that the level of performance of normal elderly participants over about a 30-year period was similar to scores observed in those with cerebral damage.

Reed and Reitan (1963a) and Bak and Greene (1980) examined patterns of impairment on HRB subtests by comparing old and older groups. Both studies found the TPT to be particularly problematic, followed by Trails B. Performance on SSPT, Seashore Rhythm, and finger tapping were less impaired. On the Seashore Test, older subjects actually performed better than their younger counterparts (Reed & Reitan, 1963a). Hom (1992) looked at HRB subtests performance in an AD sample and found the following rankings of the tests from most impaired to least: Category Test, Trails B, TPT-Memory, HII, TPT-Location, TPT-Total, Seashore, SSPT, Trails A, Finger Tapping, and Grip Strength. This is consistent with other researchers (Bak & Greene, 1980; Butters et al., 1998; Reed & Reitan, 1963a; Reed & Reitan, 1963b) who have also found that

performance on motor tests was less impaired, while tasks requiring complex, abstraction abilities were more impaired.

Moehle and Long (1989) found that the most age-sensitive HRB subtest was TPT-Location. It clearly divided the research sample into three distinct age groups. Other tests showed age effects including Trails B, SSPT, the Rhythm Test, and TPT-Memory. These results point to a "specific decline model nested within a general decline model" (Moehle & Long, 1989, p. 176). Overall, the findings support a decrease in test performance for older adults with some tests being particularly sensitive. Elias, Robbins, Walter, and Schultz (1993) extended the previous work to include the Category Test and rankings of the tests that best discriminated among six age groups.

Additional research points to specific abilities that are lacking in the elderly. Reed and Reitan (1963a) hypothesized that age related changes documented on psychological tests were the result of underlying physiological changes affecting brain function. On all but two measures, including Seashore Rhythm Test, subjects over 50 performed worse than the younger group, aged 40-49. Overall, few abilities were spared by the aging process, but some skills were more affected than others. Changes did appear to be linked to organic deficits in the brain. Reed and Reitan (1963b) investigated why older subjects perform worse than younger ones on certain tests. Younger subjects performed better than older subjects on tasks requiring immediate adaptive ability. However, older subjects were slightly better than young subjects on tests using stored memory, although the difference was not statistically significant.

Research suggests that AD results in different impairments, both quantitatively and qualitatively, than other types of dementia. Compared to those with vascular

dementia, AD produces more deficits in language function and non-verbal memory (Ballion et al., 2003). Russell and Polakoff (1993) point to a specific pattern of deficits for AD patients on the HRB and WAIS that is different than patterns seen in their MID participants. They found early memory and cognitive impairments, while deficits in motor function occur later in the disease process.

The Trail Making Test has been documented as among the most difficult tests for the elderly and those with dementia. It has been found that dementia affects performance both in errors and time required regardless of age (Rasmusson, Zonderman, Kawas, & Resnick, 1998). Further analysis of errors suggests that it may be due to inefficient inhibitory mechanisms. Researchers found that 67% of errors in those with AD were due to inability to inhibit versus 24% in matched elderly controls. Additional research indicates that impairment on Trails may also be impacted by the necessity of concurrently manipulating information (Baillon et al., 2003). In general, it appears that AD patients have difficulty with manipulating and suppressing irrelevant information during the task, which leads to poorer performance (Amieva et al., 1998).

Gender Differences on the HRB

The influence of subject variables has been a methodological concern for neuropsychological research (Parsons & Prigatano, 1978). Extraneous variables can significantly impact test performance. Presentation format, subject gender, examiner gender, age, and education have been found to influence scores (Chavez, Trautt, Brandon, & Steyaert, 1983; Kupke, 1983; Seidenberg et al., 1984; Ruff & Parker, 1993). For example, the interaction between subject gender and examiner gender has been shown to affect performance. Kupke (1983) found that opposite-gender pairs of subjects

and examiners scored higher than same-gender pairs on TPT-Location and TPT-Memory subtests.

Repeated studies have demonstrated that certain HRB subtests have pronounced gender differences. Grip strength and Finger Tapping are well known for their male-superior test performance (Chavez et al., 1983; Dodrill, 1979; Gordon & O'Dell, 1983; Morrison, Gregory, & Paul, 1979; Ruff & Parker, 1993; Seidenberg et al., 1984; Yeudall, Reddon, Gill, & Stefanyk, 1987). Another study showed that women perform worse than men on Finger Tapping and, in contrast to men, get substantially slower with age (Ruff & Parker, 1993).

Chavez, Schwartz, and Brandon (1982) found that females had higher TPT-Location scores than males. Gordon & O'Dell (1983) demonstrated female-superiority on both TPT-Location and TPT-Memory scores. Dodrill (1979) found the expected male-superior performance on Finger Tapping and Grip Strength, but observed differences between neurological and non-neurological groups. Gender differences were more pronounced in the non-neurological group, suggesting that as brain functioning is affected, the variability from gender differences decreases. It seems as though gender differences decrease in importance when other variables, such as brain impairment, are introduced.

Alzheimer's Disease (AD)

As the United States population ages, recognizing and accurately diagnosing illnesses prevalent in the elderly become increasingly important. One of the most common mental health problems seen in the elderly is dementia. Dementia can be considered to be the loss or deterioration of cognitive abilities, which results in

impairment of functioning in daily living activities (Green, 1995). While there are many etiologies of dementia, the focus of this study is one of the most frequent, Alzheimer's disease.

Originally identified in 1906 by Alois Alzheimer, AD has become a growing problem in the United States, with 8-15% of individuals over the age of 65 having the disease. Individuals with AD are usually 60 years of age or older (Strub & Black, 1988; Victor & Ropper, 2001). Psychological deficits are the result of physiological changes in the brain including decrease in brain mass, increase in ventricular space, atrophy, and cell loss (Adams, Parsons, Culbertson, & Nixon, 1996; Victor & Ropper, 2001). AD is an illness that affects cortical functioning and a broad range of abilities, including those in emotional, social, and cognitive areas. Initial symptoms progress to a chronic course of overall deterioration in broad areas of cognitive functioning (Strub & Black, 1988). Mental status exams indicate a specific pattern of deficits in social abilities, an absence of signs indicating impaired consciousness as seen in individuals with delirium, and decreased functioning in several cognitive areas (Strub & Black, 1988). Specifically, AD is characterized by an insidious onset of memory problems, difficulties in problem-solving and executive functioning, distractibility, and failure to react to environmental stimuli with usual speed and accuracy (Green, 1995; Snyder & Nussbaum, 1998). Behavioral symptoms are also prominent in individuals with AD and include depression, psychosis, and agitation (Surgeon General, n.d.).

Understanding neuropsychological change associated with AD is necessary to determine current cognitive status and patient prognosis. The estimation of premorbid abilities allows the measurement of changes in neuropsychological functioning related to

brain injury (Kareken et al., 1995). The interpretation of test performance as declined, improved, or unchanged is ultimately linked to how the individual functioned in the past (Gladsjo, Heaton, Palmer, Taylor, & Jeste, 1999). Regardless of the source of brain impairment, determining the resulting cognitive decline is essential because it allows deficits to be understood by considering previous abilities as an intellectual and functional baseline. This enables clinicians to evaluate the extent of impairment by comparing current deficits to estimates of previous functioning, providing a more accurate understanding of impairments independent of premorbid functioning.

AD is the most prevalent type of dementia and is the most frequent cause of institutionalization of the elderly (Nolan, Swihart, & Pirozzolo, 1986). Although AD is common, there is still much that is unknown. There is no cure for AD, and it cannot be positively diagnosed until death. The problem of identification is compounded by the characteristics it shares with other diseases and makes differentiating it from other types of dementia especially difficult.

Research suggests that there are both genetic and environmental components in the development of AD. Although it appears that the disease is influenced by genetic factors, with familial occurrence in 1% of all cases, more than one genetic factor may be needed to develop the disease (Victor & Ropper, 2001). Risk factors include birth order, mother's age at birth, advanced age, and family history. Reports suggest a relationship between AD and Down's syndrome, Parkinson's disease, and previous head injury (Snyder & Nussbaum, 1998; Victor & Ropper, 2001). Overall, females are more likely to be diagnosed and those with onset before age 70 are more likely to have relatives with AD (Li et al., 1995; Surgeon General, n.d.). This finding supports the hypothesis that

family history of AD is related to its development, especially in conjunction with early onset.

The etiology of AD is unknown and diagnosis is by identification of clinical features and exclusion of other causes of dementia. However, different dementias have features in common and comorbid disorders can complicate diagnosing. McKhann et al. (1984) report the findings of a task force assembled to define the criteria for diagnosing AD. The requirements of diagnosis are divided into three categories: probable, possible, and definite. The criteria for probable AD includes verification of dementia through clinical examination and neuropsychological tests, deficits in two or more areas of cognition, progressive worsening of memory, no disturbance of consciousness, onset between 40 and 90 years of age, and absence of other systemic disorders which could better account for symptoms. A diagnosis of possible AD is made in the presence of the dementia syndrome, which is characterized by decline of memory and other cognitive functions. Other systemic or psychiatric diseases cannot better account for symptoms or if present, they must be insufficient to produce observed clinical symptoms. To diagnose definite AD, the criteria for probable AD must be met along with histological evidence from a biopsy or autopsy.

Anatomical changes resulting from AD are characterized by the presence of neurofibrillary tangles and neuritic plaques. Overall changes to the brain include narrowed cerebral convolutions and wider sulci throughout the cerebral cortex (Victor & Ropper, 2001). Computer Assisted Topography (CAT) and Magnetic Resonance Imaging (MRI) scans indicate granulovascular changes, atrophy, cell loss, and ventricular dilation (Adams et al., 1996). Cell loss results in fewer working neurons and disruption of

connections between them, which hinders the transmission of neural messages and produces cognitive impairments. Atrophy of brain tissue results in larger ventricles and overall brain shrinkage, which can decrease brain matter up to 15-20% at the end stages of the disease (Adams et al., 1996; Victor & Ropper, 2001). Atrophy is regional and typically involves the amygdala and much of the frontal, temporal, and parietal lobes with extreme atrophy seen in the hippocampus (Victor & Ropper, 2001). These anatomical changes in the brain result in a pattern of psychological deficits unique to AD and represent significant deviations from normal aging, both quantitatively and qualitatively. The pattern of impairment is exclusive to AD and differs from other diseases such as MID and Parkinson's. Further investigation is needed to understand the psychological aspects of this disease in order to elucidate the disease process.

Research shows that deficits occur early in AD, even when observable signs of impairment are not present. Neuropsychological and cognitive tests are able to identify those who appear asymptomatic but will later develop dementia (Cervilla, Prince, Joels, Lovestone, & Mann, 2004). Neuropsychological data of asymptomatic relatives of AD patients shows lower functioning in several cognitive areas, suggesting that impairment occurs prior to full manifestation of the disease (Hom, Turner, Risser, Bonte, & Tintner, 1994). Those with suspected dementia perform worse than normal subjects on the Trail Making Test, even when social and functional impairment are absent (Rasmusson et al., 1998). Additionally, a longitudinal study using pre-AD patients demonstrated that MMSE scores, recall of organizable words, facial recognition, and letter fluency were predictors for those who would develop dementia during a three-year time period (Small, Herlitz, Fratiglioni, Almkvist, & Backman, 1997).

Despite early studies that viewed AD as a disease characterized by overall cognitive dysfunction, later research indicates that initial amnesic deficits may be present for years before other cognitive domains such as language, semantic memory, and visuospatial functioning are significantly affected (Perry & Hodges, 1999). However, even in early stages AD impacts a wide range of cognitive functions (Hom, 1992). AD affects not only memory, but other cognitive abilities such as attention (Lezak, 1995; Parasuraman & Haxby, 1993; Perry & Hodges, 1999), executive functioning (Lafleche & Albert, 1995), and language (Adams et al., 1996; Huber, Shuttleworth, & Freidenberg, 1989; Snyder & Nussbaum, 1998). Motor and sensory skills appear to be largely unimpaired in AD patients (Hom, 1992).

Determination of change in neuropsychological functioning is accomplished by understanding brain-behavior relationships as they relate to AD. This requires a thorough assessment of a broad range of behaviors and abilities. Much of the previous research on AD has focused on specific cognitive deficits while ignoring general functions that depend on the integrity of the brain. Impairment can be divided into general and specific functions, which are based on the location, laterality, and the reliance on specific areas within the brain that depend on the integrity of a particular hemisphere (Hom, 1992). General functions are incidental memory, attention/concentration, and abstract reasoning ability. Specific brain functions include semantic memory, language, and academic/ verbal learning (Hom, 1992). Ignoring the wide range of impairments associated with AD results in an incomplete understanding of change, leading to an inability to determine whether deficits are premorbid.

The most obvious deficit associated with AD concerns memory, which is highly correlated with the progression of the disease and one of its earliest observable symptoms. As memory begins to fail, recall of recent events is affected first and is related to the magnitude of the underlying organic cause (Victor & Ropper, 2001). Difficulty in the formation of new memories is common; however, long-term memories appear to be well preserved. Therefore, it is likely that an individual with AD could go into great detail about incidents from childhood, but could not tell you what he had for breakfast. It is problematic to assess the veracity of remote memories and confabulations are common as patients attempt to cover the gaps (Victor & Ropper, 2001).

There are two overarching memory systems: declarative or explicit memory and non-declarative, also called implicit or procedural memory. Implicit memory contains information that has been learned, such as habits, skills, procedures, and abilities (Adams et al., 1996). This is the type of memory that tells how to do something. Explicit memory is knowledge about facts or events (Adams et al., 1996). It includes time, place, and emotions. Explicit memory can be divided into episodic and semantic memory. Episodic memory contains personal experiences and their relationships to each other (Victor & Ropper, 2001). Semantic memory is comprised of perceptual and factual knowledge, which makes it possible to understand language (Victor & Ropper, 2001). Semantic memory also contains information about meanings, historical figures, and events.

One type of measured memory is incidental memory, which is assessed unexpectedly when the participants are unaware that their memory of the task will be tested. On the HRB, TPT - Memory and TPT-Location scores measure incidental memory. Elias et al. (1996) report that performances on these tasks do not decline over

time, which suggests that they are stable in a non-neurological population. Hom (1992) found that these scores are significantly lower in those with AD compared to normal elderly participants. This indicates that incidental memory is one aspect of higher cognitive function that is impaired in AD.

Semantic memory is a specific memory function that depends on intact hemispheres within the brain and appears to be impaired in those with AD (Hom, 1992). The involvement of medial temporal structures influences early hippocampal dysfunction and appears to be one of the primary physiological causes of memory impairment, particularly in the formation of new episodic memories (Parasuraman & Haxby, 1993; Perry & Hodges, 1999). Elderly participants perform equally to younger individuals on tasks of semantic memory, which include measures of vocabulary and general knowledge. Semantic material is well learned, which may explain why these skills are stable despite brain dysfunction (Snyder & Nussbaum, 1998). Examining memory loss to determine whether aspects of memory are differentially affected, Nebes et al. (1984) found that compared to normals, AD patients performed significantly worse on episodic memory tasks; however, they were equal to controls in tests of semantic memory. This is consistent with Ober, Shenaut, and Reed (1995), who found that semantic memory appears to be preserved in the early stages of the disease. However, others have found that those with early stage AD show deficits in explicit semantic and implicit memory functioning (Green, 1995; Heindel, Salmon, Shults, Walicke, & Butters, 1989; Monti et al., 1996), suggesting that it is inaccurate to state broadly that overall memory is impaired because there are several separate components of memory that are differentially affected.

AD patients perform poorly on tests of memory, whether rote recall of information, recall in sentence form, or in the context of a story (Lezak, 1995; Nebes et al., 1984). Strub and Black (1988) report that patients often receive low scores on paragraph reading involving logical memory, paired associated words, and visual memory. Additionally, difficulty generating word lists often precedes other language deficits, with AD patients scoring below normals despite cues to facilitate recognition (Lezak, 1995). Johnstone et al. (2002) examined the neuropsychological deficit profile for AD and determined that attention and memory are the most significant problems for these individuals, which is consistent with the observation that memory impairments are one of the first deficits to develop and one of the most pronounced problems in AD patients.

Problems with attention and concentration follow memory impairments as the disease progresses (Parasuraman & Haxby, 1993; Perry & Hodges, 1999). There is evidence that difficulties with attention occur early in AD and may be the first indicator of neocortical dysfunction (Hom, 1992; Perry & Hodges, 1999). However, similar to impairment of memory, areas of attention are differentially affected. While AD patients have problems with shifting and dividing attention, focused attention appears to be only marginally affected (Perry & Hodges, 1999). Other areas of attentional dysfunction seen in AD include reduced attention span, inability to focus, and decreased reaction time (Lezak, 1995). Within the HRB, the SSPT, Seashore Rhythm Test, and Trails A assess attention and concentration. Reed and Reitan (1963a) found that abilities required on the Seashore test seemed to be well preserved in a dementia sample. However, it has been

shown that AD patients perform worse than normals on Trails, even in the absence of social dysfunction (Amieva et al., 1998; Rasmusson et al., 1998).

Another area that is impacted by AD is abstract reasoning. The HRB measures this skill on the Category Test, TPT-Total, and Trails B. Hom (1992) found that these abilities were significantly impaired in those with AD. The Category Test is the single most effective test in the HRB for identifying cerebral impairments (Reitan & Wolfson, 1993). It has been found that elderly normals and individuals with brain damage perform equally poorly and obtain similar patterns of deficits on the Category Test, suggesting that impairments seen in brain injury are also reflected in those with normal aging (Bigler et al., 1981; Mack & Carlson, 1978; Reitan & Wolfson, 1986b).

Additional studies show that AD patients have difficulties with executive functioning prior to problems with language and visuospatial tasks (Perry & Hodges, 1999). Lafleche and Albert (1995) compared AD patients to controls on several tests of executive functioning. Those with AD performed significantly worse on tasks that required concurrent manipulation of information. Tasks that include simple concept formation, figure copying, attention and naming, and cue-directed behavior did not differentiate between the two groups. They concluded that AD patients have problems of executive functioning independent of memory impairments.

Specific functions affected by AD include language abilities. Language skills appear to be differentially affected, with dysnomia and dyslexia occurring more often in those with AD compared to normals (Hom, 1992). Even early in the disease, changes are evident in the quality, quantity, and meaningfulness of speech, as well as verbal comprehension abilities. The disease disrupts the linguistic features of speech, affects

content of spontaneous speech and confrontational naming, and results in poor performance in tasks assessing semantic and letter fluency (Adams et al., 1996; Huber et al., 1989; Lezak, 1995; Snyder & Nussbaum, 1998). Focusing on written linguistic output, Kemper et al.(1993) found that when AD patients were asked to construct sentences, the length of clauses, informational content, and quality of sentence composition decreased as severity of dementia increased. However, even though individuals who were considered mild or moderately demented produced grammatically correct sentence structure, it was to a lesser degree than non-demented individuals. This suggests that language skills are affected at earlier stages of the disease, but may not be significant enough to be evident in daily activities and interactions with others.

