The Construct Validity of the Clock Test
in Normal and Demented Adults

by

David Kurzman
B.Sc., McGill University, 1989

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

We accept this thesis as conforming to the required standard

[redacted]

Dr. Roger A. Dixon, Supervisor
(Department of Psychology)

[redacted]

Dr. Louis Costa, Departmental Member
(Department of Psychology)

[redacted]

Dr. Elaine Gallagher, Outside Member
(School of Nursing)

[redacted]

Dr. Brian Harvey, External Examiner
(Department of Psychological Foundations)

University of Victoria

Supervisor: Dr. Roger A. Dixon

Abstract

The Clock Test is a multidimensional measure comprised of Clock Drawing, Clock Setting, and Clock Reading. Recently components of this test, which were designed to assess visuospatial construction, visual perception, abstract conceptualization of time, and general cognitive status, have received much empirical attention. The present study examined the following properties of the Clock Test: (a) reliability (i.e., internal consistency), (b) the effect of age, education, and diagnostic group, (c) group differences, and classification of the dimensions of Clock Drawing, and (d) construct validity, or the relationship of the Clock Test to other neuropsychological tests. In addition, a series of Confirmatory Factor Analyses were conducted in order to examine the factor structure of the Clock Test relative to other neuropsychological measures.

The sample consisted of 19 individuals with a diagnosis of non Alzheimer's dementia, 100 individuals with a diagnosis of Possible Alzheimer's disease, 146 individuals with a diagnosis of Probable Alzheimer's disease, 90 individuals who complained of memory problems but were not demented, and 61 normal community-dwelling adults over the age of 50. The participants were further classified into 267 demented and 151 relatively normal adults ranging in age from from 54 to 84 years ($\underline{M}$ = 71.1 years).

Results demonstrated that whereas the Clock Setting (α = .84) and Clock Reading (α = .83) showed high reliabilities for the demented participants, the reliabilities were lower for the not demented adults (α = .84, .57; Clock Setting and Clock Reading, respectively). Unlike the demented adults whose performance was unrelated to age, performance decreased as a function of age for the relatively normal adults. Education was unrelated to performance for both groups of subjects. In addition, differences were observed in the performance of the demented and relatively normal groups, when each of the seven error categories of Clock Drawing were examined separately. Further, only the first category of errors alone was sufficient to classify individuals into their groups, since the remaining categories did not significantly improve the classification of individuals into groups.

When the Clock Test components were compared to a battery of valid and reliable neuropsychological tests, they were correlated positively with the convergent validity measures (r = .27 - .77); in contrast, they did not show a strong relationship, to the discriminant validity measures (r = .04 - .27). Whereas the pattern of correlations was replicated for the demented participants, they revealed only limited support for the relatively normal group. A Confirmatory Factor Analysis was conducted on a four factor model. This model was tested separately for the demented and not demented groups and although none of the models satisfied a conventional fit, the best-fitting model was one in which the Multi-focus Assessment Scale (MAS) was dropped and various error parameters were free to be estimated.

Discussion focuses on the limitations of the Clock Test as a diagnostic tool for dementia. In addition, suggestions for future research, as well as potential modifications to the Clock Test are proposed.

Examiners:

[redacted]

Dr. Roger A. Dixon, Superviser (Department of Psychology)

[redacted]

Dr. Louis Costa, Departmental Member (Department of Psychology)

[redacted]

Dr. Elaine Gallagher, Outside Member (School of Nursing)

[redacted]

Dr. Brian Harvey, External Examiner (Department of Psychological Foundations)

Table of Contents

List of Tables

List of Figures

Acknowledgements

This research project could not have been completed without the help of a number of people. First, I would like to thank my supervisor Roger A. (Allen) Dixon for his help and support during my time as a student at UVic. I would also like to thank my committee members: Dr. Louis Costa, Dr. Elaine Gallagher, and Dr. Brian Harvey for their helpful comments and criticisms, and for their encouragement during my defense. In the presence of rigid deadlines, I am grateful for the speed with which they were able to read the manuscripts. I would also like to thank Holly Tuokko and everyone at the Alzheimer's Disease and Related Disorders Clinic for their help during my frequent trips to the clinic.