Difficulties with academic/verbal learning are related to specific brain functions and are especially relevant for AD because language disturbances result in problems with reading, particularly reading aloud. Since these skills impact both the NART and WRAT-Reading, there are significant implications for using these instruments with AD patients who have language difficulties. There are few studies that have examined dementia and the WRAT. For those with suspected dementia in the moderate to severe range, the WRAT- Reading is highly correlated with the WAIS (Margolis et al., 1985). In healthy subjects, the WRAT- Reading accounted for significant variance in FSIQ, VIQ, and PIQ; however, the NART appeared to account for more variance than the WRAT (Kareken et al., 1995).

It has not been determined at what point during the disease process reading abilities are affected. Some researchers have found that the ability to read aloud is well maintained until severe stages of AD, supporting an absence of relationship between the

reading of irregular words and dementia severity (Cummings et al., 1986). However, others examining language disturbances in AD have determined that the NART is affected by disease severity. Patterson, Graham, and Hodges (1994) divided AD patients into groups based on severity and found that reading was impaired in the moderate AD group and related to problems with semantic memory. This finding is consistent with others who have identified difficulties with semantic memory in AD patients (Green, 1995; Heindel et al., 1989; Monti et al., 1996).

The progression and the manner in which cognitive structures and processes are affected are unique to AD. In order to facilitate differential diagnosis between AD and other dementias, research has focused on the sequence of pathological changes and the ways that these changes manifest into the behavioral symptoms associated with AD. These differences can be seen in neuropsychological testing. There is a neuropsychological deficit pattern specifically associated with AD and it is distinct from other disorders such as TBI, chronic pain, and E. Lupus (Johnstone et al., 1995; Johnstone et al., 2002; Skeel, Johnstone, Yangco, Walker, & Komatireddy, 2000). Using the WRAT as an estimate of baseline functioning, it was determined that cognitive flexibility (Trails B) was most impaired. This was followed by speed of processing (Trails A), attention/memory (Wechsler Memory Scale), and intelligence (WAIS). These results suggest a distinct pattern of impairments seen in AD that can ultimately be linked to specific changes in brain pathology.

Hypotheses

Cognitive abilities are affected differentially by the brain dysfunction associated with specific disease processes, with some skills being more impacted than others.

Therefore, measures of these abilities will differ in their sensitivity to changes in brain functioning. Some subtests assess abilities that are less impacted by brain dysfunction and can be considered "hold" tests, in that performance will not be impacted significantly as severity increases. However, other subtests measure abilities that are not resilient to brain dysfunction and are considered "don't hold" tests. Performance on these tests will be affected negatively as severity increases.

Previous research is conflicted concerning the ability of tests to measure change in cognitive status, particularly as brain dysfunction becomes severe (Fromm et al., 1991; Paolo et al., 1997; Stebbins et al., 1990a; Stebbins et al., 1990b). One objective of the present study is to determine whether several common neuropsychological measures are sensitive to differences in severity of brain dysfunction. Specifically, this study investigated the ability of the WRAT- Reading and HRB subtests to detect differences in severity using an AD sample. This was accomplished by using a correlation, r, to determine the strength of relationship between specific subtests and a measure of severity, Mini Mental State Exam (MMSE) score. The subtests were ranked according to the magnitude of relationship with MMSE. Subtests with larger magnitude correlations to severity were considered "don't hold" tests, while those with smaller magnitude correlations were considered "hold" tests. These correlations allow indirect inferences about the ability of the WRAT - Reading and HRB subtests to detect changes in neuropsychological functioning associated with AD. The criterion variable used was MMSE score.

Hypothesis One

The first hypothesis stated that the HRB and WRAT - Reading subtests would be a mixture of "hold" and "don't hold" tests. This hypothesis was tested by computing the Pearson correlation between each of the HRB and WRAT subtests and MMSE score. The greater the magnitude of the correlation between the given HRB or WRAT subtest and the criterion variable, the stronger the evidence that the subtest is a "don't hold" test. Conversely, the smaller the magnitude of the correlation between a given subtest and the MMSE, the stronger the evidence that the subtest is a "hold" test. The magnitude of the correlations for the 18 subtests were ranked to provide a continuum describing "hold/ don't hold" status for the entire sample, not divided by severity group.

Hypothesis 1A. HRB subtests that assess abstraction and problem solving (e.g., the Category Test) would be affected significantly by severity of AD as measured by the MMSE. Therefore, it was hypothesized that the Category Test would one of the strongest "don't hold" test for both males and females, which would be indicated by the greatest magnitude correlation.

Hypothesis 1B. Tests that rely heavily on attention and concentration (e.g., SSPT, Rhythm Test, and Trails A) would be the second most compromised group of subtests for both males and females.

Hypothesis 1C. Tests that assess incidental memory (e.g., Memory) would be the third most compromised group subtests for both males and females.

Hypothesis 1D. In contrast, the motor tests, Grip Strength and Finger Tapping, will be strong "hold" tests for both the dominant and non-dominant hands for both males and females.

Hypothesis 1E. The WRAT-Reading subtest would be a strong "hold" test for both males and females.

Hypotheses 1A through 1E were tested by examination of the Pearson correlations between the subtests and the MMSE score.

Hypothesis Two

A smaller subset of "don't hold" tests would be significant predictors of severity status as measured by the MMSE. It was hypothesized that the Category Test would be a significant predictor, as well as at least one of the TPT subtests. It was hypothesized that the motor tests would not accurately predict MMSE. Additionally, it was hypothesized that the subset of significant tests would differ based on gender. This hypothesis was tested by stepwise Multiple Linear Regression (MLR) with MMSE score as the criterion.

Hypothesis Three

The third hypothesis is that a smaller subset of "don't hold" tests would be significant predictors of severity group membership and these would vary by gender. This hypothesis was tested by using MLR and stepwise discriminant analysis to determine the ability of the WRAT - Reading and HRB subtests to discriminate between two severity groups of AD participants. Using MMSE scores, the sample was divided into two severity groups: normal/mildly impaired and moderate/severely impaired. The analyses were conducted using both the full sample and the sample divided into male and female groups. The 18 HRB and WRAT- Reading subtests were used as the predictors while severity group membership will be the criterion, or dependent variable. The results of the discriminant analysis were used to classify participants into two severity groups and the proportion of hits/ misses were examined.

Hypothesis Four

The fourth hypothesis is that the neurophysiological differences between males and females would result in significant gender differences in test performance on several of the HRB subtests (Chavez et al., 1983; Dodrill, 1979; Morrison et al., 1979; Seidenberg et al., 1984; Yeudall et al., 1987). Specifically, females should have lower scores on Grip Strength and Finger Tapping, with both their dominant and non-dominant hands. This hypothesis was tested using an independent samples t-test to determine whether there are significant group differences on specific subtests.

Chapter 2: Method

The purpose of the present study was to examine the ability of the WRAT- Reading and HRB subtests to detect differences in neuropsychological functioning in a sample of clinical patients. Specifically, it allowed inferences about the sensitivity of subtests to changes in severity in a sample of AD patients. Statistical analyses were used to determine the sensitivity of the WRAT-Reading and HRB subtests to detect changes in functioning, represented by increasing severity of brain impairment due to chronic illness. The study used subtest scores on the HRB and the WRAT- Reading subtest (Wilkinson, 1993) as independent variables. The HRB and WRAT- Reading are examples of psychological measures that assess a range of cognitive abilities differentially affected by brain dysfunction. The dependent, or criterion, variable was MMSE scores, or in the case of hypothesis three, impairment groups derived from the MMSE score distribution.

Participants

Archival data from individuals diagnosed with AD were used to assess indirectly neuropsychological subtests' sensitivity to change in illness severity. The AD sample consisted of 151 participants with a mean age of 72.89 years (SD = 7.65) and a mean education level of 12.80 (SD = 3.32). Education was the only demographic variable that

showed significant gender differences. Males had a higher number of years of education (p> .005). The majority of participants were female (N = 105) and Caucasian (N=137).

The demographic information for the study's participants is given in Table 1. The majority of participants were patients at The Alzheimer's Disease Center at University of Texas Southwestern. These participants were referred to the Center because of significant AD related impairments and were evaluated extensively, including neurological workup and neuroimaging. This results in a more thorough clinical evaluation than is afforded most patients seen in private practice. The remaining participants were evaluated by a neuropsychologist in private practice. All subjects were evaluated by a Board-certified neurologist and had a diagnosis of "probable Alzheimer's Disease" as based on the National Institute of Neurological and Communicative Diseases and Stroke- Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). The criteria for this diagnosis include verification of dementia through clinical examination and neuropsychological tests, deficits in two or more areas of cognition, progressive worsening of memory, no disturbance of consciousness, onset between 40 and 90 years of age, and the absence of other systemic disorders which could better account for symptoms. Participants who had a history of other neurological or psychiatric disorders were excluded from the study. All information was held confidential, no identifying information was used, and all data were viewed only by the primary researcher. The collection and use of the archival data has been approved by the Human Use Committee at Louisiana Tech University (approval #HUC-202).

Table 1

Mean and Standard Deviation of Demographic Variables

	Sample Size	Mean	Standard Deviation
Male			
Ethnicity			
Caucasian	41		
African American	1		
Other	0		
Unreported	4		
Total	46		
Age	46	71.65	7.19
Education	46	13.93	3.29
MMSE	46	19.54	5.29
Female			
Ethnicity			
Caucasian	96		
African American	3		
Other	1		
Unreported	5		
Total	105		
Age	105	73.44	7.82
Education	105	12.30	3.32
MMSE	105	18.82	4.92

Instrumentation

Mini Mental State Exam (MMSE)

The MMSE is designed to assist in the examination of an individual's cognitive state. It consists of a series of questions grouped into 11 categories: orientation to time, orientation to place, registration, attention and calculation, recall, naming, repetition, comprehension, reading, writing, and drawing (Folstein, Folstein, McHugh, & Fanjiang, 2001). Advantages of the MMSE include its ease of administration, brevity, and easy score calculation. Scores are calculated from the number of correct items and can range from 0 to 30. Any items that are incorrectly answered are scored as a 0. The MMSE can be used to classify the severity of cognitive impairment of both dementia and medical patients (Folstein et al., 2001). The authors make the following recommendations for cutoff scores: >27 indicates normal functioning, mild impairment is signaled by scores between 21 and 26, moderate impairment is indicated by scores 11 to 20, and scores 10 or below suggest severe impairment (Folstein et al., 2001). The MMSE is the most often used measure of severity in studies of demented patients and is considered the "Gold Standard" for the measurement of severity in AD patients. (R. D. Vanderploeg, personal communication, July 6, 2006)

Wide Range Achievement Test-III (WRAT-III)

The WRAT-III is an achievement test whose first edition was introduced in 1936. Since that time, it has undergone six revisions; most of these have been due to re-norming (Jastak & Wilkinson, 1984; Reynolds, 1986). There are significant advantages to the WRAT including simplicity, easy administration, comprehensive norms, and availability

(Kareken et al., 1995). However, others (Reynolds, 1986) point to problems with previous editions of the WRAT, including questionable evidence of reliability and validity.

The WRAT is designed to assess academic achievement and can be used from ages 5 to 74 years (Wilkinson, 1993). It was intended to be an adjunct test to intelligence and behavioral measures. It consists of three subtests measuring spelling, reading, and arithmetic abilities. Reading incorporates recognizing/naming letters and pronouncing words that are out of context. Respondents are asked to pronounce a list of 42 isolated words that increase in difficulty. The spelling subtest includes writing 40 words to dictation after hearing them in the context of a sentence. The arithmetic portion consists of 40 problems intended to assess abilities such as counting, solving oral problems, and completing written computations (Wilkinson, 1993).

There are two forms, which can be utilized individually, in a pre-test/post-test format, or combined to create a more comprehensive evaluation. Both forms were given to norming participants in a counterbalanced design to control for differences associated with order of administration. The two WRAT forms were developed from a common list of items that were used to create equivalent forms. Items were designed to measure the full range of the domain without significant duplication (Wilkinson, 1993).

The standardization of the WRAT-III was done in 1992 and 1993 and utilized 4,433 individuals (Snelbaker, Wilkinson, Robertson, & Glutting, 2001). The sample was stratified according to census data to control for age, gender, ethnicity, region, and socioeconomic status (Wilkinson, 1993). The standardization sample included 50.7%

males and 49.3% females. The ethnicity of the participants was 71.07% White, 13.6% Black, 10.7% Hispanic, and 3.9% Other (Snelbaker et al., 2001).

Reporting the psychometric properties of a test is necessary to demonstrate its utility and soundness as a measure. Reliability describes an assessment's ability to measure traits or skills in a consistent manner. The coefficient alpha, a measure of internal consistency, ranges from .85 to .95 across the WRAT subtests (Wilkinson, 1993). An alternate form correlation, another measure of reliability, had a median correlation of .92 for the reading subtest, a median of .93 for the spelling subtest, and the arithmetic subtest median was reported at .89 (Wilkinson, 1993). The test-retest reliability is a method of assessing reliability by administering a measure to a group of examinees on two separate occasions and computing the correlation between their scores (Aiken, 2003). The stability, or test-retest, coefficient for the WRAT ranges from .91 to .98 across the subtests (Snelbaker et al., 2001). Additionally, standard errors of measurement were calculated for standard scores, which provide information about the error in interpreting an individual's score. It was found that the Reading subtest has a mean standard error of measurement ranging from 4.5 to 4.9 (Snelbaker et al., 2001). These indices suggest that the WRAT has good reliability across its three subtests individually and in its combined form.

Validity is concerned with whether a test measures the construct it was designed to measure. Content validity for the WRAT was assessed by evaluating the item selection of the subtests, which should consist of the items ranging from easy to more difficult. This is determined by an analysis of each set of items to establish whether the goal was reached. Content validity for the WRAT is reported by the Rasch statistic of item

separation, which assumes that all items are equally discriminating, and is concerned with item separation and person separation (Aiken, 2003; Snelbaker et al., 2001). Item separation concerns how well the test items define the measured variable, while person separation is concerned with how well items differentiate between individual performances. The WRAT had an item separation of 1.00, which is the highest possible score, and a person separation of .98 to .99 (Snelbaker et al., 2001). These indices suggest that the WRAT has a satisfactory representative sample of the domain, while also providing evidence that it differentiates between individuals. Concurrent validity can be assessed by determining how well a test is correlated with other tests that purport to measure the same or similar constructs. The WRAT is highly correlated with other measures of academic achievement. The correlation between the WRAT and Comprehensive Test of Basic Skills-Fourth Edition was .69, the California Achievement Test-Form E was .72, and the Stanford Achievement Test was .87 (Snelbaker et al., 2001). This indicates that the WRAT is a satisfactory measure of achievement while also measuring constructs that other tests do not.

The Halstead-Reitan Battery (HRB)

The HRB is one of the most widely used and well-researched neuropsychological instruments available with high, established validity (Dean, 1985; Russell, 1998). It is comprised of 8 subtests that originated from a variety of sources including both neurology and psychology. The HRB began with 27 measures used by Halstead to examine biological intelligence and brain function. Over time, these tests yielded 10 measures that were included in the HRB. Halstead selected the 10 tests that best differentiated between brain injured and normal controls (Russell, Neuringer, &

Goldstein, 1970). Three tests were dropped and seven are currently used to calculate the HII. Later Reitan discarded several of the original measures due to lack of differentiation between normal and brain-injured populations (Jarvis & Barth, 1984). Reitan's contributions included adding WAIS subtests, standardizing techniques, and other methods of inference, which resulted in a more powerful assessment instrument (Schludermann et al., 1983).

Historically, brain damage diagnosis was accomplished by using an inferential procedure, which consisted of identifying brain deficits and drawing conclusions about functioning by comparing the brain damaged patient to individuals with normal brains. There is a fundamental problem inferring dysfunction solely by comparing scores to normal populations. To address this problem, the HRB was standardized using both neurologically normal and brain injured individuals. The result is a standardized neuropsychological battery that was developed so that a variable such as brain lesions could differ while dependent variables, such as test results, could be held constant (Reitan & Wolfson, 1993). This allows a more accurate identification of deficits because subjects are being compared to others with similar injuries. However, a limitation of this approach is the variability of abilities in those with both normal and damaged brains (Reitan & Wolfson, 1986a; Reitan & Wolfson, 1993).

Initially, psychologists attempted to develop a single test of organicity that could accurately identify brain dysfunction. Halstead believed that one test could not possibly capture the multidimensional nature of brain impairment. Overall, brain injured patients had difficulties in problem solving, logical analysis, and reaching conclusions from their observations (Reitan & Wolfson, 2004). A series of tests was needed in order to identify

both general and specific brain functions. He observed brain injured patients during their daily activities and recorded the range of behavioral responses. Additionally, he noted aspects of their behavior that differed from normal individuals. This began a radically different way of evaluating brain dysfunction (Reitan, 1994; Reitan & Wolfson, 2004). The result was the development of a battery of tests that examined various aspects of brain function leading to a more comprehensive picture of impairments that is based on empirical criteria rather than theoretical orientation (Reitan & Wolfson, 1993; Schludermann et al., 1983).

Halstead's original assessments included in the HRB were his estimates of the tests that would be sensitive to brain damage based on existing research and practice (Schludermann et al., 1983). Once appropriate tests were identified, additional research was done on over 8,000 individuals including brain injured and control subjects. Each participant was given the series of tests, which generated a neuropsychological interpretation about brain functioning. These results were compared to neurological evidence from neurologists, neurosurgeons, and neuropathologists (Reitan & Wolfson, 1993; Reitan & Wolfson, 2004). The correspondence between several methods of evaluation gave credibility to the use of the HRB in determining brain functioning.

This research prompted the use of the HRB in localizing brain lesions, at a time when technology was unable to fulfill this need. Currently, the focus of neuropsychology has shifted from localizing brain damage to evaluating brain-behavior relationships. However, it is important to recognize that imaging procedures and neuropsychological evidence are each independent components of a comprehensive evaluation of brain injury. Imaging results are highly correlated with HRB results, but the HRB yields further

information about higher brain functions, intelligence, and adaptive activities that imaging cannot provide (Reitan & Wolfson, 2004).

The unique method of development and norming of the HRB resulted in tests that were psychometrically different from other tests of the time, in that they required the subject to go beyond solving the immediate problem. Here, it was necessary to consider the nature of the problem, analyze its elements, and use this information to solve it (Reitan & Wolfson, 2004). Halstead included only tests that could differentiate between brain injured and control subjects while additional tests sensitive to brain dysfunction were added later (Schludermann et al., 1983). Halstead introduced 10 tests that reflected his concept of biological intelligence, but only 7 of these have remained in the battery. These are the principle components of the test and are used to calculate the HII (Reitan & Wolfson, 2004). The goal of the battery was to include tests that could sample various perceptual, motor, and cognitive functioning without redundancy and extensively cover all aspects that could be affected by brain damage in the shortest time possible (Schludermann et al., 1983). These characteristics enable the HRB to provide a comprehensive view of brain-behavior relationships.

Inferential pattern analysis is one of the Reitan's most important contributions to the HRB (Schludermann et al., 1983). Inferential pattern analysis is characterized by the relationship between tests as compared to an individual's score or functioning on single tests (Schludermann et al., 1983). Pattern analysis allows conclusions about an individual's condition to be drawn on their unique pattern of scores on the test battery. This reflects a fixed battery approach since pattern analysis requires that all components of a battery be administered. The HRB is comprised of a set of individual tests that are to

be given as a battery, which is necessary to utilize pattern analysis (Schludermann et al., 1983). The HRB is often not used in this manner, which results in the loss of information about deficits. The pattern analysis approach is contrasted with the hypothesis testing method. Here, a neuropsychologist selects and administers tests based on the history and referral question, or hypothesis. Tests are viewed as a way to gather the necessary information to address the needs of the individual (Schludermann et al., 1983). Reitan and Wolfson (1996) cite the advantages of a fixed battery over a flexible one. They highlighted that fixed batteries are validated as compared to flexible batteries, which are casually composed. Also, because test choices are based on patient self-report, they are subject to biases and may serve only to confirm the patient's self-diagnosis (Reitan & Wolfson, 2004).

The HRB uses several different tests to gather information about the brain's functional status. They are divided into five distinct categories: input measures; tests of verbal abilities; tests of spatial, sequential, and manipulation abilities; measures of abstraction, reasoning, and concept formation; and output measures (Reitan & Wolfson, 1986a). Overall, the HRB looks at brain-related abilities and central processing using a hierarchical methodology. Reitan and Wolfson (1986a) described the levels assessed by the HRB. The first level consists of measures of attention, memory, and concentration. These components are distributed throughout the measures of the HRB, but are particularly relevant in the SSPT and the Rhythm Test (Reitan & Wolfson, 2004). The second level of central processing is concerned with the differential functions of the two brain hemispheres. Certain tests rely on the ability of one hemisphere over the other (Reitan & Wolfson, 2004). For example, language and verbal measures usually depend

more on the left hemisphere, whereas spatial relationships normally utilize the right hemisphere more. The highest level of processing examines reasoning and concept formation. The best test of this within the HRB is the Category Test, which contributes to its high sensitivity to brain damage (Reitan & Wolfson, 2004).

The HRB has implemented several measurement strategies in order to identify brain damage. The level of performance is how well the patient performs on each individual test. Due to the variability of performance among subjects, a single level of performance strategy does not accurately diagnose cerebral damage (Reitan & Wolfson, 2004). However, one can use the patient's performance on individual tests to note impaired brain functions. Another approach involves identifying deficits that occur primarily among those with brain damage and are rarely seen in normal individuals, known as pathognomonic signs. This also can yield information about localization (Reitan & Wolfson, 2004). Finally, the HRB examines sensory and motor functioning on both sides of the body using the individual as his/her own control. This yields information about each hemisphere as well as areas within them and denotes possible involvement of the two hemispheres, as performance differences between the two sides of the body suggest contra-lateral cerebral involvement (Reitan & Wolfson, 1993; Reitan & Wolfson, 2004).

Halstead (1947) introduced the HRB as a battery of tests that was able to differentiate brain-damaged individuals from those without such damage. One of the first Halstead studies demonstrated the effectiveness of the HRB by comparing individuals with and without frontal damage to a non-neurological control group. The results were highly significant and indicated that those with frontal damage performed more poorly

than those without, while both did worse than the control group. Overall, the study demonstrated the ability of the HRB to differentiate between neurological and non-neurological groups, as well as its ability to identify different types of brain impairment.

One of the first studies conducted by Reitan on the HRB compared patients with cerebral damage to a non-neurological control group matched on ethnicity and gender. They were also closely related in age and education. Results achieved a higher level of significance than in Halstead's original study. Strongly significant group differences were noted on 7 of the 10 tests, with the largest difference observed on the Category Test and the HII (Reitan, 1955a). These studies highlight the validity of the HRB at differentiating brain-damaged individuals from those without such damage.

These early findings are supported by more recent investigations. Goldstein and Shelly (1972) found that the HRB correctly classified brain damaged and non-brain damaged individuals at a rate of 71.06% with 27.56% false positives and 29.71% false negatives, which is significant beyond the .01 level. Additionally they found that the HRB was able to differentiate between those with lateral, diffuse, and no brain damage. Correct classification percentages of these individuals are 66.7% for the left hemisphere, 52.38% for the right hemisphere, 42.54% for diffuse damage, and 65.38% for those without brain damage. This is significant beyond the .001 level. Russell (1995) reviewed the available studies that examined the validity of the HRB and found that it has been repeatedly validated to identify the presence of brain damage. The accuracy of the indices ranged from 58 to 92%. However, the overall index was approximately 80% correct.

Halstead (1947) performed the initial factor analysis of the HRB, which resulted in the four factors of verbal learning, abstraction, attention, and perceptual/motor skills.

Goldstein and Shelly (1972) also examined the HRB and found four factors. These were language skills, which included SSPT and the Aphasia Test; perceptual skills whose major factors were finger agnosia and finger-tip number writing and whose minor factors were Trails B and SSPT; complex non-verbal problem solving that includes the Category Test, and TPT speed, memory, and location scores; and motor speed, which consisted of the Finger Tapping Test. Aftanas and Royce (1968) used factor analysis and obtained three factors: perceptual organization, which included the Category Test and TPT; perceptual motor speed that included Trails; and temporal perceptual resolution reflecting organic integrity. Overall these studies suggest that brain damage is reflected on psychological tests in a quantitative rather than qualitative way with the major area assessed being cognitive functioning. Testing reveals verbal and non-verbal skills reflecting distinct dimensions vulnerable to brain damage (Schludermann et al., 1983).

Subtests of the HRB. The Finger Tapping Test requires patients to use their index finger to tap as quickly as possible on a mechanical counter. The goal is to get five 10-second trials on each hand. The score is the average number of taps for the dominant and non-dominant hands. Overall this test is a measure of motor speed and coordination that yields information about the motor components of each cerebral hemisphere (Jarvis & Barth, 1984; Reitan & Wolfson, 1993; Reitan & Wolfson, 2004). Performance on motor tests such as the Finger Tapping test have been found to be less impaired than on other tests of brain function in AD samples (Bak & Greene, 1980; Butters et al., 1998; Reed & Reitan, 1963a; Reed & Reitan, 1963b).

The Grip Strength Test assesses strength using a hand dynamometer. Patients are asked to hold the instrument by their side and squeeze as hard as possible. The goal is to

get alternate trials using the dominant and non-dominant hands. The result is a total of two trials with each hand. There are noted gender differences in this subtest, with men having greater grip strength on average than women. The test is a measure of motor strength without consideration of any other behavioral components (Jarvis & Barth, 1984). It gives information regarding the motor areas of each hemisphere (Reitan & Wolfson, 2004). The neuropsychologist notes discrepancies between the two sides, which may be indicative of contralateral cerebral hemisphere dysfunction (Reitan & Wolfson, 1993). This test of motor strength is also less impacted by AD (Bak & Greene, 1980; Butters et al., 1998; Reed & Reitan, 1963a; Reed & Reitan, 1963b).

The Seashore Rhythm Test is adapted from the Seashore Tests of Musical Ability (Jarvis & Barth, 1984). In this test patients are asked to differentiate between pairs of rhythmic beats. They are to indicate "same" if the two beats are identical and "different" if they are not. The score is the number correct, which is used in calculating the HII (Jarvis & Barth, 1984). The test is designed to measure alertness to nonverbal stimuli, ability to maintain attention and concentration, and the ability to compare various rhythmic sequences (Jarvis & Barth, 1984; Reitan & Wolfson, 1993; Reitan & Wolfson, 2004). The test is an indicator of general cerebral functioning and has no significance with regard to lateralization (Reitan & Wolfson, 1993). Decrease in performance on this subtest suggests deficits of attention, concentration, and coordination (Jarvis & Barth, 1984). Dodrill and Dikmen (1978) found this measure differentiated between neurological and non-neurological groups without undue overlap with other tests. Reed and Reitan (1963a) found that on the Seashore Rhythm Test, subjects over 50 years of age performed better than the younger group, aged 40-49. Abilities required on this test,

attention to non-verbal stimuli and sustained attention, seemed well preserved in this dementia sample.

The Speech Sounds Perception Test (SSPT) uses 60 nonsense words to measure attention and concentration in first level central processing (Reitan & Wolfson, 1993). Patients listen to a recording of spoken words and are required to select the correct word from four alternative written words. The score is the number of errors; the score contributes to the calculation of the HII (Jarvis & Barth, 1984). This test shares commonalities with the Seashore Rhythm Test, but differences are that the SSPT is at a slower pace, it is simpler due to cues, and the stimuli are verbal in contrast to the Seashore Rhythm Test in which they are nonverbal (Jarvis & Barth, 1984). Bornstein (1982) found the split-half reliabilities for the SSPT were .74 and .87 for two independent samples, which correctly classified 96% and 90% of the samples. This test is a good discriminator of brain function regardless of type of lesion or disorder, independent of age and gender (Reitan & Wolfson, 1989). However, due to its dependence on attention and concentration, Reitan and Wolfson (1990) found that it is not as effective on those who have left cerebral damage.

Although it is not factored into the HII, the Trail Making Test is one of the most sensitive to brain damage partly because it requires the utilization of both right and left hemispheres (Reitan & Wolfson, 2004). It was originally a performance subtest of the Army Individual Test (Gaudino, Geisler, & Squires, 1995; Reitan, 1955b). It is composed of two parts, A and B. On part A of the test, patients are required to draw a line through 24 consecutive numbers, in circles and spaced randomly on a page, as quickly as possible. On part B, patients are given both numbers and letters. They are to draw a line

through consecutive numbers and alphabetical letters in an alternating sequence (i.e., 1-A-2-B-3-C...) as quickly as possible. On each section, the examiner immediately corrects errors and patients are instructed to begin again at the point of the error. Scores are the total time required to complete the task and number of errors made on the task. The test measures patients' scanning ability, visual attention, motor speed and coordination, and contains perceptual and problem solving requirements (Gaudino et al., 1995; Jarvis & Barth, 1984). Additionally, part B assesses number-letter recognition and flexibility in completing the alternating series under time pressure while maintaining attention on both aspects of the presented stimulus (Jarvis & Barth, 1984; Reitan & Wolfson, 1993).

In one of the first investigations of Trails, Reitan (1955b) matched brain damaged and non-brain damaged individuals on gender, ethnicity, age, and education and found that the test correctly differentiated between the groups at $p<.001$. Others have found that AD patients perform worse than normals on this test, even when obvious signs of social dysfunction are absent in the AD group (Amieva et al., 1998; Rasmusson et al., 1998). Lamberty, Putnam, Chatel, Bieliaukas, and Adams (1994) found that Trails differentiated between clinical and normal groups. The AD group performed more slowly than other groups and age was correlated with score. Boll and Reitan (1973) found that for both brain injured and non-brain injured, Trails is not associated with age, but does correlate with Verbal Performance and Full Scale IQ WAIS scores.

The Tactile Performance Test requires the use of a form board and ten blocks of various shapes. The patient is blindfolded and never sees the apparatus being used for the test. The task is to place the blocks in the board as quickly as possible. The first trial uses the dominant hand only, the second trial uses the non-dominant hand only, and the third

trial utilizes both hands. Following the third trial, the board is removed and the blindfold is removed. At this point, patients are asked to draw an outline of the board and the correct position of the blocks on the board. The test yields eight subtest scores, which are derived from time (TPT- Dominant, TPT- Non-dominant, TPT- Both, Total Time); number of blocks correctly placed (Blocks- Dominant, Blocks- Non-dominant, and Blocks-Both); and number of blocks correctly recalled (Memory) and located (Location) (Jarvis & Barth, 1984). The scores of Total Time, Memory, and Localization contribute to the HII. The TPT requires complex problem solving abilities and yields information about the intactness of the hemispheres. It enables the clinician to compare the efficacy of the two hemispheres and provides details regarding general brain functions (Reitan & Wolfson, 2004). This test specifically measures strength and speed of movement, abstraction, ability to utilize tactile perception, and ability to form a mental map of the board. It also assesses incidental memory since patients are not told that they will be asked to draw the board later (Jarvis & Barth, 1984; Reitan & Wolfson, 2004).

The Category Test is one of the most researched subtests of the HRB. It was developed from a card sorting task that Halstead found differentiated between normal brain injured subjects (Halstead, 1940). The original version contained 360 items in 9 subtests (Choca, Laatsch, Wetzel, & Agresti, 1997). The current test is comprised of 208 stimulus slides of geometric shapes and letters divided into 7 subtests that are serially projected onto a screen and increase in difficulty. Patients are told that each stimulus slide will remind them of a number between 1 and 4. They respond by pressing a button that corresponds to the suggested number. A correct answer yields a bell, while an incorrect answer results in a noxious buzzer (Boyle, 1986).

The patient is informed that The Category Test is divided into 7 subtests. The first subtest requires matching to Roman numerals while the second is based on the number of items presented (Reitan & Wolfson, 1993). Even individuals with severe brain impairments frequently perform the first and second subtests well, and these subtests have been found too easy to yield any information about the patient (Choca et al., 1997). Other criteria of subsequent subtests are uniqueness, quadrants, and proportions (Reitan & Wolfson, 1993). The score is comprised of the total number of errors on the test.

Overall, The Category Test is a measure of problem solving, judgment, abstract reasoning, concept formation, mental flexibility, and mental efficiency (Boyle, 1986; Jarvis & Barth, 1984). It requires the ability to note aspects of the stimulus material, postulate hypotheses about similarities and differences, use feedback about these hypotheses by receiving positive or negative information, and adapt future responses based on this feedback (Reitan & Wolfson, 2004). It is one of the best indicators of brain damage and the single most effective test of the HRB in detecting cerebral impairment (Reitan & Wolfson, 1993). This is primarily because of its measuring the subject's ability to alter performance based on positive and negative feedback (Reitan & Wolfson, 2004). It has been found that elderly normals and individuals with brain damage perform equally poorly and obtain similar pattern of deficits on the Category Test. This suggests that impairments seen in brain injury are also reflected in those with normal aging (Bigler et al., 1981; Mack & Carlson, 1978; Reitan & Wolfson, 1986b).

There are several variables that affect an individual's performance on the Category Test. Researchers have found that both age and education influence scores. Leckliter and Matarazzo (1989) calculated the correlation between age and test score as

.54, while the correlation between age and education was -.31. Additionally, there appears to be an interaction between the two. Heaton, Grant, and Matthews (1986) found that, prior to the age of 60, less educated individuals show more impairment, but after 60 all subjects perform equally poorly. Test-retest reliability of the Category Test is .60 for those without brain damage, perhaps due to practice effects. However, correlations for those with brain damage range from .82 to .96 (Choca et al., 1997; Matarazzo, Matarazzo, Wiens, Gallo, & Klonoff, 1976). Russell (1992) determined that test-retest correlations were .89, which indicates that the test is highly reliable. Similarly Kilpatrick (1970) found a split-half reliability of at least .90. Using factor analysis, Fischer and Dean (1990) identified three factors including attention and incidental memory. With regards to convergent validity, there are modest correlations with other assessments (Choca et al., 1997). Correlations with other HRB measures are .53 with the TPT, .58 with Trails B, and SSPT ranges from .22 to .43 (Choca et al., 1997; Goldstein & Shelly, 1972; Ryan, Larsen, & Prititera, 1978). One limitation of the Category Test is its excessive length and time of administration. Because of this limitation, researchers have attempted to construct an abbreviated version that retains the strong psychometric properties of the original test (Boyle, 1986; Russell & Levy, 1987).

The Aphasia Screening Test is a modification of the Halstead-Wepman Aphasia Screening Test and provides a comprehensive overview of aphasic and related deficits (Jarvis & Barth, 1984; Reitan & Wolfson, 1993). Using stimuli in a spiral bound book, patients are asked to perform several simple tasks such as naming objects, spelling, reading, writing, calculations, comprehending language, and copying figures (Jarvis & Barth, 1984; Reitan & Wolfson, 1993; Reitan & Wolfson, 2004). Results yield

information about dysnomia, dyslexia, spelling dyspraxia, dyscalculia, and constructional dyspraxia and may be affected by patient's educational or psychiatric status (Jarvis & Barth, 1984).The test is scored qualitatively. Due to the simple requirements of this test, performance failure on any item suggests the presence of cerebral impairment (Reitan & Wolfson, 1993).

Procedure

Participants were outpatients who received a comprehensive neuropsychological evaluation at The Neuropsychology Center in Dallas, Texas. Patients were referred for evaluations primarily to determine diagnosis or extent of brain impairment. Trained psychometrists who were supervised by licensed neuropsychologists administered all tests. Subjects were given a battery of tests including the MMSE and all subtests of the HRB and the WRAT. The HRB includes the Halstead Category Test, Trails A and B, TPT, SSPT, Seashore Rhythm Test, Finger Tapping, and Grip Strength. The Reading subtest was used from the WRAT. MMSE scores were used as a measure of severity of brain impairment.

Data Analysis

Collected archival data were analyzed to determine the ability of the HRB and WRAT-Reading subtest to detect differences in neuropsychological functioning in a clinical AD population. To test hypothesis one, data were analyzed separately by gender using a Pearson correlation to determine the relationship between each subtest and MMSE score. The rankings were divided into six sets of three subtests. This is a convenient method to facilitate the comparisons of the subtests rankings. These correlation coefficients were then rank ordered from greatest magnitude, indicating the

strongest "don't hold" test, to smallest, indicating the strongest "hold" test. Regarding hypothesis two, MLR was used to determine whether a smaller subset of tests is sufficient to predict MMSE score. For the third hypothesis, MLR and discriminant function analysis (DFA) determined how well the WRAT and HRB subtests classified participants in the two severity groups. The stepwise discriminant analysis tested the ability of the selected subtests to adequately classify participants into the two severity groups. In this stepwise discriminant analysis, the *F* to enter a predictor variable was set at $p<.10$. The fourth hypothesis used an independent samples t-test to determine gender differences with regard to subtest scores.

Several statistical procedures were employed to test the hypotheses. First, a Pearson correlation was used to assess the relationship between each subtest and MMSE score. This correlation was used to rank "hold/ don't hold" tests by gender. Second, it was determined which "don't hold" tests are significant predictors of severity status as measured by the MMSE. This was accomplished by using all of the HRB and WRAT subtests as predictors in a stepwise regression model with the MMSE as the criterion variable. Subtest scores were added into the model until they no longer provided a pre-selected increment ($p<.10$) in the prediction of the dependent variable, MMSE. Subtests that did not aid in prediction were deleted from the regression model. These analyses were conducted separately by gender.

Third, stepwise discriminant analysis was used to determine a smaller subset of "don't hold" subtests that were significant predictors of severity group membership. Using the MMSE score distribution, two severity groups were selected: normal/ mildly impaired and moderately/ severely impaired. The normal/ mildly impaired group

consisted of respondents with MMSE scores between 30 and 21. The moderately/ severely impaired group consisted of participants with MMSE scores below 20. Then, the results of the discriminant analysis were used to classify participants into the two severity groups. The proportion of hits, an accurate classification, and misses, those participants who were incorrectly classified, were recorded.