Many friends have lent their support to this project. Special thanks go to Dave Almeida and Jennifer Maggs (my LISREL buddies), Richard Chadwick (overall mainframe help), Tracey Chapman and Jennifer Frood (for helping me finish all of the tables), Odette Gould, Giselle Kolaric, and Leslie McDonald-Miszczak (for overall confidence), and lastly Ingrid C. Friesen for listening to my endless comments about the Clock Test, and for learning more about clocks then she ever wanted to.

I would also like to thank my family for their confidence in me, and for their overall support and encouragement. Most of all, I would like to thank my fiancee, and soon to be wife, Arlene Berman for giving me strength to endure our three years apart, and for giving me the confidence to get the job done.

Lastly I would like to thank the Sara Spencer Research Foundation for supporting this research project.

Construct Validity of the Clock Test in Normal and Demented Adults

Chapter 1
Introduction

The present study examines the construct validity of the Clock Test as a diagnostic tool for dementia. Shulman, Shedletsky, and Silver (1986), were the first to attempt a systematic validation of Clock Drawing. Using a five-point hierarchy to rank severity of errors, Shulman et al. (1986) reported a significant relationship between Clock Drawing and two measures of cognitive functioning. This study will be described further in a subsequent chapter. The present research focuses on the Clock Test (Tuokko, Horton, Hadjistavropoulos, & Beattie, 1990a), which consists of Clock Drawing, Clock Setting, and Clock Reading. These tasks were designed according to the suggestions proposed by Shulman et al. (1986), in which scoring would both provide a standardized method of quantifying Clock Drawings, and maintain the qualitative assessment of the process of the task. In addition to visuospatial and constructional abilities, the comprehension of time concepts are required in order to complete successfully the test. The present study investigated the extent to which the Clock Test is both a valid and reliable test for dementia.

The second chapter begins with a brief overview of the definitions of dementia, with particular emphasis on Alzheimer's disease (AD). In this section, the current definitions of dementia and Alzheimer's disease are presented, and the criteria used for its diagnosis are described, and the prevalence rate of AD is discussed. This is followed by a description of the literature dealing with visuospatial deficits and drawing abilities in dementia. In general, tests of dementia are often based on measures of verbal ability and therefore do not include measures of visuospatial abilities (Cummings & Benson, 1983). Specifically, the research of Moore and Wyke (1984) is described. These authors found significant correlations between tests of intelligence and drawing abilities in demented patients. They also proposed various explanations for their findings, concluding that demented individuals perform poorly on drawing abilities due to an attentional deficit. The subsequent sections provide a detailed review of the clock literature as it is related to both the normal aged and to those who suffer from such dementing disorders as Alzheimer's disease. Specifically, various scoring systems and presentation methods which are currently used in clinical practice are described in detail. In addition, a summary of the use of the Clock Test as an assessment tool for dementia is provided. Finally, the goals and hypotheses of the present study are presented in detail.

In the third chapter, the participants used in the present study as well as the procedure used to administer and score the Clock Test were described in detail. Specifically, the data from 456 participants were examined. These participants included referrals to the University of British Columbia's Clinic for Alzheimer's Disease and Related Disorders (who were classified into one of four diagnostic groups) and a sample of community dwelling adults who served as normal controls. The methods used to diagnose an individual along with a variety of demographic information are described. The subsequent sections describe the procedure used to administer each component of the Clock Test, the scoring criteria used to evaluate performance, and a description of the inter-rater and test-retest reliabilities. The final section describes the neuropsychological test battery which was presented to the participants of the study.