Discriminant analysis tested whether groups differ on the mean of a subtest; then that information was used to predict group membership (Hair, Anderson, Tatham, & Black, 1995). Dummy coding was used to identify group membership. The severity group is a dichotomous, mutually exclusive, and exhaustive dependent variable and the HRB and WRAT subtests are the independent variables. The subtests scores were used to calculate a discriminant model that optimally classified AD patients into two groups, normal/mildly impaired and moderately/severely impaired. The method used involves computing the discriminant function so that all of the independent variables are considered in a stepwise fashion (Hair et al., 1995). The "hit" rate is the percentage of patients that were correctly classified into severity groups. The discriminant function yielded weighted *F* statistics, which gave the significance level of the discriminant function overall, and for each subtest. Subsequently, subtests were grouped in order of decreasing ability to differentiate between severity groups. From severity group, change in cognitive status can be inferred.

The result is a description of the relationship between severity group, which is a single, nonmetric, categorical dependent variable, and the subtests of the WRAT and HRB, which are metric independent variables. The groups were *a priori* defined by the MMSE and the analyses derived the linear combination of variables that best

discriminates these groups. Predictive accuracy, which is known as a "hit", is determined by how many participants the subtest correctly classifies into groups. The hit ratio is the percentage that was correctly classified and indicates how well the discriminant function identified group membership. The hit ratio is conceptually equivalent to R^2 and indicates how much variance was explained by the regression equation. Stepwise analyses were used, which means that the independent variables were entered in a predetermined sequence according to the greatest increment to predictability. Statistical significance was tested, as is conventional, by Wilks' lambda.

Fourth, the data examined gender differences with regard to test performance. An independent samples t-test was used to show where significant gender differences existed regarding specific subtests. An alpha level of .005 was used in all analyses to determine gender differences. This more conservative alpha level was used, rather than the conventional alpha of $p < .05$, to adjust for the possibility of incremental Type I error due to the use of multiple t-tests.

Hypotheses

Hypothesis One

The first hypothesis stated that the HRB and WRAT subtests will be a combination of "hold" and "don't hold" tests.

Hypothesis 1A. With regard to the HRB subtests, it was hypothesized that tests that assess abstraction and problem solving (e.g., the Category Test) would be most affected by severity. Therefore, it was hypothesized that the Category Test would be one of the strongest "don't hold" test for both males and females.

Hypothesis 1B. Tests that rely heavily on attention and concentration (e.g., SSPT, Rhythm Test, and Trails A) were hypothesized as the second most compromised group of subtests for both males and females.

Hypothesis 1C. Tests that assess incidental memory (e.g., TPT-Memory) were hypothesized as the third most compromised group of subtests for both males and females.

Hypothesis 1D. In contrast, it was hypothesized that the motor tests of Grip Strength and Finger Tapping would be strong "hold" tests for both the dominant and non-dominant hands for both males and females.

Hypothesis 1E. It was further hypothesized that the WRAT-Reading subtest would be a strong "hold" test for both males and females.

Hypothesis Two

The second hypothesis stated that a smaller subset of "don't hold" tests would be sufficient to predict severity status as measured by MMSE score. It was hypothesized that the Category Test would be a significant predictor along with at least one of the TPT subtests. It was predicted that the motor tests would not be significant predictors of MMSE. It was also predicted that the subset of sufficient tests would differ based on gender. This hypothesis was tested using stepwise MLR analyses separately for both genders.

Hypothesis Three

The third hypothesis was that a smaller subset of "don't hold" tests would be significant predictors of severity group, and these would vary depending on gender. Two groups were established using MMSE score, normal/ mildly impaired and moderately/

severely impaired. The results of the discriminant analysis were used to classify participants into two severity groups. This hypothesis was tested using stepwise MLR and discriminant analysis.

Hypothesis Four

The fourth hypothesis was that there would be significant differences between males and females in test performance on certain of the HRB subtests. Specifically, females would have lower scores on Grip Strength and Finger Tapping, with both their dominant and non-dominant hands. An independent samples t-test was conducted to identify specific subtests that had significant gender differences with regard to test performance. An alpha level of .005 was used as the criterion for significance to account for possible incremental Type I error using multiple t-tests.

Chapter 3: Results

The archival data were prepared for analysis by deleting participants with missing subtest scores. A total of 16 participants were deleted. Severity of impairment is only one of many reasons that participants might have incomplete data. Given this, it is incorrect to assume that a participant did not take a test because he or she was too impaired. Therefore, analyses were conducted using only those participants who had completed all HRB and WRAT subtests. The TPT scores for dominant, non-dominant, and both hands were converted to a score of total time. The scores using blocks for the dominant, non-dominant, and both hands were converted to a total number of blocks. These two scores were divided to calculate a score representing minutes per block.

When using statistical procedures, several factors may negatively impact calculation and interpretation. Multicollinearity occurs when there is a high correlation between several of the predictor variables that are theoretically independent. Multicollinearity becomes a problem if several of the variables significantly overlap; this indicates that the predictor variables are not independent. The statistic used to measure multicollinearity is tolerance. Tolerance is the proportion of variance associated with a selected predictor variable that is not due to the other predictor variables (Hayes, 1994). Variables with low tolerance add little information to the prediction model. An examination of tolerance statistics for all subtests (see Table 2) reveals that there are two subtests, the Rhythm Test and Trails A, with

tolerance levels in the .60s, which can be problematic. However, the tolerances of these two variables are in the high to mid .60s and still account for at least 60% of variance not predicted by other variables. Thus, it was judged reasonable to include all of these subtests in the final analysis. The tolerance of each of the HRB and WRAT subtests is presented in Table 2. The intercorrelation matrix for the subtests is in Table 3.

Other factors to consider when analyzing data are the skewness and kurtosis of the distribution. The skewness refers to the degree of asymmetry of a distribution. When one side of a distribution contains a higher frequency of scores than the other, the distribution is skewed. Skewness statistics below –1 and above +1 can be problematic. Kurtosis describes the shape of a distribution; high kurtosis indicates that there is not much spread within the scores. A leptokurtotic plot is very narrow and peaked. A platykurtotic plot is wide and low, reflecting scores low in kurtosis. In the present sample several subtests, TPT-Minutes/Block, TPT-Location, SSPT, Trails A, and Trails B, have skew outside the ideal range. However, the skewness statistics for all variables but Location are close to 1. Location is significantly skewed because it has a restricted range of 0 to 4 even though the range of possible scores is 0-10. Additionally, 77% of participants scored 0 on this subtest. These scores emphasize the significant problems that these AD participants had on this subtest. It is probably due to the large memory component of the test, an ability that is significantly compromised in AD patients. The information concerning the skewness and kurtosis of the present distribution is given in Table 2. Overall, the kurtosis levels are satisfactory, with Location being the only significantly affected subtest.

Table 2

Descriptive Statistics of HRB and WRAT Subtests

Subtest	Mean	Standard Deviation	Skewness	Kurtosis	Tolerance
WRAT-Reading	93.93	13.98	-.45	-.33	.93
Category Test	114.30	22.72	-.29	-.71	
TPT-Minutes/Block	6.78	8.49	2.28	4.95	.79
TPT- Memory	2.25	2.02	.79	-.15	.77
TPT- Location	.28	.60	2.90	11.49	.92
Rhythm Test	14.97	8.04	.01	-.86	.68
Speech Sounds Perception Test	19.16	11.12	1.13	1.24	.77
Finger Tapping-Dominant	39.68	8.54	-.22	.14	.87
Finger Tapping-Non-dominant	37.13	7.16	-.55	.35	.92
Trails A	103.99	81.17	1.55	1.38	.65
Trails B	273.95	77.21	-1.69	1.71	.82
Grip Strength - Dominant	27.66	11.06	.98	.64	.95
Grip Strength - Non-dominant	24.67	10.36	.87	.54	.95
MMSE	19.04	5.03	-.47	-.02	

Table 3

Pearson Product-Moment Intercorrelations between subtests of the HRB and WRAT For the Full Sample (N = 151)

Subscale	1	2	3	4	5	6	7	8	9	10	11	12	13
1. WRAT	1.0	-.26	-.04	.05	.03	.36	-.56	-.41	-.20	.28	.27	-.09	.13
2. Cat Test		1.0	.46	-.48	-.29	-.57	.48	-.59	.42	-.36	-.28	-.22	-.22
3. TPT Minute/ Block			1.0	-.36	-.25	-.36	.29	.52	.21	-.22	-.15	-.19	-.15
4. TPT- Memory				1.0	.47	.42	-.24	-.27	-.39	.19	.07	.15	.11
5. TPT- Location					1.0	.28	-.18	-.20	-.18	.04	.04	.19	.17
6. The Rhythm Test						1.0	-.50	-.54	-.38	.33	.21	.21	.19
7. Speech Sounds Perception Test							1.0	.68	.28	-.34	-.29	-.08	-.06
8. Trails A								1.0	.37	-.35	-.33	-.14	-.12
9. Trails B									1.0	-.23	-.13	-.06	-.08
10. Finger Tapping-Dominant										1.0	.79	.39	.35
11. Finger Tapping-Non-dominant											1.0	.40	.40
12.Grip Strength- Dominant												1.0	.91
13. Grip Strength-Non-dominant													1.0

Hypotheses

Hypothesis One

The first hypothesis stated that the HRB and WRAT subtests would be a combination of "hold" and "don't hold" tests.

Hypothesis 1A. With regard to the HRB subtests, it was hypothesized that tests that assess abstraction and problem solving (e.g., the Category Test) would be among those most affected by severity. Therefore, the Category Test would be one of the strongest "don't hold" test for both males and females. Hypothesis 1A was supported. The Category Test was found to be a strong "don't hold" test for both males ($r = -.64$) and females ($r = -.52$), ranking 2nd highest correlation for males and 3rd highest for females. This places the Category Test in the first set of subtests for both genderes. However, it was not the strongest "don't hold" subtest for either gender.

Hypothesis 1A was tested separately for males and females. Pearson correlations were used to determine the relationship between each WRAT and HRB subtest and MMSE score. The results show that the rankings for males, from most to least discriminating are: Category Test, TPT-Memory, Trails A, Trails B, TPT-Minutes/Block, SSPT, Rhythm Test, WRAT-Reading, TPT-Location, TPT-Both, Grip- Non-dominant, Tapping-Dominant, Grip-Dominant, and Tapping- Non-dominant. The rankings for females are: Rhythm Test, SSPT, Category Test, Trails A, WRAT-Reading, TPT-Memory, TPT-Location, Trails B, TPT-Minutes/Block, Tapping-Dominant, Tapping-Non-dominant, Grip-Non-dominant, and Grip-Dominant. To facilitate understanding the

results, the rankings have been equally divided into six sets of three subtests each. The rankings for both males and females are given in Table 4.

Hypothesis 1B. Tests that rely heavily on attention and concentration (SSPT, Rhythm Test, and Trails A) will be the second most compromised group of subtests for both males and females. Hypothesis 1B was partially supported. For males, the three subtests fall in the middle of the rankings. Trails A score is based on time and is ranked 3rd (r = -.51), SSPT was 6th (r = -.46), and the Rhythm Test ranked 7th (r = .44). This places Trails A in the first set of rankings, while SSPT is in the second set and the Rhythm Test is in the third set of rankings. Although SSPT was in the middle of the rankings, Trails A and the Rhythm Test were not. For females, the hypothesis was not supported. The Rhythm Test was ranked 1st (r = .57), SSPT was ranked 2 (r = -.52), and Trails A was ranked 4th (r = -.51). This places both the Rhythm Test and SSPT in the first set of rankings, while Trails A is in the second set. Findings suggest that only Trails A was in the second most compromised group for females. These subtests were consistently "don't hold" tests for both groups.

Hypothesis 1C. Tests that assess incidental memory (TPT-Memory) will be the third most compromised group of subtests for both males and females. Hypothesis 1C was not supported. For males, TPT-Memory was one of the strongest "don't hold" subtests, with a ranking of 2nd (r = .52). This places it in the first set of subtest rankings. For females, the hypothesis was supported as Memory was ranked 6th (r = .37), placing it in the third most compromised group overall.

Table 4

HRB and WRAT Subtest Rankings by Gender Using Pearson r Correlation With MMSE Score

Males (N= 46)				Females (N= 105)			
Subtest	Coefficient	Ranking		Subtest	Coefficient	Ranking	
---	---	--	--	---	---	---	Set 1
Category Test	-.64**	1		Rhythm Test	.57**	1	
TPT- Memory	.52**	2		Speech Sounds Perception Test	-.52**	2	
Trails A	-.51**	3		Category Test	-.52**	3	
---	---	--	--	---	---	---	Set 2
Trails B	-.50**	4		Trails A	-.51**	4	
TPT-Minute/ Block	-.48**	5		WRAT-Reading	.41**	5	
Speech Sounds Perception Test	-.46**	6		TPT- Memory	.37**	6	
---	---	--	--	---	---	---	Set 3
Rhythm Test	.44**	7		TPT- Location	.28**	7	
WRAT-Reading	.40**	8		Trails B	-.25*	8	
TPT- Location	.25	9		TPT-Minute/ Block	-.23*	9	
---	---	--	--	---	---	---	Set 4
Grip Strength- Non-dominant	.10	10		Finger Tapping-Dominant	.18	10	
Finger Tapping-Dominant	.09	11		Finger Tapping-Non-dominant	.13	11	
Grip Strength-Dominant	.06	12		Grip Strength-Non-dominant	-.04	12	
---	---	--	--	---	---	---	Set 5
Finger Tapping-Non-dominant	-.05	13		Grip Strength -Dominant	-.02	13	

* p<.05 **p<.01

Hypothesis 1D. It was hypothesized that the motor tests of Grip Strength and Finger Tapping will be strong "hold" tests for both the dominant and non-dominant hands for both males and females. Hypothesis 1D was supported in both groups. For males Grip Strength placed 10^{th} (r = .10) and 12^{th} (r = .06) while for females it was ranked 12^{th} (r = -.04) and 13^{th} (r = .02). This places it in the fourth and fifth sets for males. For females, Grip Strength also placed in the fourth and fifth sets of subtests. This indicates that it was a strong "hold" test for both groups. Finger Tapping was ranked 11^{th} (r = .09) and 13^{th} (r = -.05) for males. It was also a strong "hold" test for females and was ranked 10^{th} (r = .18) and 11^{th} (r = .13), making it slightly less of a "hold" test as compared to the male sample. Finger tapping placed in the fourth and fifth sets for males and in the fourth set for females. Overall, Finger Tapping was a strong "hold" test for both groups.

Hypothesis 1E. It was also hypothesized that the WRAT-Reading subtest will be a strong "hold" test for both males and females. Hypothesis 1E was not supported. In the male sample the WRAT was a moderately strong "don't hold" test (r = .40), ranked 8^{th}. For the female sample the WRAT was ranked 5^{th} (r = .41), suggesting that it is also a moderately strong "don't hold" test. This places it in the third set for males and in the second set for females. Although the rankings are different, the correlations indicate that the direction and magnitude of the relationship between the WRAT and MMSE are equivalent for both.

Hypothesis Two

The second hypothesis stated that a smaller subset of "don't hold" tests would be sufficient to predict severity status as measured by MMSE score. It was hypothesized that the Category Test would be a significant predictor along with at least one of the TPT

subtests. It was predicted that the motor tests would be least able to predict MMSE. It was also predicted that the subset of sufficient tests would differ based on gender. The second hypothesis was partially supported. For males, the strongest predictors of MMSE were the Category Test, TPT-Memory, and Trails A, which contributed significant variance to the prediction model. The summary of the regression model for males is in Table 5. For females there were four subtests that significantly predicted MMSE score within the regression model: the Rhythm Test, the Category Test, and WRAT-Reading. SSPT was included, but was deleted from the model in step 4 due to shared variance with other subtests. The summary of the regression model for females is presented in Table 6.

Hypothesis Three

The third hypothesis was that a smaller subset of "don't hold" tests would be significant predictors of severity group, and these would vary depending on gender. The third hypothesis was partially supported. Examining the results using the full sample shows that participants in the normal-mildly impaired group were correctly classified 49 out of 63 times, a 77.8% hit rate. For participants in the moderately-severely impaired range, correct classification occurred 67 out of 88 times, a 76.1% hit rate.

A cross-validation was performed using a leave-one-out procedure. This is useful when an independent sample is not available for cross-validation. The classification is based on all cases but one and that one is based on the discriminant function. This procedure is repeated until all cases have been left out once. The results allow an estimate of the accuracy of classification on a new sample, when a totally new hold-back sample is not available.

Table 5

Summary of Multiple Linear Regression Analysis Using HRB and WRAT Subtests to Predict MMSE Score for Male Participants (N=46)

Variable	Beta	Standard Error Beta	Standardized Beta
Step 1			
Category Test	-.15	.03	-.64
Step 2			
Category Test	-.12	.03	-.50
TPT-Memory	.83	.37	.28
Step 3			
Category Test	-.08	.03	-.33
TPT-Memory	.97	.36	.33
Trails A	-.02	.01	-.29

Note. R^2= .39 for Step 1; ΔR^2= .06 for Step 2; ΔR^2= .06 for Step 3 (*ps* <.05).

For the cross-validated sample, participants in the normal-mildly impaired group were correctly classified 47 out of 63 times, which is a 74.6% hit rate. For participants in the moderately-severely impaired range, correct classification occurred 67 out of 88 times, a 76.1% hit rate. The results of the discriminant analysis of the full sample are in Table 7.

Table 6

Summary of Multiple Linear Regression Analysis Using HRB and WRAT Subtests to Predict MMSE Score for Female Participants (N= 105)

Variable	Beta	Standard Error Beta	Standardized Beta
Step 1			
Rhythm Test	.34	.05	.57
Step 2			
Rhythm Test	.24	.06	.40
Speech Sounds Perception Test	-.13	.04	-.31
Step 3			
Rhythm Test	.18	.06	.30
Speech Sounds Perception Test	-.11	.04	-.26
Category Test	-.04	.02	-.21
Step 4			
Rhythm Test	.15	.06	.25
Speech Sounds Perception Test	-.06	.04	-.15
Category Test	-.06	-.02	-.25
WRAT-Reading	.07	.03	.19
Step 5			
Rhythm Test	.17	.06	.28
Category Test	-.07	.02	-.30
WRAT-Reading	.09	.03	.26

Note. R^2= .32 for Step 1; ΔR^2= .07 for Step 2; ΔR^2= .02 for Step 3; ΔR^2= .02 for Step 4; ΔR^2= -.01 for Step 5 (*ps* <.05).

Looking at the results of the sample divided by gender indicates that males who are in the normal-mildly impaired group were correctly classified 22 out of 25 times, an 88% hit rate. Males in the moderately/severely impaired group were correctly classified 17 out of 21 times, an 81% hit rate. In the cross-validated sample, participants in the normal- mildly impaired group were correctly classified 21 out of 25 times, an 84% hit rate. For participants in the moderately-severely impaired range, correct classification occurred 17 out of 21 times, an 81% hit rate.