In the fourth chapter, the statistical analyses used and the results of the study were described in detail. This section is divided into three parts. The first part deals with issues encompassing the structure of the Clock Test. Specifically, the internal consistency of the Clock Test, the effects of age, education, and diagnostic group on performance were determined, and the differential performance, by the diagnostic groups, on Clock Drawing were examined. Furthermore, using the three components of the Clock Test, classification into groups based on cognitive status was determined. The second part deals exclusively with the construct validity of the Clock Test components. In this section, performance on each component of the Clock Test was compared to the performance on other validated neuropsychological measures. The final part uses confirmatory factor analytic procedures in order to examine the factor structure of the Clock Test relative to other neuropsychological measures.

The final chapter provides a summary of the results, in reference to the hypotheses presented in Chapter 2. In addition, a critical evaluation of the present study and the limitations of the Clock Test as an assessment tool are addressed. Lastly, suggestions for future research and modifications to improve the scoring schema of the Clock Test are presented.

Chapter 2
Literature Review

The present chapter reviews the literature regarding the three components of the Clock Test. Specifically, various scoring systems and the reported research findings will be described in detail. In addition, the present chapter includes a brief review of the nature Alzheimer's disease, and a description of the visuospatial deficits associated with dementia. Finally, the potential benefits of using the Clock Test are described, followed by the hypotheses of the present study.

Nature of the Deficit

Current definitions of dementia emphasize a general decline in intellectual competence after maturity, with symptoms specifically affecting memory, orientation, abstract thinking, judgement, personality, emotional control, and functional capacity (Diagnostic and Statistical Manual - Third edition revised, DSM-III-R; American Psychological Association, 1987). The National Institute for Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Diseases Association Work Group (NINCDS-ADRDA; McKhann, Drachman, Katzman, Price, & Stadlan, 1984) define Dementia as "the decline of memory and other cognitive functions in comparison with the patient's previous level of function as determined by a history of decline in performance and by abnormalities noted from clinical examination and neuropsychological tests" (p. 940).

Alzheimer's Disease, which was first described by Alois Alzheimer in 1907, is the most well known of the dementing disorders and probably accounts for more cases of dementia than any other single disease entity (Cummings &

Benson, 1983). It has the added characteristics of unknown etiology, an age-related incidence, a distinctive neuropathology (i.e., diffuse degeneration and atrophy of brain tissue with excessive numbers of neurofibrillary tangles and senile plaques concentrated in the cerebral cortex and hippocampal regions of the brain), an insidious onset, a chronic course, irreversible progression, and premature death (Terry & Katzman, 1983). In addition, AD is two to three times more common in women than in men (Cummings & Benson, 1983; Terry & Katzman, 1983). Death generally occurs 6 to 12 years after onset (Cummings & Benson, 1983) although a rapid progressive course can result in a quick death. Belsky (1984; see also Katzman, 1986) estimated both that Alzheimer's Disease may be responsible for up to 50 percent of all cases of dementias, and that approximately twenty percent of all institutionalized psychiatric patients have Alzheimer's Disease (Dewis & Baumann, 1982). Gurland and Cross (1986), however, estimated that AD accounts for approximately 75% of all cases of dementia in the general population.

Jorm, Korten and Henderson (1987) reviewed 47 studies which examined the prevalence rates for severe dementia among people aged 65 and over. The prevalence rates varied according to study, from a low of 0.6% to a high of 5.1%. The rate for mild dementia ranged from 1.6% to 13.1%. All of the studies acknowledged that the prevalence rates increase greatly with age. Specifically, there is a relationship between prevalence and age, such that there is a doubling of prevalence with every 5.1 years of age over 65 (Jorm et al., 1987). AD may be the fourth or fifth cause of death among older adults, although it is difficult to confirm since it is not reported as a person's official cause of death (Blieszner & Shifflett, 1990). Smith (1989, cited in the Canadian Collaborative Study, 1990) reported a prevalence rate of 22,000 cases of AD in Canada for 1987. He stated, however, that, since death certificates often fail to report Alzheimer's disease as the cause of death, a more realistic prevalence figure would be around 300,000 cases per year. Apart from the effect of age, little is known about the etiology of the disease although such causes as changes in immune functioning, aluminum toxicity, genetic factors and viral transmission have been explored (Gurland & Cross, 1986).