Table 7

Summary of Discriminant Analysis Using the HRB and WRAT Subtests to Classify Severity Group Membership in an Undivided Sample

	Normal/mildly impaired group			Moderate/severely impaired group		
	Correctly Classified	Total	Percentage Correct	Correctly Classified	Total	Percentage Correct
All participants	49	63	77.8%	67	88	76.1%
Cross-validation	47	63	74.6%	67	88	76.1%

Females who are normal-mildly impaired were correctly classified 30 out of 38 times, a 78.9% hit rate. For females in the moderately/severely impaired group, correct classification occurred 51 out of 67 times, a 76.1% hit rate. In the cross-validated sample, participants in the normal-mildly impaired group were correctly classified 30 out of 38 times, a 78.9% hit rate. In the moderately-severely impaired range, correct classification occurred 50 out of 67 times, a 74.6% hit rate. The results of the discriminant analysis for the sample divided by gender are given in Table 8.

Table 8

Summary of Discriminant Analysis Using the HRB and WRAT Subtests to Classify Severity Group Membership in a Sample Divided by Gender

	Normal/mildly impaired group			Moderate/severely impaired group		
	Correctly Classified	Total	Percentage Correct	Correctly Classified	Total	Percentage Correct
Males	22	25	88%	17	21	81%
Females	30	38	78.9%	51	67	76.1%
Cross-validation						
Males	21	25	84%	17	21	81%
Females	30	38	78.9%	50	67	74.6%

Hypothesis Four

The fourth hypothesis was that there would be significant differences between males and females in test performance on certain of the HRB subtests. Specifically, females would have lower scores on Grip Strength and Finger Tapping, with both their dominant and non-dominant hands. The fourth hypothesis was supported. Results show that test performance was significantly affected by gender. Grip and Finger Tapping, both dominant and non-dominant, were significant at the .005 level, with males having higher scores. However, the Category Test, the Rhythm Test and TPT-Minutes/Block had a non-significant trend with male-superior performance on the Rhythm Test and TPT-Minutes/ Block. All of the t-scores for gender differences on the HRB and WRAT are given in Table 9.

Table 9

Summary of an Independent Samples t-test for Gender Differences on HRB and WRAT Subtests

Subtest	*p* value
WRAT-Reading	.22
Category Test	.05*
TPT-Minute/ Block	.08*
Memory	.52
Location	.40
Rhythm Test	.07*
SSPT	.98
Tapping-Dominant	.00**
Tapping- Non-dominant	.00**
Trails A	.37
Trails B	.78
Grip-Dominant	.00**
Grip- Non-dominant	.00**

*- non-significant trend, **- significant at the .005 level
All significant subtests had male-superior performance. Due to the problem of incremental t error, $p<.005$ was selected as criterion for statistical significance.

Chapter 4: Discussion

The discussion of the current study begins with its major purpose: to determine the effectiveness of the HRB and WRAT-Reading subtests to assess differences in illness severity using a sample of AD patients. Four hypotheses concerning the following were held: (1) the ranking of the HRB and WRAT subtests using a "hold/ don't hold" continuum; (2) whether a subset of "don't hold" subtests will predict severity; (3) whether subtests can accurately classify participants into severity groups; (4) gender differences in test performance. The four formal hypotheses are then introduced and discussed individually. A general discussion of the results follows, highlighting the significant findings and implications. This will be followed by a discussion of the limitations of the study. Finally, suggestions for future research will be given.

General Overview

The results of this study indicate that the HRB is a combination of "hold/don't hold" tests. Specifically, the Category Test was one of the strongest "don't hold" tests and motor tests were strong "hold" tests for both males and females. However, in many cases the "hold" or "don't hold" status of subtests varied by gender. For females, Grip Strength was a stronger "hold" test than Finger Tapping, but these subtests were equivalent in males. In general, subtests measuring attention/concentration (e.g., SSPT, Rhythm Test, and Trails A) were not

as strong "don't hold" tests in males compared to females. Also, TPT-Memory appears to be more related to severity for males. Additionally, the WRAT-Reading test was a "don't hold" test for both males than females. Examining overall test performance suggests that it is influenced by gender, with motor subtests being the most affected. However, there was an interesting non-significant trend for male-superior performance on the Rhythm Test and TPT-Minutes/Block. Additionally, there was a non-significant trend for female-superior performance on the Category Test. The cause of this trend toward gender differences is unclear. It is indeterminable whether this reflects a true difference in test performance or is due only to the characteristics of this sample.

Other results indicate that a smaller subset of "don't hold" tests is sufficient to predict severity. The subtests vary by gender, with the Category Test and TPT-Minutes/Block being significant for males. For females there were three subtests that significantly predicted MMSE score: the Rhythm Test, the Category Test, and WRAT-Reading. Additionally, subtest scores accurately classified participants into severity groups, particularly for males in the normal/mildly impaired group and for males and females in the moderately/severely impaired group.

Hypothesis One

The first hypothesis is a general hypothesis followed by five sub-hypotheses. The general hypothesis states that the HRB will be a combination of both "hold" and "don't hold" tests. Hypothesis 1A states that the Category Test will be one of the strongest "don't hold" subtests for both males and females. The results show that the Category Test was one of the strongest "don't hold" subtests, ranking 1st ($r = -.64$) for males and 3rd ($r = -.52$) for females. The correlation between the Category Test and MMSE is negative,

indicating that as errors on the Category Test increase, MMSE score decreases. Specifically, deficits in abstraction and problem solving, as measured by the Category Test, are strongly related to MMSE score in this AD sample. This relationship is slightly stronger for males; however, the differences between males and females are not significant. Consistent with the literature (Reitan & Wolfson, 2004) the Category Test is a strong indicator of brain impairment and is closely related to severity status as measured by the MMSE.

The Category Test is an accurate measure of abstraction, reasoning, and logical analysis. Individuals with memory problems, such as those with AD, often have poor performance on this test (Reitan & Wolfson, 2004). Impairment of memory and reasoning skills is evident in AD patients early in the disease process, resulting in poor neuropsychological test performance (Barth & Macciocchi, 1986; Lezak, 1995). There is a dearth of research using the Category Test with AD patients. Due to its reliance on higher order cognitive processes, the Category Test is particularly difficult for individuals with AD. Consistent with the present results, Storrie & Doerr (1980) found the Category Test the most difficult subtest for AD patients within the HRB. Additionally, these deficits are seen in daily activities of individuals with AD, including problems with organization, dealing with novel situations, and in recognizing cause and effect relationships. Hom (1992) found that the Category Test had the largest discrepancy in test performance between AD patients and normal elderly.

Studies show that the Category Test is sensitive to both brain impairment as well as the effects of normal aging. Reed and Reitan (1963a) originally examined the hypothesis that age-related changes in neuropsychological test performance were the

result of physiological changes in the brain. Overall test performance of older normal subjects mimicked the scores of younger participants with brain damage, suggesting an underlying organic process (Reed & Reitan, 1963a). The Category Test is one of the most sensitive to the organic condition of the brain and Subtests III and IV were particularly problematic for this group, probably due to the degree of task complexity (Mack & Carlson, 1978).

Looking at tests that differentiate elderly normals from younger normals, Reed and Reitan (1963b) found that the Category Test was accurate in its discrimination. The results suggest significant limitations in adaptive abilities in the older group, whereas tasks relying on stored memory were less able to distinguish the groups (Reed & Reitan, 1963b). Additionally, Prigatano and Parsons (1976) found that age is correlated with several HRB subtests, including the Category Test, and performance is negatively impacted by aging.

The present sample primarily consists of elderly individuals with a mean age of 73 diagnosed with AD. The Category Test is impacted by both age and brain dysfunction. It is unclear how each factor has influenced participants' performance in this study. Overall, these findings highlight the strong relationship between errors on the Category Test and MMSE, supporting its status as a strong "don't hold" test and the continued use of the HRB Category Test in the identification of cognitive impairment.

Hypothesis 1B stated that subtests measuring attention and concentration, including SSPT, the Rhythm Test, and Trails A, would be the second most compromised group of subtests. Results show that for males, the subtests were ranked in sets 1, 2, and 3 suggesting that they are moderately strong "don't hold" tests. Given this variability, they

were not the second most compromised group overall. For male participants, Trails A ranked 3rd, SSPT ranked 6th, and the Rhythm Test ranked 7th. In females, the Rhythm Test ranked 1st and SSPT ranked 2nd, making them the strongest "don't hold" tests. Trails A was less correlated with severity than the others and was ranked 4th.

These findings indicate that these subtests were strong "don't hold" tests for both groups and point to a strong relationship between deficits in attention-concentration and MMSE score, particularly for female participants. Tests assessing attention-concentration were the two strongest "don't hold" subtests for females. Male participants had a weaker relationship between the subtests and MMSE score, suggesting that performance on these tests is less affected by increasing severity deficits. All subtests measuring attention-concentration were moderately related to severity, being significant at the .01 level.

Physiological changes associated with AD result in significant decline in attention/concentration and related cognitive areas. Although memory impairment is evident early on in the process of AD, decline in attention and concentration soon follows. This may be the first indication of neocortical dysfunction (Parasuraman & Haxby, 1993). Attention deficits are important because they are likely the cause of deficits in daily living skills, seen even in mildly demented AD patients (Perry & Hodges, 1999). These are some of the most problematic deficits for both AD patients and their caregivers. They are often the first undeniable signs that something is wrong.

Similar to research on memory, attention can be divided into sub-types, which are differentially affected by the disease process (Parasuraman & Haxby, 1993; Perry & Hodges, 1999). Although divided attention, set shifting, and response selection are significantly affected during AD, sustained attention appears to remain well preserved in

early stages (Perry & Hodges, 1999). The HRB, the SSPT and Rhythm Test require sustained attention to stimulus materials, which has been shown to be less affected by severity (Perry & Hodges, 1999; Reitan & Wolfson, 1993). In this area, gender differences were found. For females, measures of sustained attention were the most strongly associated with severity status. This is in contrast to Perry and Hodges (1999) who found that this type of attention was not affected until later in the disease process. For males, the opposite was found. Trails A ranked significantly higher than SSPT and the Rhythm Test, suggesting that these two subtests were less affected by severity, in support of Perry and Hodges (1999).

Trails A assesses attention-concentration through visual scanning ability under time pressure (Reitan & Wolfson, 1993). Overall, problems with attention in AD patients are supported by the present study, which finds that aside from the complex tasks of the Category Test, subtests relying on attention-concentration are among the most compromised groups. There are slight gender differences evident in the current study, but they are not pronounced. It appears that attention/concentration may be slightly more related to severity for females.

The majority of research on HRB subtests assessing attention-concentration involves examining the Trail Making Test. Lafleche and Albert (1995) compared AD patients and controls on several tests, including Trails. They found that Trails A did not differentiate the groups, but Trails B did. This suggests that skills assessed by Trails B, such as concurrent manipulation of information, were significantly compromised by AD. Also, the visual scanning ability needed in Trails A was not declined in those with AD. This finding was not supported by the present study, which found that Trails A

discriminated between the two severity groups better than Trails B, for both males and females. The difference in results may be due to combining normal and mild individuals in the current study, while Lafleche and Albert (1995) compared AD patients to normal controls.

AD patients have been found to take longer to complete and commit more errors on Trails than normal or other clinical populations, including TBI and neuropsychiatric patients (Amieva et al., 1998; Lamberty et al., 1994; Rasmusson et al., 1998). Heun et al. (1998) found that Trails successfully discriminated between demented and non-demented controls. Chen et al. (2000) demonstrated that Trails is useful in predicting those who would later develop dementia, which implies early executive dysfunction including problems with attention/concentration. Longitudinal data show that Trails declines more than other neuropsychological measures during a 10 year period in those identified as cognitively impaired (Ratcliff et al., 2003). Johnstone et al. (2002) examined the neuropsychological deficit profile for AD and found that Trails B had the poorest performance, followed by Trails A; however, both had lower scores than tests assessing memory.

The current research emphasizes the problems that AD patients have on Trails, indicating significant deficits in these skills. Impairments in these areas are also found in the AD sample used in the current study. HRB subtests assessing abilities in attention-concentration reflect impairment early in the disease process and are some of the best subtests at discriminating between levels of severity, regardless of gender.

Hypothesis 1C stated that the HRB TPT-Memory subtest would be in the third most compromised group of subtests. Results show that for males, Memory was one of

the strongest "don't hold" subtests, with a ranking of second. This indicates a strong relationship between the Memory subtest and MMSE score for this group, in that as severity increases, or the MMSE score declines, Memory performance decreases. However, for males it was not the third most compromised subtest. For females, Memory was ranked sixth, placing it in the second most compromised group. This suggests a moderate relationship between Memory and severity; however, one that is still significant at the .01 level. These findings indicate that memory, as assessed by this subtest, is more closely related to MMSE score, or severity, in males than females.

Memory impairments are highly associated with the progression of AD and one of the first symptoms to emerge during the course of the disease (Green, 1995; Parasuraman & Haxby, 1993). The types of memories most affected by AD are those related to new learning and delayed recall (Green, 1995). Other types of memory deficits are working and secondary memory, especially when a distracter is involved (Lezak, 1995). Due to the overlap between mildly demented and normal aged individuals, it is difficult to determine whether deficits are organic or due to normal aging (Green, 1995).

The Memory subtest of the HRB is one of several generated by the TPT. Memory scores range from 0 to 10 and is simply the number of shapes correctly remembered. The subtest assesses incidental memory, as participants are unaware that they will be asked to draw the shapes after the first trials (Hom, 1992). The Memory subtest has been found to discriminate AD patients from elderly controls, suggesting that even very mildly demented participants have problems with incidental memory deterioration (Hom, 1992). This finding is supported by the present study, as Memory scores were found to be closely associated with severity for both males and females. However, Memory scores

were more strongly correlated with severity for males, indicating that they are more likely to have more significant deficits in memory as severity increases.

Johnstone et al. (2002) found that the neuropsychological deficit profile for AD included impaired memory, but other skills, such as speed of processing and cognitive flexibility, displayed more decline. This is in contrast to the present study, as Memory was correlated with severity more than Trails A and B for males and Trails B for females. However, Johnstone et al. (2002) used the Wechsler Memory Scales (WMS), which is a more comprehensive memory assessment than that used in the present study. The use of different measures of memory certainly affected the results and may explain the discrepancy between findings.

Memory is differentially affected during the AD disease process. Semantic memory is defined as remembering of facts and concepts and has been examined with regard to AD. Studies show that semantic memory is largely preserved in AD patients (Nebes et al., 1984; Ober et al., 1995). However, episodic memory, which is the recollection of events, is severely impaired in AD as compared to normal elderly participants (Nebes et al., 1984). Baillon et al. (2003) found that non-verbal memory is also impaired in AD patients, more significantly than another clinical population including those with vascular dementia. These studies demonstrate the various aspects of memory and how they can be affected differently by the disease process. AD is unique in the significant memory deficits that are seen. The present study examined incidental learning-memory in AD patients and found evidence that this type of memory is also impaired in this population.

Hypothesis 1D states that the motor tests of Grip Strength and Tapping will be "hold" tests for both males and females. Results show that Grip was a strong "hold" test for males. In males, Grip Strength for the non-dominant hand ranked 10^{th}, ($r = .10$). Grip Strength for males using the dominant hand ranked 12^{th}, ($r = .06$). Although Finger Tapping was also a strong "hold" test for males, for the non-dominant hand it ranked 13^{th}, ($r = -.05$). Using the dominant hand, it ranked 11^{th}, ($r = .09$).

Similar results were found for women, Grip Strength ranked 12^{th} for the non-dominant hand, ($r = -.04$). For females using the dominant hand, the ranking was 13^{th}, ($r = -.02$). Finger Tapping for the female group was ranked 11^{th} for the non-dominant hand, ($r = .13$). Finger Tapping for the dominant hand ranked 10^{th}, ($r = .18$). The reported correlations demonstrate the weak relationship between motor ability and MMSE score. This indicates that both Grip Strength and Finger Tapping are strong "hold" tests for both males and females. Whereas both motor tests appear equivalent for males, Grip Strength seems to be a slightly stronger "hold" test than Finger Tapping in the female group. Overall, this finding indicates that motor tests are strong "hold" tests and are largely unrelated to severity status for both males and females.

Results of the present study are consistent with previous literature regarding motor tests and AD patients (Barth & Macciocchi, 1986; Bigler et al., 1981; Storrie & Doerr, 1980). Although AD patients have decreased performance on other cognitive measures, tasks assessing motor abilities appear to be largely unaffected (Barth & Macciocchi, 1986; Bigler et al., 1981). Reed and Reitan (1963a) found that motor tests are strong "hold" tests, in that they do not discriminate between young controls and brain damaged subjects. This indicates that they remain intact despite brain impairment.

Results also suggest that motor ability is not influenced by age, as motor tests were unable to discriminate between old (mean age of 45) and older normal subjects (mean age of 55). However, it has been shown that as motor tasks become more complex, normal elderly subjects begin to demonstrate some impairment (Meyerink, 1982).

The extent to which AD patients retain motor functioning, despite increasing cognitive dysfunction, is disease-specific. Comparing AD patients with other illnesses affecting brain function, such as Huntington's disease and Multi Infarct Dementia, indicates those with AD retain greater motor skills (Butters et al., 1998; Russell & Polakoff, 1993). The present study supports this previous research suggesting that motor tests are not impacted by severity in AD patients, making them strong "hold" tests.

Hypothesis 1E stated that the WRAT-Reading test would be a strong "hold" test for both males and females. Contrary to hypotheses, results show that it is a strong "don't hold" test for males ranking 8^{th} ($r = .40$), significant at the .01 level. For females it also appears to be a strong "don't hold" test ranking 5^{th} ($r = .41$), significant at the .01 level. The findings indicate that there is a strong, positive correlation between MMSE score and WRAT performance. Therefore, as severity increases, or MMSE score declines, WRAT performance decreases. This indicates that reading ability, or WRAT score, is affected by severity. By definition, a "hold" test is a test that measures an ability that is less impacted by brain dysfunction. These results suggest that reading ability is impacted by increasing brain impairment in AD participants. Therefore, the WRAT is not a "hold" test because WRAT performance is moderately correlated with severity.

Additionally, the correlations indicate that the relationship between MMSE score and the WRAT is similar for both males and females. There is a difference in the WRAT

ranking for males and females. However, the difference in rankings is a result of males having higher correlations for their first two subtests. This indicates that males have a stronger relationship with their first two subtests than females have with their two highest tests. This results in differences within the rankings, although no gender differences are present in the correlation between the WRAT and MMSE.

Historically, there have been many methods exploring ways to document change in neuropsychological functioning. Determining whether the WRAT is a "hold" test is important because past research has indicated that "hold" tests are good measures of premorbid functioning. Brain dysfunction is characterized by change, specifically how present abilities relate to previous ones. This issue is especially relevant to diagnosing dementia because present abilities must fall below what would be expected with normal aging, with consideration of previous ability levels. Therefore, problems arise when previous abilities are unknown and must be estimated. Research about the ability of the WRAT to be a "hold" test has been conflicted, especially in brain damaged populations (Crawford et al., 1988; Griffin et al., 2002; Hart et al., 1986; Johnstone & Wilhem, 1996; Nebes et al., 1984; Nelson & O'Connell, 1978; O'Carroll et al., 1995; Patterson et al., 1994; Stebbins et al., 1990a; Stebbins et al., 1990b; Storandt et al., 1995). It is still unclear whether the WRAT and other similar measures are negatively impacted by increasing severity. If the tests are compromised, the WRAT is not a "hold" test and will underestimate previous abilities. Underestimation can lead to inaccuracies with regard to the progression of dementia.

Previous research has found the WRAT to be a more accurate estimate of premorbid ability in dementia and TBI patients than other measures, such as vocabulary

(Nelson & McKenna, 1975; Orme et al., 2004). However, a trend for greater score change with increasing severity, indicating that scores are influenced by severity, prompted the authors to suggest confirming estimates through the use of multiple sources (Orme et al., 2004). This trend is consistent with the present findings, which indicate that the WRAT is influenced by severity and is not a "hold" test.

Because much of the research involving reading and severity is inconsistent, investigators are seeking alternative ways of making premorbid estimates more accurate. Estimates have been developed using WAIS subtests and demographic variables and have generally been found to be more accurate than either method alone (Krull, Scott, & Sherer, 1995; Schoenberg, Duff, Scott, & Adams, 2003; Schoenberg, Duff, Dorfman, & Adams, 2004; Scott, Krull, Williamson, Adams, & Iverson, 1997). Similarly, studies have also examined methods using both the NART and demographic variables, finding that the combination is better than either method alone (Crawford et al., 1989; Crawford et al., 1990a; Crawford et al., 1990b; Gladsjo et al., 1999). The present results indicate that the WRAT should not be considered a "hold" measure, and other methods using a combination of variables should result in more accurate estimations of premorbid abilities, particularly when individuals have increasing brain dysfunction.

Hypothesis Two

The second hypothesis stated that a smaller subset of "don't hold" tests would be significant predictors of severity status. Specifically, the Category Test would be a significant predictor along with at least one of the TPT subtests. Additionally, motor tests would not be a significant predictor for either gender. Results show that although both the Category Test and one of the TPT subtests were significant for males and females,

overall regression models showed significant gender differences. For males, the significant subtests were the Category Test, TPT-Memory, and Trails A, suggesting that these subtests were the best at predicting severity status. In contrast, the significant subtests for females were the Rhythm Test, the Category Test, and WRAT-Reading. SSPT was in the model, but was dropped in step 4 due to shared variance with other subtests. Motor tests were not significant predictors for either gender.

The results indicate that the Category Test, which measures reasoning and abstraction, reliably discriminates between severity levels for both males and females. Similarly, the Rhythm Test is the strongest "don't hold" subtest for females and has a strong correlation with MMSE. Therefore, it is expected that it will be a strong predictor in the regression equation.

In contrast to other research (Crawford et al., 1988; Hart et al., 1986; Nebes et al., 1984), the present study found that the WRAT, i.e. reading ability, is not a "hold" test and is moderately correlated with severity for both males and females. The fact that the WRAT is a significant predictor of severity in females emphasizes its status as a strong "don't hold" test.

Although MLR uses the correlation between subtest and MMSE to calculate prediction rates, it also considers the contribution that each individual subtest makes to the regression model. Given this, some subtests share so much variance with other subtests that they do not add to the prediction even though they are highly correlated with MMSE, or severity. Overall, the regression shows that the few subtests that are significant predictors are all that is needed to differentiate AD patients based on severity status. This suggests that valuable information concerning severity can be gleaned by

administering only a few subtests of the HRB. However, in most situations other relevant clinical information is needed, not just the state of severity.

One disadvantage of the HRB is the lengthy administration time, which often requires a day of testing. This amount of time is not always feasible for patients, particularly the elderly or disabled. Given this constraint, researchers have investigated ways to shorten administration time, while retaining the comprehensive clinical information of a full battery. Abbreviated tests have been shown to be effective at identifying brain impairment including information regarding lateralization, location, and severity (Erickson, Calsyn, & Scheupbach, 1978; Golden, 1976; Storrie & Doerr, 1980).

One of the first researchers to examine the utility of an abbreviated battery was Golden (1976). He found that along with other tests, the HRB subtests of Trails, SSPT, Aphasia Screening Test, and the Rhythm Test were sufficient to differentiate normal from brain-damaged participants 93% of the time. This suggests that these subtests are particularly good at identifying brain dysfunction, likely because of the complexity of the tasks. In the present study, the Rhythm Test was shown to be a significant predictor of severity in females. However, the other subtests used by Golden did not make a significant contribution to the regression model in this study. Additionally, Golden (1976) did not report information regarding participant gender. The results of the present study indicate that gender significantly impacts the subtests that predict severity, highlighting the importance of examining gender differences.

Focusing specifically on AD patients, Storrie and Doerr (1980) found that an abbreviated battery consisting of the WAIS, TPT, Trails, the Category Test, and Finger Tapping, discriminated well between normal, aged males and AD subjects. Individuals

with AD demonstrated significant problems with the Category Test and Trails B. None of the TPT subtests differentiated between AD and controls. However, this appears due to early termination stemming from frustration resulting from early errors. Also, the study had a very small number of AD subjects. Therefore, it is unknown how the TPT would have discriminated between subjects. In the present study, one TPT subtest was a significant predictor of severity status, TPT-Memory for males.

Searight, Dunn, Grisso, Margolis, and Gibbons (1989) investigated the relationship between neuropsychological functioning and daily living skills in geriatric patients with suspected dementia. They found that four HRB subtests were most strongly associated with daily functioning: SSPT, the Rhythm Test, TPT-Memory, and Finger Tapping- Dominant. These findings suggest that these four subtests assess skills that are closely related to AD patients' ability to perform tasks of daily living. In contrast to the present study, Searight et al. (1989) examined HRB subtests' ability to predict daily living skills, which have both a cognitive and physical component. Given this, it was expected that motor tests would predict daily functioning. The present findings focused on discriminating between two levels of severity, which relies less on motor ability.

Several studies (Golden, 1976; Searight et al., 1989) indicate the importance of the Rhythm Test in an abbreviated battery. This is supported by the present findings, which demonstrate that it is the most important predictor of severity status in female AD patients. The Rhythm Test has been independently examined for its efficacy in diagnosing brain function. Reitan and Wolfson (1989) compared control and brain damaged subjects on the Rhythm Test. Those with brain dysfunction performed significantly worse than controls. Lesion placement, age, and gender were not

contributing factors to subtest scores. These findings suggest that the Rhythm Test is able to discriminate between normal subjects and those with brain impairment, regardless of damage location or gender. Unlike Reitan and Wolfson (1989), the present study found gender differences with regard to the ability of the Rhythm Test to distinguish between two severity levels.

Hypothesis Three

The third hypothesis states that a smaller subset of "don't hold" tests will be significant predictors of severity group membership and classification accuracy will vary by gender. Results show the accuracy of classification for the undivided sample is equivalent in the normal/mild (77.8% hits, correctly classified) and moderate/ severe groups (76.1% hits). When the sample was divided by gender, normal/mild males had more accurate classification than normal/mild females (88% vs. 78.9%). In the moderate/severe group, males and females were approximately equally well classified (81% vs. 76.1%). Overall, males had slight reduction in classification accuracy from normal/mild group to moderate/severe group (88% vs. 81%). However, for females, accuracy was equivalent from normal/mild to moderate/severe (78.9% vs. 76.1%).

The findings suggest that HRB subtests can accurately classify AD participants into severity groups, especially when the sample is divided by gender. Overall, classification accuracy improved as severity decreased, with the normal/mild group having more hits than the moderate/severe group. This indicates that there are significant test performance differences between the two groups. With regard to gender differences, males and females show improvement in classification accuracy as severity decreases, but males demonstrate more classification accuracy. This indicates that discrepancies in

test performance between severities are more pronounced in males. This suggests greater variance in test performance for males in the moderate/severe group as compared to the more consistent poor scores in the normal/mild group.

There is a dearth of research available using the HRB to classify participants into severity groups. Most of the studies using discriminant analysis examine the ability of subtests to differentiate between individuals with and without brain damage. The HRB has two indices that describe severity of impairment, the Impairment Index and the General Neuropsychological Deficit Scale (GNDS). The Impairment Index uses the results of 7 subtests as an indication of the proportion of individual tests that are in the impairment range (Reitan & Wolfson, 1993). In contrast, the GNDS uses 42 variables, including level of performance, pathognomonic signs, patterns and relationships among test results, and right-left differences (Reitan & Wolfson, 1993). This yields a measure of impairment based on more extensive data. The present study examined the ability of HRB and WRAT-Reading subtests to classify AD participants into two severity groups, normal-mild and moderate-severe.

Early on, researchers described the advantages of using neuropsychological tests instead of the WAIS to discriminate those with brain-damage from normal populations (Goldstein & Shelly, 1984). The HRB has been shown to discriminate between those with normal functioning and those with disorders affecting the brain, including alcoholism, epilepsy, and psychiatric conditions (Goldstein & Shelly, 1972; Goldstein & Shelly, 1984; Kupke & Lewis, 1986; O'Leary, Donovan, Chaney, Walker, & Schau, 1979; Shearn, Berry, & Fitzgibbons, 1976). Shearn et al. (1976) demonstrated that the HRB subtests of TPT-Total time, Trails A, Trails B, and Finger Tapping-Right hand

could distinguish between psychiatric patients with and without suspected mild brain damage better than a psychiatric evaluation. O'Leary et al. (1979) ranked HRB and WAIS subtests using F-ratio and subjected those that most highly differentiated between alcoholics and non-alcoholic to a discriminant analysis. Of the HRB subtests, the Category Test, TPT-Both, and Trails B were found to distinguish between participants. These HRB subtests and six WAIS subtests resulted in correct classifications 74.7% of the time. This finding is similar to what was obtained in the present study; however, unlike the present sample, their sample was not divided by gender.

Looking at individuals with epilepsy, Kupke and Lewis (1986) found that the HRB subtests of Trails B, the Rhythm Test, and Finger Tapping discriminate between mildly and moderately impaired participants. However, for the control group and those who were mildly impaired, scores were not significantly different, suggesting that those with mild dysfunction related to epilepsy retain many of the task-related abilities. Additionally, Grip Strength discriminated between normal and severely impaired individuals, with normal and mildly impaired participants showing no significant differences. This is in contrast to those with AD, as their motor functioning is not significantly affected even at more severe impairment levels. Overall, subtests have been shown to differentiate between individuals with mild and moderate brain impairment. These findings are supported by the present study, which indicated that HRB subtests can successfully discriminate between normal/mild and moderate/severe AD groups. The present study demonstrated that the HRB subtests that function as "don't hold" tests vary by the physiological characteristics of the disease (O'Leary et al., 1979).

Goldstein and Shelly (1972) found that not only did the HRB discriminate between brain-damaged and non brain-damaged participants; it also differentiated brain-damaged subjects with lateralized and diffuse damage. However, it could not discriminate between right hemisphere and diffuse lesion. The findings indicate the degree of specificity that the HRB can identify impairment. Lateralization was not examined in the present study as AD patients have more diffuse impairments.

Hypothesis Four

Finally, the fourth hypothesis stated that there would be significant gender differences in test performance on some of the HRB subtests. Specifically, females will have lower scores on Grip Strength and Finger Tapping, with both their dominant and non-dominant hands. Results show that there were significant gender differences only on motor subtests. Females had significantly poorer performance on both Grip Strength and Tapping, with both hands. It is unclear whether this gender difference is due to physical strength, hand size, or a combination of the two. This finding is consistent with previous literature (Chavez et al., 1983; Dodrill, 1979; Morrison et al., 1979; Seidenberg et al., 1984; Yeudall et al., 1987) which found male-superior performance on HRB motor tests. Male-superior performance has been documented across test modalities and ages (Batchelor & Dean, 1990; Christianson & Leathem, 2004). The present finding suggests that gender differences observed in younger, normal individuals also characterize those with AD.

Gender differences on motor tests have been observed in both normal and clinical populations. Dodrill (1979) examined gender differences within a non-neurological sample and in those with neurological conditions, consisting primarily of individuals with

seizure disorders. Although gender differences were present in both groups, he found they were more prominent in non-neurological participants. This suggests that when a more significant variable, such as brain impairment, is introduced, variance related to gender becomes less important. This may be due to the overwhelming influence of brain impairment on test performance, which overshadows lesser gender differences. However, there is no measure of severity in this study. It is possible that more significant brain impairment would result in no observed gender differences on the motor tests. The result demonstrates that gender differences on motor tests are evident in both normal individuals and those with brain dysfunction, although sometimes to a lesser extent.

This conclusion is supported by Seidenberg et al. (1984) who also found gender differences on motor tests in those with seizure disorders. Several other clinical populations have demonstrated poorer performance on motor skills tasks. Individuals with TBI (Geldmacher & Hills, 1997) and schizophrenia (Flashman, Flaum, Gupta, & Andreasen, 1996) have been found to have lower scores on tasks of motor speed and coordination. The average MMSE score of AD participants in the present sample was 19, which indicates moderate impairment. However, the sample's dysfunction may not be significant enough to counteract more prominent gender differences.

Studies have shown that numerous factors influence performance on the HRB motor tests (Chavez et al., 1983; Elias et al., 1993; King, Hannay, Masek, & Burns, 1978; Ruff & Parker, 1993; Wefel, Hoyt, & Massman, 1999). In a study examining gender differences and age, Ruff and Parker (1993) found that age negatively impacted females' performance on Finger Tapping; however, males' scores were unaffected by age.

However, another study (Elias et al., 1993) found that females performed worse than males across six age groups, but no Age x Gender interaction was noted.

Other studies examining neurological samples have shown that motor ability is not related to age. Prigatano and Parsons (1976) examined neuropsychological test performance in those with documented brain damage and found that their scores on Finger Tapping were not correlated with age. Additionally, tapping ability appears to be relatively unaffected by depressive symptoms and test anxiety, but poorer performance has been linked to trait anxiety state (Chavez et al., 1983; King et al., 1978; Wefel et al., 1999). It cannot be determined from the present study whether these elderly AD participants' scores were adversely influenced by age or severity of brain dysfunction.

An interesting finding of this study was a non-significant trend for female-superior performance on the Category Test and male-superior performance on the Rhythm Test and TPT-Minutes/Block. This result suggests that female AD participants are less impaired on reasoning and complex problem-solving tasks. Additionally, better performance on the Rhythm Test by males indicates fewer deficits in attention/ concentration and alertness to non-visual stimuli (Jarvis & Barth, 1984; Reitan & Wolfson, 1993; Reitan & Wolfson, 2004). Male-superior performance on the TPT-Minutes/Block emphasizes higher levels of psychomotor speed. This may be due to the nature and magnitude of premorbid differences between males and females.

Female-superior performance on the Category Test and male-superior performance on the Rhythm Test and TPT-Minutes/Block have not been documented in previous literature. Other studies (Chavez et al., 1982; Gordon & O'Dell, 1983) have found significant gender differences on Location and Memory scores, with female-

superior performance. These findings were not supported by this study, as both of these HRB subtests were equivalent for males and females. In a study examining the Category Test, Elias et al. (1993) found that, in addition to age-related linear decline on all HRB subtests, females performed worse on this test with age. The present study does not support this finding based on the non-significant trend for female-superior performance on the Category Test.

General Implications

The present study has significant strengths, which enable it to contribute to the general knowledge regarding AD and how it impacts individual neuropsychological functioning. The extensive neurological evaluation of the study's participants is a significant advantage. These participants have had more thorough evaluations and diagnoses compared to general clinical practice. Additionally, the use of archival data has distinct advantages, both in the information generated and in statistical analyses. The data used in this study were archival data that were part of a larger set of administered tests. The nature of this type of testing results in a standardized testing environment that is similar across participants. This is important because it creates equivalent testing conditions for all participants. Additionally, other advantages of archival data include the reduction of threats to internal validity, such as reactivity and expectancy.

One method of examining change is to compare current deficits to previous ones. However, previous functioning is often unknown and must be estimated. Reading has been considered relatively resilient to brain impairment, and reading tests rely on these skills to function as "hold" tests. Review of other study results indicated that reading tests are not good "hold" tests for AD patients (Fromm et al., 1991; Stebbins et al., 1990a;

Storandt et al., 1995). Although the present study found no gender differences with regard to test performance, findings do support other studies and indicate that the WRAT is clearly a moderate "don't hold" test with this patient population. These findings imply that the WRAT cannot be a good estimator of premorbid functioning, as it is moderately related to severity. Use of the WRAT to estimate previous abilities in AD patients would be unwise and would likely underestimate premorbid levels. Other research suggests that although reading may be an effective measure of previous abilities, the NART's format of irregular words may be a better estimator (Crawford et al., 1988; Nebes et al., 1984; O'Carroll & Gilleard, 1986). However, even when using the NART, other information such as demographic variables should be included to increase the accuracy of estimates of premorbid functioning (Grober & Sliwinski, 1991; Willshire et al., 1991).

Studies have shown that abbreviated HRB batteries can successfully differentiate brain damaged patients from normal subjects (Erickson et al., 1978; Golden, 1976; Storrie & Doerr, 1980). The present study substantiates a related finding, determining that a small subset of subtests is sufficient to differentiate individuals in two severity groups. This suggests that these tests might be sufficient to yield information concerning severity. However, only using a few subtests would result in significant limitations regarding other clinical information and use of pattern analysis, which requires administering the full HRB battery.

The physiological changes related to the AD disease appear clearly reflected in neuropsychological performance. AD patients show significant deficits in abstraction, attention/concentration, and memory skills. The present study corroborates previous research findings, which suggest that these skills are negatively impacted as severity

increases (Fromm et al., 1991; O'Carroll et al., 1995; Paolo et al., 1997). However, motor skills appear to be largely unaffected by AD, even at more advanced disease stages. This neuropsychological deficit profile is unique to AD and varies from other diseases disrupting brain function.

The gender differences on the subtests are important and can inform clinical practice. The observed differences suggest that AD likely affects males and females very differently. Although both genders have problems with abstraction and reasoning at higher levels of severity, males had difficulty with Blocks-Both. This has a significant memory component, suggesting that these skills are compromised in males at higher levels of severity. In contrast, the subtests that best discriminate between severity levels in females assessed reading ability, psychomotor speed, and attention/ concentration. These results emphasize the significant problems that females have in these areas when severity is increased. Overall, the findings suggest a distinct neuropsychological profile associated with AD that may have additional gender differences. More research is needed to determine whether the observed gender differences are due to the AD disease process itself or the nature of premorbid abilities.

Assessing change in neuropsychological functioning is essential to determine the progression of disease. It has implications for understanding the physiological course of the disease and how it affects individual patients. Perhaps the most significant implication of the present study is its contribution to the understanding of the overall AD disease process and its effect on individual neuropsychological functioning.

Limitations

Although sound methodologies were used in this study, there are several limitations that may have impacted the results and should be considered. Some of the limitations concern the characteristics of the sample and others are related to the assessments utilized. The limitations and potential implications of this study are thoroughly discussed to prevent improper generalization of the results.

As with all samples, its characteristics significantly limit generalization to other populations. The results reflect only the characteristics of the present sample and may only be applicable to other, similar AD patients. First, the sample is drawn from a private practice in a large, southern state within the United States. All participants can be considered aged (mean = 73 years) and results or generalizations must be guarded with respect to other age groups or individuals. The sample consists primarily of females (70%) and Caucasians (97%); therefore, the study significantly represents only this population. Overall, females are more likely to be diagnosed with AD, and this is reflected in the present study's sample, which is disproportionately female. The ethnic composition of the sample is restricted, which affects the ability to generalize to other ethnic groups. The restriction in gender and ethnicity may be influenced by patients' willingness and ability to access neuropsychological services.