The NINCDS-ADRDA work group was established to describe clinical criteria for the diagnosis of Probable, Definite and Possible Alzheimer's disease. Although an in depth discussion of these criteria are beyond the scope of the present study, the following brief description is included. In order to establish a diagnosis of Probable Alzheimer's disease: (a) dementia must be established by a clinical examination and neuropsychological testing, (b) a progressive worsening of memory and other cognitive functioning must be observed, (c) there must be no disturbance of consciousness, (d) onset must occur between the ages of 40 and 90 (although most often after age 65), (e) deficits must be observed in two or more areas of cognition, and (f) there must be an absence of systemic disorders or other brain diseases that could account for the progressive deficits in memory and cognition. Definite Alzheimer's disease can be confirmed only with histopathological evidence, obtained from a biopsy or autopsy, which shows pathognomic changes in brain tissue (Lezak, 1983). In addition, the individual must have demonstrated the clinical criteria for a diagnosis of Probable Alzheimer's disease. In order to establish a diagnosis of Possible Alzheimer's disease: (a) the dementia syndrome must be present in the absence of other neurologic, psychiatric or systemic disorders sufficient to cause dementia, (b) variations in the onset, presentation, or in the clinical course of the dementia syndrome must be present, and (c) a second systemic or brain disorder sufficient to produce dementia, but which is not deemed to be the cause of the dementia, may be present. In clinical practice, Alzheimer's disease is diagnosed by the exclusion of other medical or psychological disorders. Specifically, when all other diagnostic possibilities have been eliminated and the patient displays the behaviourial alterations characteristic of Alzheimer's disease, this becomes the suspected diagnosis.

Memory and disturbed visuospatial discrimination is impaired with the onset of AD (Gainotti, Caltagirone, Masullo, & Miceli, 1980). Most neuropsychological research with demented participants, however, emphasizes verbally mediated episodic and semantic memory dysfunctions rather than visuospatial abnormalities (Cummings & Benson, 1983). In addition, visuospatial deficits may be the primary deficit in some patients in the early stages of the illness (Moore & Wyke, 1984). The following section describes the visuospatial deficits associated with dementia.

Visuospatial Deficits

The existence of a deficit of drawing ability associated with dementia has been widely accepted (Grossi & Orsini, 1978; Grossi, Orsini, & Michele, 1978; Moore & Wyke, 1984; Yaguchi, 1981). Dementia patients, therefore, are routinely assessed with tasks that require the copying of drawings which vary in degrees of complexity (Moore & Wyke, 1984). There is, however, no clear indication as to the extent and nature of such a deficit.

Disturbances in organizing objects in space are considered to be an early sign of the decline of the superior cortical functions in dementia (Grossi et al., 1978). In addition, demented participants show a high degree of disturbance in the ability to arrange objects (relative to one another) in space (Grossi & Orsini, 1978). For example, Yaguchi (1981) asked demented and normal participants to draw trees. Tree drawings by demented participants were more constricted and lower, in addition to being straighter and more centered on the page than normal participants. Moore and Wyke (1984) compared spontaneous drawings to copies drawn from pictures. Demented participants produced smaller and simplified representations for spontaneous drawings, whereas normal controls tended to draw extra lines. When spontaneous drawing were compared to copies, demented participants' copies were larger and contained a greater number of defining features than their spontaneous drawings. The copies, however, tended to be fragmented with incorrectly positioned features. These errors were absent from the drawings of normal controls. The demented participants' drawings included omissions of essential features, additions of labels as a substitute for some of the features of the drawing, perseverations across a series of drawings, and drawings that were small and cramped in appearance. Copies showed improved performance in terms of the number of essential details included, but these details were frequently positioned incorrectly, giving a scattered and fragmented appearance. Copies were also larger in size.

Moore and Wyke (1984) found highly significant correlations between four standardized scales of intelligence and drawing ability. They stated that demented participants' performance (in which they failed to include essential features in their drawings) cannot be accounted for in terms of a motor deficit, since the participants could copy simple geometric shapes accurately. In addition, performance could not be entirely accounted for in terms of a planning deficit, which has been used to explain performance by patients with cerebral lesions (Warrington, 1969). This explanation would predict that copies would include more essential features than spontaneous drawings and that accurate spatial relationships would be preserved in copies. Incorrectly positioned features are a typical finding in copies (Albert & Moss, 1984; Moore & Wyke, 1984). The omission of essential features in drawings could be accounted for by a memory impairment. It cannot, however, account for the lack of accurate spatial relationships in copies. Moore and Wyke (1984) provide a possible explanation in terms of an attentional deficit at the level of control processing.

Individuals fail to integrate separate features into a coherent whole, focusing attention instead on one feature at a time. The individual thereby fails to integrate information contained in separate aspects of the task.

The concept of time and the ability to comprehend and manipulate time as represented on a clock face involve highly abstract and complex higher order functioning. Clock Drawing, Clock Setting, and Clock Reading require the comprehension of time concepts, conceptual reasoning, and visuospatial skills for successful execution. The following sections investigate the literature on Clock Drawing, Clock Setting, and Clock Reading with particular emphasis on the different scoring systems that are presently being used.

Clock Drawing

Clock drawing ability was first referred to in the neurology and psychology literatures as a possible indicator of cerebral organic dysfunction (Albert & Kaplan, 1980; Albert & Moss, 1984; Critchley, 1953). It was used to investigate deficits associated with constructional apraxia (Benton & Fogel, 1962; Critchley, 1953). In addition, the ability to draw successfully a clock was described as a function of the parietal lobes (Goodglass & Kaplan, 1983). Critchley (1966) defined constructional apraxia as a "difficulty in putting together one-dimensional units so as to form two-dimensional figures or patterns" (p. 172). Van der Horst (cited in Critchley, 1966) described it as an interference with such actions as drawing from a copy, putting figures of a puzzle together, or building with bricks. Essentially, constructional apraxia is an executive defect within a visuospatial domain and one which is associated with other spatial disorders such as acalculia (a disturbance in doing arithmetic calculation) and dysgraphia. Benton and Fogel (1962) stated that constructional apraxia is a broad concept which has been applied to a number of different types of activities which have the common characteristic that they require an individual to assemble, join, or articulate parts in order to form a single unitary structure (or whole). The activities may differ, however, in the complexity, type of movement and degree of motor dexterity required to achieve the task, in the demands made on the higher intellectual functions, and in whether or not they involve construction in two or three spatial dimensions. Critchley (1966) described the following characteristics of drawings of constructional apraxia: (a) drawings would be crowded onto one of the corners of the sheet (i.e., there is a neglect of space to one side of the midline), (b) the copy would be considerably smaller than the design, (c) lines would be wavy, (d) lines would not accurately meet each other, (e) scribbles, resulting from several lines put in where there should only be one would be present, and (f) large sections of the design may be omitted altogether.