All participants were referred to either a neuropsychologist in private practice or an Alzheimer's Clinic for an evaluation. This impacts the type of patient represented in the sample, as these individuals are probably more severe. Those with very mild impairment are likely not symptomatic enough to warrant an evaluation referral to a

neuropsychologist. Also, neuropsychological evaluations are expensive; therefore, patients in the sample likely are from disproportionally higher socioeconomic groups.

Some of the assessments used within this study also present limitations. Although the HRB and the WRAT have long histories of use and have been well researched, these measures suffer from limitations. The study was restricted by test selection. Of the HRB subtests, only 7 subtests were included in the present study. Other subtests may have been applicable to examining "hold/don't hold" abilities in AD patients. Additionally, the patient population used may present some limitations evident during testing. AD patients may be disoriented and easily fatigued, especially in more severe illness stages. Therefore, the results may not represent their optimal effort. To account for this tendency during testing, any participant with missing data was eliminated from the study. Also, some subtests may be too difficult for those at advanced levels of severity and may result in a floor effect.

This study is limited by how participants are classified into severity groups. Any limitation of the MMSE, will also be a limitation of the study. The present study's participants are individuals with significant dementia symptoms and extensive neurological evaluations. Therefore, it can be assumed that the extent of their impairments will negatively impact their cognitive functioning. However, 8 participants scored in the normal range and 55 scored in the mildly impaired range on the MMSE. It is surprising that any of the participants had scores in the normal range on this test, considering their significant deficits. Although it is the standard for the diagnosis of severity in neuropsychological functioning, it may be that other severity scales may be more accurate in classification. Therefore, it is important to examine in the usefulness of

the MMSE as it commonly used in clinical practice and is an essential component of the neurological exam.

Another limitation of the study is the statistics used. Some of the hypotheses rely on dividing the sample by gender, since the proportion of males to females is unequal, this may have negatively impacted the validity of findings concerning males. Also, several of the hypotheses have been tested using a correlational design, which limits the conclusions that can be drawn. The nature of correlational design prevents determination of causality. Additionally, the present study used a cross-sectional design. This entails comparing test results from different people at varying levels of severity, rather than test results from the same individual over time, as found in a longitudinal design. There are inherent limitations associated with a cross-sectional study. The most significant is that individuals are compared to other individuals, instead of their own previous performance.

Suggestions for Future Research

Alzheimer's Dementia is a significant problem in the United States and negatively impacts the quality of life for many aged individuals. There is no current effective treatment for AD and much remains to be learned about this disease. The current study found that the HRB is a combination of "hold" and "don't hold" subtests. These findings support previous research suggesting that the Category Test and subtests measuring attention/concentration are strong "don't hold" tests and are sensitive to increases in severity in those with AD (Hom, 1992; Storrie & Doerr, 1980). Also, motor tests appear to be relatively unaffected by AD disease process, even at advanced severity levels (Chavez et al., 1983; Dodrill, 1979; Hom, 1992; Morrison et al., 1979; Seidenberg et al., 1984; Yeudall et al., 1987). The WRAT has conflicted research regarding its utility as a

good estimator of premorbid functioning, or a "hold" test (Crawford et al., 1988; Griffin et al., 2002; Hart et al., 1986; Johnstone & Wilhem, 1996; Nebes et al., 1984; Nelson & O'Connell, 1978; O'Carroll et al., 1995; Patterson et al., 1994; Stebbins et al., 1990a; Stebbins et al., 1990b; Storandt et al., 1995). The present study indicates that it is not a good "hold" test for this AD sample.

The sensitivity of a subtest likely varies depending on type of brain dysfunction. Physiological changes resulting from a disease process will dictate neuropsychological test performance. Given this, additional research is needed to determine how severity and other types of illness impact a tests' ability to be a "hold/don't hold" test. Future research efforts may be directed at examining the "hold/don't hold" status of other diseases.

An additional area for future research is the investigation of other neuropsychological measures and how they are impacted by severity. The results of the present study indicate that tests vary in their sensitivity to changes in severity. Therefore, future research should examine the effect of severity on other commonly used neuropsychological tests, including other HRB subtests. Other scales that are measures of severity should be examined to determine the efficacy of the MMSE. This will enable conclusions to be drawn regarding "hold/don't hold" status of other neuropsychological measures. Additional research should also be conducted using samples with greater diversity to allow generalization to other ethnic groups.

Future research can also address the design limitations of the present study by employing a longitudinal design. This will enable participants to serve as their own comparison, in that neuropsychological testing will occur over time. This allows conclusions regarding the way that severity impacts individual test performance. Also, a

design that permits other types of analyses, other than correlations will be able to make determinations concerning causality. This will allow for a direct measure of change instead of measuring it indirectly as done in the present study.

There is a paucity of research available on using the WRAT as an estimator of premorbid functioning. The WRAT is more commonly used in clinical practice; therefore, it is more likely that patients may have taken it during a pre-injury evaluation. Most of the research uses the NART because it was developed as an estimator of premorbid functioning. Since the NART is used primarily for premorbid estimation and is not commonly used for other evaluations, making it unlikely that patients will have taken it before. Since it is more probable that pre-injury WRAT scores may be available, it is important to understand how its performance is impacted by increasing impairment severity. It is unclear how its format of irregular words impacts its estimation efficacy. Underestimation of previous functioning can result in inappropriate treatment goals and misunderstanding regarding disease progression. Future research should focus on using the WRAT with AD patients and determining whether adding additional information, such as demographic variables, will increase the WRAT's estimation accuracy.

References

Adams, R. L., Parsons, O. A., Culbertson, J. L., & Nixon, S. J. (1996). *Neuropsychology for clinical practice.* Washington, DC: American Psychological Association.

Aftanas, M. S. & Royce, J. R. (1968). Analysis of brain damage tests administered to normal subjects with factor score comparisons across ages. *Multivariate Behavior Research, 4,* 459-481.

Aiken, L. R. (2003). *Psychological testing and assessment.* Boston: Allyn and Bacon.

Amieva, H., Lafont, S., Auriacombe, S., Rainville, C., Orgogozo, J., Adigues, J., et al. (1998). Analysis of error types in the trail making test evidences an inhibitory deficit in dementia of the Alzheimer type. *Journal of Clinical and Experimental Neuropsychology, 20,* 280-285.

Babcock, H. (1930). An experiment in the measurement of mental deterioration. *Archives of Psychology, 117,* 1-105.

Baillon, S., Muhommad, S., Marudkar, M., Suribhatla, S., Dennis, M., Spreadbury, C. et al. (2003). Neuropsychological performance in Alzheimer's disease and vascular dementia: Comparisons in a memory clinic population. *International Journal of Geriatric Psychiatry, 18,* 602-608.

Bak, J. S. & Greene, R. L. (1980). Changes in neuropsychological functioning in an aging population. *Journal of Consulting and Clinical Psychology, 48,* 395-399.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52,* 885-887.

Barth, J. T. & Macciocchi, S. N. (1986). Dementia: Implications for clinical practice and research. In S.B. Filskov & T.J. Boll (Eds.), *Handbook of clinical neuropsychology* (pp. 398-425). New York: Wiley.

Batchelor, E. S. & Dean, R. S. (1990). Gender differences in neuropsychological performance for children with reading deficits. *International Journal of Neuroscience, 50,* 95-102.

Bigler, E. D., Steinman, D. R., & Newton, J. S. (1981). Clinical assessment of cognitive deficit in neurologic disorder: I. Effects of age and degenerative disease. *Clinical Neuropsychology, 3,* 5-13.

Blair, J. R. & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3,* 129-136.

Boll, T. J. & Reitan, R. M. (1973). Effect of age on performance of the Trail Making Test. *Perceptual and Motor Skills, 36,* 691-694.

Bornstein, R. A. (1982). Reliability of the Speech Sounds Perception Test. *Perceptual and Motor Skills, 55,* 203-210.

Boyle, G. J. (1986). Clinical neuropsychological assessment: Abbreviating the Halstead category test of brain dysfunction. *Journal of Clinical Psychology, 42,* 615-625.

Bright, P., Jaldow, E., & Kopelman, M. D. (2002). The National Adult Reading Test as a measure of premorbid intelligence: A comparison with estimates derived from demographic variables. *Journal of the International Neuropsychological Society, 8,* 847-854.

Butters, M. A., Goldstein, G., Allen, D. N., & Shemansky, W. J. (1998). Neuropsychological similarities and differences among Huntington's disease, Multiple Sclerosis, and Cortical Dementia. *Archives of Clinical Neuropsychology, 13,* 721-735.

Canter, A. (1966). A background interference procedure to increase sensitivity of the Bender-Gestalt Test to organic brain disorder. *Journal of Consulting Psychology, 30,* 91-97.

Cervilla, J., Prince, M., Joels, S., Lovestone, S., & Mann, A. (2004). Premorbid cognitive testing predicts the onset of dementia and Alzheimer's disease better than and independently of APOE genotype. *Journal of Neurology, Neurosurgery, and Psychiatry, 75,* 1100-1106.

Chavez, E. L., Schwartz, M. M., & Brandon, A. (1982). Effects of gender of subject and method of block presentation on the Tactual Performance Test. *Journal of Consulting and Clinical Psychology, 50,* 600-601.

Chavez, E. L., Trautt, G. M., Brandon, A., & Steyaert, J. (1983). Effects of test anxiety and gender of subject on neuropsychological test performance: Finger Tapping, Trail Making, Digit Span, and Digit Symbol Test. *Perceptual and Motor Skills, 56,* 923-929.

Choca, J. P., Laatsch, L., Wetzel, L., & Agresti, A. (1997). The Halstead Category Test: A fifty year perspective. *Neuropsychology Review, 7,* 61-75.

Cockburn, J., Keene, J., Hope, T., & Smith, P. (2000). Progressive decline in NART score with increasing dementia severity. *Journal of Clinical and Experimental Neuropsychology, 22,* 508-517.

Christianson, M. K. & Leathem, J. M. (2004). Development and standardisation of the Computerised Finger Tapping Test: Comparison with other finger tapping instruments. *New Zealand Journal of Psychology, 33,* 44-49.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21-49). Hove: Lawrence Erlbaum.

Crawford, J. R., Parker, D. M., & Besson, J. A. O. (1988). Estimation of premorbid intelligence in organic conditions. *British Journal of Psychiatry, 153,* 178-181.

Crawford, J. R., Stewart, L. E., Parker, D. M., Besson, J. A. O., & Cochrane, R. H. B. (1989). Estimation of premorbid intelligence: Combining psychometric and demographic approaches improves predictive accuracy. *Personality and Individual Differences, 10,* 793-796.

Crawford, J. R., Cochrane, R. H. B., Besson, J. A. O., Parker, D. M., & Stewart, L. E. (1990a). Premorbid IQ estimates obtained by combining the NART and demographic variables: Construct validity. *Personality and Individual Differences, 11,* 209-210.

Crawford, J. R., Nelson, H. E., Blackmore, L., Cochrane, R. H. B., & Allan, K. M. (1990b). Estimating premorbid intelligence by combining the NART and demographic variables: An examination of the NART standardisation sample and supplementary equations. *Personality and Individual Differences, 11,* 1153-1157.

Cummings, J. L., Houlihan, J. P., & Hill, M. A. (1986). The pattern of reading deterioration in dementia of the Alzheimer type: Observations and implications. *Brain and Language, 29,* 315-323.

Dean, R. S. (1985). Review of the Halstead-Reitan Neuropsychological Test Battery. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 642-646). Highland Park, NJ: The Gryphon Press.

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

Dodrill, C. B. (1979). Gender differences on the Halstead-Reitan Neuropsychological Battery and on other neuropsychological measures. *Journal of Clinical Psychology, 35,* 236-241.

Dodrill, C. B. & Dikmen, S. S. (1978). The Seashore Tonal Memory Test as a neuropsychological measure. *Journal of Consulting and Clinical Psychology, 46,* 192-193.

Elias, M. F., Robbins, M. A., Walter, L. J., & Schultz, N. R. (1993). The influence of gender and age on Halstead-Reitan Neuropsychological Test performance. *Journal of Gerontology, 48,* 278-281.

Elias, M. F., Robbins, M. A., & Elias, P. K. (1996). A 15-year longitudinal study of Halstead-Reitan neuropsychological test performance. *Journal of Gerontology: Psychological Sciences Series B, 51B,* 331-334.

Erickson, R. C., Calsyn, D. A., & Scheupbach, C. S. (1978). Abbreviating the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 34,* 922-926.

Fischer, W. E. & Dean, R. S. (1990). Factor structure of the Halstead Category Test by age and gender. *International Journal of Neuropsychology, 12,* 180-183.

Flashman, L. A., Flaum, M., Gupta, S., & Andreasen, N. C. (1996). Soft signs and neuropsychological performance in schizophrenia. *American Journal of Psychiatry, 153,* 526-532.

Folstein, M. F., Folstein, S. E., McHugh, P. R., & Fanjiang, G. (2001). *Mini-Mental State Examination.* Odessa, Fl.: Psychological Assessment Resources, Inc.

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12,* 711-738.

Fromm, D., Holland, A. L., Nebes, R. D., & Oakley, M. A. (1991). Longitudinal study of word-reading ability in Alzheimer's disease: Evidence from the National Adult Reading Test. *Cortex, 27,* 367-376.

Gallese, A. J. (1956). Spiral afteraffect as a test of organic brain damage. *Journal of Clinical Psychology, 12,* 254-258.

Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the Trail Making Test: What makes Part B harder? *Journal of Clinical and Experimental Neuropsychology, 17,* 529-535.

Geldmacher, D. S. & Hills, E. C. (1997). Effect of stimulus number, target-to-distractor ratio, and motor speed on visual spatial search quality following traumatic brain injury. *Brain Injury, 11,* 59-66.

Gladsjo, J. A., Heaton, R. K., Palmer, B. W., Taylor, M. J., & Jeste, D. V. (1999). Use of oral reading to estimate premorbid intellectual and neuropsychological functioning. *Journal of the International Neuropsychological Society, 5,* 247-254.

Golden, C. J. (1976). The identification of brain damage by an abbreviated form of the Halstead-Reitan Neuropsychological Battery. *Journal of Clinical Psychology, 32,* 821-826.

Goldstein, G. & Shelly, C. H. (1972). Statistical and normative studies of the Halstead neuropsychological test battery relevant to a neuropsychiatric hospital setting. *Perceptual Motor Skills, 34,* 603-620.

Goldstein, G. & Shelly, C. H. (1984). Discriminative validity of various intelligence and neuropsychological tests. *Journal of Consulting and Clinical Psychology, 52,* 383-389.

Gordon, N. G. & O'Dell, J. W. (1983). Gender differences in neuropsychological performance. *Perceptual and Motor Skills, 56,* 126.

Graham, F. K. & Kendall, B. S. (1946). Performance of brain-injured cases on a Memory-for-Designs Test. *Journal of Abnormal and Social Psychology, 41,* 303-314.

Green, R. C. (1995). Alzheimer's disease and other dementing disorders in adults. *Clinical Neurology, 3,* 1-84.

Griffin, S. L., Mindt, M. R., Rankin, E. J., Ritchie, A. J., & Scott, J. G. (2002). Estimating premorbid intelligence: Comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology, 17,* 497-507.

Grober, E. & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology, 13,* 933-949.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Halstead, W. (1940). Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. *American Journal of Psychiatry, 96,* 1263-1294.

Halstead, W. C. (1947). *Brain and intelligence.* Chicago: University of Chicago Press.

Hayes, W. L. (1994). *Statistics.* Fort Worth: Harcourt College Publishers.

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and gender. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment in neuropsychiatric disorders: Clinical methods and empirical findings* (pp.100-120). New York: Oxford University Press.

Heindel, W. C., Salmon, D. P., Shults, C. W., Walicke, P. A., & Butters, N. (1989). Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *The Journal of Neuroscience, 9,* 582-587.

Hewson, L. R. (1949). The Wechsler-Bellevue Scale and The Substitution Test as aids in neuropsychiatric diagnosis. *Journal of Nervous and Mental Disease, 109,* 158-183.

Hom, J. (1992). General and specific cognitive dysfunctions in patients with Alzheimer's disease. *Archives of Clinical Neuropsychology, 7,* 121-133.

Hom, J., Turner, M. B., Risser, R., Bonte, F. J., & Tintner, R. (1994). Cognitive deficits in asymptomatic first-degree relatives of Alzheimer's disease patients. *Journal of Clinical and Experimental Neuropsychology, 16,* 568-576.

Huber, S. J., Shuttleworth, E. C., & Freidenberg, D. L. (1989). Neuropsychological differences between the dementias of Alzheimer's and Parkinson's diseases. *Archives of Neurology, 46,* 1287-1291.

Hunt, H. F. (1943). A practical, clinical test for organic brain damage. *Journal of Applied Psychology, 27,* 375-386.

Jarvis, P. E. & Barth, J. T. (1984). *Halstead-Reitan test battery: An interpretative guide.* Odessa: Fl.: Psychological Assessment Resources, Inc.

Jastak, J. & Wilkinson, G. (1984). *The Wide Range Achievement Test: Manual of instructions.* Wilmington, DE: Jastak Associates.

Johnstone, B., Callahan, C. D., Kapila, C. J., & Bouman, D. E. (1996). The comparability of the WRAT-R Reading test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Archives of Clinical Neuropsychology, 11,* 513-519.

Johnstone, B., Hexum, C. L., & Ashkanazi, G. (1995). Extent of cognitive decline in traumatic brain injury based on estimates of premorbid intelligence. *Brain Injury, 9,* 377-384.

Johnstone, B., Hogg, J. R., Schopp, L. H., Kapila, C., & Edwards, S. (2002). Neuropsychological deficit profiles in senile dementia of the Alzheimer's type. *Archives of Clinical Neuropsychology, 17,* 273-281.

Johnstone, B. & Wilhelm, K. L. (1996). The longitudinal stability of the WRAT-R Reading subtest: Is it an appropriate estimate of premorbid intelligence? *Journal of the International Neuropsychological Society, 2,* 282-285.

Kareken, D. A., Gur, R. C., & Saykin, A. J. (1995). Reading on the Wide Range Achievement Test- Revised and parental education as predictors of IQ: Comparison with the Barona formula. *Archives of Clinical Neuropsychology, 10,* 147-157.

Kareken, D. A. & Williams, M. (1994). Human judgment and estimation of premorbid intellectual function. *Psychological Assessment, 6,* 83-91.

Kaufmann, A. (1968). The substitution test. *Cortex, 4,* 47-63.

Kemper, S., LaBarge, E., Ferraro, R., Cheung, H., Cheung, H., & Storandt, M. (1993). On the preservation of syntax in Alzheimer's disease. *Archives of Neurology, 50,* 81-86.

Kilpatrick, D. G. (1970). The Halstead Category Test of brain dysfunction: Feasibility of a short-form. *Perceptual and Motor Skills, 30,* 577-578.

King, G. D., Hannay, H. J., Masek, B. J., & Burns, J. W. (1978). Effects of anxiety and gender on neuropsychological tests. *Journal of Consulting and Clinical Psychology, 46,* 375-376.

Klesges, R. C. & Troster, A. I. (1987). A review of premorbid indices of intellectual and neuropsychological functioning: What have we learned in the past five years? *The International Journal of Clinical Neuropsychology, 9,* 1-11.

Klesges, R. C., Wilkening, G. N., & Golden, C. J. (1981). Premorbid indices of intelligence: A review. *Clinical Neuropsychology, 3,* 32-39.

Krull K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist,* 9, 83-88.

Kupke, T. (1983). Effects of subject gender, examiner gender, and test apparatus on Halstead Category and Tactual Performance Tests. *Journal of Consulting and Clinical Psychology, 51,* 624-626.

Kupke, T. & Lewis, R. (1986). Differential sensitivity of the WAIS and a modified Halstead-Reitan battery to severity of brain dysfunction in epilepsy. *Archives of Clinical Neuropsychology, 3,* 197-207.

Laatsch, L. & Choca, J. (1991). Understanding the Halstead Category Test by using item analysis. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3, 701-701.

Lafleche, G. & Albert, M. S. (1995). Executive function deficits in mild Alzheimer's disease. *Neuropsychology, 9,* 313-320.

Lamberty, G. J., Putnam, S. H., Chatel, D. M., Bieliaukas, L. A., & Adams, K. M. (1994). Derived trail making test indices: A preliminary report. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 7,* 230-234.

Leckliter, I. N. & Matarazzo, J. D. (1989). The influence of age, education, IQ, gender, and alcohol abuse on Halstead-Reitan Neuropsychological Test Battery performance. *Journal of Clinical Psychology, 45,* 484-512.

Lezak, M. D. (1976). *Neuropsychological Assessment.* New York: Oxford University Press.

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

Li, G., Silverman, J. M., Smith, C. J., Zaccario, M. L., Schmeidler, J., Mohs, R. C., et al. (1995). Age at onset and familial risk in Alzheimer's disease. *American Journal of Psychiatry, 152,* 424-430.

Mack, J. L. & Carlson, N. J. (1978). Conceptual deficits and aging: The Category Test. *Perceptual and Motor Skills, 46,* 123-128.

Maddrey, A. M., Cullum, C. M., Weiner, M. F., & Filley, C. M. (1996). Premorbid intelligence estimation and level of dementia in Alzheimer's disease. *Journal of the International Neuropsychological Society, 2,* 551-555.

Margolis, R. B., Greenlief, C. L., & Taylor, J. M. (1985). Relationship between the WAIS-R and the WRAT in a geriatric sample with suspected dementia. *Psychological Reports, 56,* 287-292.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams & Wilkins.

Matarazzo, J. D., Matarazzo, R. G., Wiens, A. N., Gallo, A. E., & Klonoff, H. (1976). Retest reliability of the Halstead Impairment Index in a normal, a schizophrenic, and two samples of organic patients. *Journal of Clinical Psychology, 32,* 338-349.

McFie, J. (1975). *Assessment of organic intellectual impairment.* New York: Academic.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services task force on Alzheimer's disease. *Neurology, 34,* 939-944.

Meyerink, L. H. (1982). Intellectual functioning: The nature and pattern of change with aging. *Dissertation Abstracts International.*

Moehle, K. A. & Long, C. J. (1989). Models of aging and neuropsychological test performance decline with aging. *Journal of Gerontology, 44,* 176-177.

Monti, L. A., Gabrieli, J. D. E., Reminger, S. L., Rinaldi, J. A., Wilson, R. S., & Fleischman, D. A. (1996). Differential effects of aging and Alzheimer's disease on conceptual implicit and explicit memory. *Neuropsychology, 10,* 101-112.

Morrison, M. W., Gregory, R. J., & Paul, J. J. (1979). Reliability of the Finger Tapping Test and a note on gender differences. *Perceptual and Motor Skills, 48,* 139-142.

Nebes, R. D., Martin, D. C., & Horn, L. C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology, 93,* 321-330.

Nelson, H. E. & McKenna P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14,* 259-267.

Nelson, H. E. & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex, 14,* 234-244.

Nolan, B. H., Swihart, A. A., & Pirozzolo, F. J. (1986). The neuropsychology of normal aging and dementia: An introduction. In D. Wedding & A. M. Horton (Eds.), *The neuropsychology handbook: Behavioral and clinical perspectives* (pp.410-440). New York: Springer Publishing Company.

O'Carroll, R. E. & Gilleard, C. J. (1986). Estimation of premorbid intelligence in dementia. *British Journal of Clinical Psychology, 25,* 157-158.

O'Carroll, R. E., Prentice, N., Murray, C., Van Beck, M., Ebmeier, K. P., & Goodwin, G. M. (1995). Further evidence that reading ability is not preserved in Alzheimer's disease. *British Journal of Psychiatry, 167,* 659-662.

O'Leary, M. R., Donovan, D. M., Chaney, E. F., Walker, D. R., & Schau, E. J. (1979). Application of discriminant analysis to level of performance of alcoholics and nonalcoholics on Wechsler-Bellevue and Halstead-Reitan Subtests. *Journal of Clinical Psychology, 35,* 204-208.

Ober, B. A., Shenaut, G. K., & Reed, B. R. (1995). Assessment of associative relations in Alzheimer's disease: Evidence for preservation of semantic memory. *Aging and Cognition, 2,* 254-267.

Orme, D. R., Johnstone, B., Hanks, R., & Novack, T. (2004). The WRAT-3 reading subtest as a measure of premorbid intelligence among persons with brain injury. *Rehabilitation Psychology, 49,* 250-253.

Paolo, A. M., Troster, A. I., Ryan, J. J., & Koller, W. C. (1997). Comparison of NART and Barona Demographic Equation premorbid IQ estimates in Alzheimer's disease. *Journal of Clinical Psychology, 53,* 713-722.

Parasuraman, R. & Haxby, J. V. (1993). Attention and brain function in Alzheimer's disease: A review. *Neuropsychology, 7,* 242-272.

Parsons, O. A. & Prigatano, G. P. (1978). Methodological considerations in clinical neuropsychological research. *Journal of Consulting and Clinical Psychology, 46,* 608-619.

Patterson, K., Graham, N., & Hodges, J. R. (1994). Reading in dementia of the Alzheimer type: A preserved ability? *Neuropsychology, 8,* 395-407.

Perry, R. J. & Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease. *Brain, 122,* 383-404.

Piotrowski, Z. (1937). The Rorschach inkblot method in organic disturbances of the cerebral nervous system. *Journal of Nervous and Mental Disease, 86,* 525-537.

Prigatano, G. P. & Parsons, O. A. (1976). Relationship of age and education to Halstead Test performance in different patient populations. *Journal of Consulting and Clinical Psychology, 44,* 527- 533.

Rasmusson, D. X., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the trail making test. *The Clinical Neuropsychologist, 12,* 169-178.

Reed, H. B. C. & Reitan, R. M. (1963a). A comparison of the effects of the normal aging process with the effects of organic brain-damage on adaptive abilities. *Journal of Gerontology, 18,* 177-179.

Reed, H. B. C. & Reitan, R. M. (1963b). Changes in psychological test performance associated with the normal aging process. *Journal of Gerontology, 18,* 271-274.

Reitan, R. M. (1955a). An investigation of the validity of Halstead's measures of biological intelligence. *Archives of Neurology and Psychiatry, 73,* 28-35.

Reitan, R. M. (1955b). The relation of the trail making test to organic brain damage. *Journal of Consulting, 19,* 393-394.

Reitan, R. M. (1955c). The distribution according to age of a psychologic measure dependent upon organic brain functions. *Journal of Gerontology, 10,* 338-340.

Reitan, R. M. (1962). The comparative psychological significance of aging in groups with and without organic brain damage. In C. Tibbits & W. Donohue (Eds.), *Social and psychological aspects of aging* (pp. 880-887). New York: Columbia University Press.

Reitan, R. (1994). Ward Halstead's contributions to neuropsychology and the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 50,* 47-70.

Reitan, R. M. & Wolfson, D. (1986a). The Halstead-Reitan Neuropsychological Test Battery. In D. Wedding, A. M. Horton, & J. Webster (Eds.), *The neuropsychology handbook: Behavioral and clinical perspectives* (pp. 134-160). New York: Springer Publishing Company.

Reitan, R. & Wolfson, D. (1986b). The Halstead-Reitan Neuropsychological Test Battery and aging. *Clinical Gerontologist, 5,* 39-61.

Reitan, R. M. & Wolfson, D. (1989). The Seashore Rhythm Test and brain functions. *The Clinical Neuropsychologist, 3,* 70-78.

Reitan, R. & Wolfson, D. (1990). The significance of the Speech-Sounds Perception Test for cerebral functions. *Archives of Clinical Neuropsychology, 5,* 265-272.

Reitan, R. & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation.* Tucson, AZ: Neuropsychology Press.

Reitan, R. & Wolfson, D. (1996). Theoretical, methodological, and validational bases of the Halstead-Reitan Neuropsychological Test Battery. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp.3-42). London: Oxford University Press.

Reitan, R. M. & Wolfson, D. (2004). The Halstead-Reitan Neuropsychological Test Battery for adults: Theoretical, methodological, and validational bases. In G. Goldstein & S. Beers (Eds.), *Comprehensive handbook of psychological assessment: Volume 1 intellectual and neuropsychological assessment* (pp.105-131). Hoboken, NJ: John Wiley and Sons.

Reynolds, C. (1986). Wide Range Achievement Test (WRAT-R): 1984 edition. *Journal of Counseling and Development, 64,* 540-541.

Ruff, R. M. & Parker, S. B. (1993). Gender- and age specific changes in motor speed and eye-hand coordination in adults: Normative values for the Finger Tapping and Grooved Pegboard Tests. *Perceptual and Motor Skills, 76,* 1219-1230.

Russell, E. W. (1972). WAIS factor analysis with brain-damaged subjects using criterion measures. *Journal of Consulting and Clinical Psychology, 39,* 133-139.

Russell, E. W. (1992). Reliability of the Halstead Impairment Index: A simulation and reanalysis of Matarazzo et al. (1974). *Neuropsychology, 6,* 251-259.

Russell, E. W. (1995). The accuracy of automated and clinical detection of brain damage and lateralization in neuropsychology. *Neuropsychology Review, 5,* 1-68.

Russell, E. W. (1998). In defense of the Halstead Reitan Battery: A critique of Lezak's review. *Archives of Clinical Neuropsychology, 13,* 365-381.

Russell, E. W. & Levy, M. (1987). Revision of the Halstead category test. *Journal of Consulting and Clinical Psychology, 55,* 898-901.

Russell, E. W., Neuringer, C., & Goldstein, G. (1970). *Assessment of brain damage: A neuropsychological key approach.* New York: Wiley-Interscience.

Russell, E. W. & Polakoff, D. (1993). Neuropsychological test patterns in men for Alzheimer's and Multi-infarct dementia. *Archives of Clinical Neuropsychology, 8,* 327-343.

Ryan, J. J., Larsen, J., & Prifitera, A. (1978). Short form of the Speech Sounds Perception Test: Further considerations. *Clinical Neuropsychology, 4,* 97-98.

Schinka, J. A. & Vanderploeg, R. D. (2000). Estimating premorbid level of functioning. In R. D. Vanderploeg (Ed.), *Clinicians guide to neuropsychological assessment* (pp.39-67). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schludermann, E. H., Schludermann, S. M., Merryman, P. W., & Brown, B. W. (1983). Halstead's studies in the neuropsychology of aging. *Archives of Gerontology and Geriatrics, 2,* 49-172.

Schoenberg, M. R., Duff, K., Dorfman, K., & Adams, R. L. (2004). Differential estimation of verbal intelligence and performance intelligence scores from combined performance and demographic variables: The OPIE-3 verbal and performance algorithms. *The Clinical Neuropsychologist, 18,* 266-276.

Schoenberg, M. R., Duff, K., Scott, J. G., & Adams, R. L. (2003). An evaluation of the clinical utility of the OPIE-3 as an estimate of premorbid WAIS-III FSIQ. *The Clinical Neuropsychologist, 17,* 308-321.

Scott, J. G., Krull, K. R., Williamson, D. J. G., Adams, R. L., & Iverson, G. L. (1997). Oklahoma premorbid intelligence estimation (OPIE): Utilization in clinical samples. *The Clinical Neuropsychologist, 11,* 146-154.

Searight, H. R., Dunn, E. J., Grisso, T., Margolis, R. B., & Gibbons, J. L. (1989). The relation of the Halstead-Reitan Neuropsychological Battery to ratings of everyday functioning in a geriatric sample. *Neuropsychology, 3,* 135-145.

Seidenberg, M., Gamache, M. P., Beck, N. C., Smith, M., Giordani, B., Berent, S. et al., (1984). Subject variables and performance on the Halstead Neuropsychological Test Battery: A multivariate analysis. *Journal of Consulting and Clinical Psychology, 52,* 658-662.

Shearn, C. R., Berry, D. F., & Fitzgibbons, D. J. (1976). A trial use of some of Reitan's neuropsychological tests to assess mild organic complications in psychiatric patients. *Journal of Clinical Psychology, 32,* 102-106.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9,* 371-377.

Skeel, R. L., Johnstone, B., Yangco, D. T., Walker, S. E., & Komatireddy, G. R. (2000). Neuropsychological deficit profiles in systemic lupus erythematosus. *Applied Neuropsychology, 7,* 96-101.

Small, B. J., Herlitz, A., Fratiglioni, L., Almkvist, O., & Backman, L. (1997). Cognitive predictors of incident Alzheimer's disease: A prospective longitudinal study. *Neuropsychology, 11,* 413-420.

Smith-Seemiller, L., Franzen, M. D., Burgess, E. J., & Prieto, L. R. (1997). Neuropsychologists' practice patterns in assessing premorbid intelligence. *Archives of Clinical Neuropsychology, 12,* 739-744.

Snelbaker, A. J., Wilkinson, G. S., Robertson, G. J., & Glutting, J. J. (2001). Wide Range Achievement Test-3 (WRAT-3). In G. Goldstein & S. Beers (Eds.), *Understanding psychological assessment* (pp.259-274). New York: Kluwer Academic/ Plenum Publishers.

Snyder, P. J. & Nussbaum, P. D. (1998). *Clinical neuropsychology: A pocket handbook for assessment.* Washington DC: American Psychological Association.

Stebbins, G. T., Gilley, D. W., Wilson, R. S., Bernard, B. A., & Fox, J. H. (1990a). Effects of language disturbances on premorbid estimates of IQ in mild dementia. *The Clinical Neuropsychologist, 4,* 64-68.

Stebbins, G. T., Wilson, R. S., Gilley, D. W., Bernard, B. A., & Fox, J. H. (1990b). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist, 4,* 18-24.

Storandt, M., Stone, K., & LaBarge, E. (1995). Deficits in reading performance in very mild dementia of the Alzheimer's type. *Neuropsychology, 9,* 174-176.

Storrie, M. C. & Doerr, H. O. (1980). Characterization of Alzheimer type dementia utilizing an abbreviated Halstead-Reitan Battery. *Clinical Neuropsychology, 2,* 78-82.

Strub, R. L. & Black, F. W. (1988). *Neurobehavioral disorders: A clinical approach.* Philadelphia: F. A. Davis Company.

Surgeon General (n.d.). *Mental health: A report of the surgeon general.* Retrieved October 9, 2004, from http://www.surgeongeneral.gov/library/mentalhealth/chapter5/sec4.html

Vanderploeg, R. D. & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology, 10,* 225-239.

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B. N. (1996). Estimation of WAIS-R premorbid intelligence: Current ability and demographic data used in a best-performance fashion. *Psychological Assessment, 8,* 404-411.

Victor, M. & Ropper, A. H. (2001). *The principles of neurology.* New York: McGraw-Hill Companies.

Wechsler, D. (1958). The measurement and appraisal of adult intelligence (4th ed.). Baltimore, MD: Williams and Wilkins.

Wefel, J. S., Hoyt, B. D., & Massman, P. J. (1999). Neuropsychological functioning in depressed versus nondepressed participants with Alzheimer's disease. *The Clinical Neuropsychologist, 13,* 249-257.

Wiens, A. N., Bryan, J. E., & Crossen, J. R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test-Revised in normal subjects. *The Clinical Neuropsychologist, 7,* 70-84.

Wilkinson, G. (1993). *The Wide Range Achievement Test- 3: Manual of instructions.* Wilmington, DE: Wide Range, Inc.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: A cross validation. *Journal of Clinical and Experimental Neuropsychology, 13,* 204-216.

Wittenborn, J. R. (1951). An evaluation of the use of difference scores in prediction. *Journal of Clinical Psychology, 7,* 108-111.

Yates, A. (1956). The use of vocabulary in the measurement of intellectual deterioration: A review. *Journal of Mental Science, 102,* 409-440.

Yeudall, L. T., Reddon, J. R., Gill, D. M., & Stefanyk, W. O. (1987). Normative data for the Halstead-Reitan Neuropsychological Tests stratified by age and gender. *Journal of Clinical Psychology, 43,* 346-368.

Yuspeh, R. L., Vanderploeg, R. D., & Kershaw, D. A. J. (1998). Normative data on a measure of estimated premorbid abilities as part of a dementia evaluation. *Applied Neuropsychology,* 5, 149-153.

Appendix

Institutional Review Board Approval

MEMORANDUM

TO: Ms. Gina Gibson-Beverly and Dr. Tony Young

FROM: Barbara Talbot, University Research

SUBJECT: HUMAN USE COMMITTEE REVIEW

DATE: 10/10/05

In order to facilitate your project, an EXPEDITED REVIEW has been done for your proposed study entitled:

"The Assessment of Change in Neuropsychological Functioning"

HUC-202

The proposed study's revised procedures were found to provide reasonable and adequate safeguards against possible risks involving human subjects. The information to be collected may be personal in nature or implication. Therefore, diligent care needs to be taken to protect the privacy of the participants and to assure that the data are kept confidential. Informed consent is a critical part of the research process. The subjects must be informed that their participation is voluntary. It is important that consent materials be presented in a language understandable to every participant. If you have participants in your study whose first language is not English, be sure that informed consent materials are adequately explained or translated. Since your reviewed project appears to do no damage to the participants, the Human Use Committee grants approval of the involvement of human subjects as outlined.

Projects should be renewed annually. ***This approval was finalized on October 5, 2005 and this project will need to receive a continuation review by the IRB if the project, including data analysis, continues beyond October 5, 2006.*** Any discrepancies in procedure or changes that have been made including approved changes should be noted in the review application. Projects involving NIH funds require annual education training to be documented. For more information regarding this, contact the Office of University Research.

You are requested to maintain written records of your procedures, data collected, and subjects involved. These records will need to be available upon request during the conduct of the study and retained by the university for three years after the conclusion of the study. If changes occur in recruiting of subjects, informed consent process or in your research protocol, or if unanticipated problems should arise it is the Researchers responsibility to notify the Office of Research or IRB in writing. The project should be discontinued until modifications can be reviewed and approved.

If you have any questions, please contact Dr. Mary Livingston at 257-4315.