Although thought to be constructional in nature, clock drawing is a highly complex task that requires a number of skills in addition to constructional abilities (e.g., Anderson, Dobbs, & Rule, 1990; Farver, 1975; Sunderland et al., 1989; Tuokko et al., 1990a). For example, Farver (1975) claimed that in order to correctly tell the time on an analog clock, one must possess the ability to do mathematical procedures. Farver adds that an inability to tell time and set clocks is closely related to acalculia since telling time involves being able to give mathematical meaning to spatial relationships (with the 12 points around the circle representing both single units or hours and multiple units or five minute intervals), and the relative lengths between the minute and hour hand. Mathematical operations can be described as "quasi-spatial" in nature (Luria, 1963) since mathematical operations are based on one's ability to place the value of digits in multiple digit numbers (e.g., 55 vs. 555). Tuokko, Hadjistavropoulos, Miller, and Beattie (1992), however, state that little attention has been directed towards determining whether or not a deficiency in clock drawing has any functional significance with respect to an individual's ability to use time concepts. Lezak (1983) maintains that instructing a patient not only to draw a clock but also to set it to a specific time provides information about the patient's ability to handle a free drawing task. In addition, as the patients' concept of time orientation and their capacity to process numbers and number/time relationships is examined (see Edlund, 1987; Goody, 1988, for a review of time-related behaviours and subjective feelings of time).

Spreen and Strauss (1991) state that, in contrast to the primarily verbal content of most dementia scales, clock drawing relies on visuospatial, constructional, and cognitive abilities. In addition, Sunderland et al. (1989) state that, while clock drawing does involve verbally mediated cognitive processes, the test reflects more visuospatial skills than it does purely verbal memory tasks. As such, clock drawing can provide complementary information about demented participants, and it can also be used to screen patients with visuospatial deficits, particularly because most available screening tests primarily assess verbal memory and orientation. Albert and Moss (1984) state that Alzheimer patients often show perseverative tendencies (e.g., repetition of numbers and hands), and are drawn to the concrete content of the instructions (e.g., place the hands on the ten and the eleven, write the words 'ten' and 'eleven'). According to these authors, performance does not result from a misperception of the stimulus, rather errors are the result of a previous type of response. If the nature of the error was overlooked and only test scores were noted, one may conclude that these individuals had visuospatial deficits. When asked to copy a picture of a clock, however, individuals with Alzeimer's disease will successfully perform this task. This allows one to conclude that the deficit may be due to something other than a visuospatial deficit.

Goodglass and Kaplan (1983) used Clock Drawing and Clock Setting (to be discussed later) as nonverbal subtests for the Boston Diagnostic Aphasia

Examination. This test is comprised of a battery of tests that measure parietal lobe functioning. The test battery was given to normal adults varying in age and in levels of education. They asked participants to: (a) draw the face of a clock, (b) place the numbers at their correct location, and (c) set the two hands at ten minutes after eleven o'clock. This time was chosen in order to increase the difficulty of the task. Previous methods required that a participant set the hands of the clock to 20 minutes after 8 o'clock (in order to have one hand in each parietal hemifield, thus having an opportunity to check for the presence of neglect). However, since there is no 20 on a clock face, it is easier to recode the 20 and to focus on minutes, than it would be for the clock setting of 10 after 11 in which both numbers are present on a clock face thus requiring that the participant comprehend the spatial nature of a clock. The clocks were scored on the basis of a three point scoring system in which one point each was given for (a) an approximately circular face, (b) symmetry of number placement, and (c) correctness of number. The setting of the hands was not scored, rather Goodglass & Kaplan (1983) stated that they "provide valuable qualitative information" (p. 58). The authors, however, did not describe the information obtained by the setting of the hands. Although unsophisticated in nature, the clock test has since been scored in a number of ways.

Shulman et al., (1986) were the first to attempt a systematic validation of the clock drawing test. They developed a hierarchical classification system of errors, based on the qualitative aspects of clock drawing, as a practical screening tool to detect the presence of cognitive impairment in old age. The qualitative categories include a range of defects in visuospatial abilities (e.g., impaired spacing of times, neglect of numbers or hands, times drawn on outside of clock), abstract thinking, right-left confusion (e.g., numbers drawn counterclockwise), and perseverations (e.g., repeats circle, numbers, or hands, continues writing numbers past 12). In this task, the participant was presented with a stimulus sheet that contained a predrawn circle. The individual was required to place the twelve numbers and two hands on the clock face, setting the time to 3 o'clock. The clocks were then scored according to a 5-point hierarchical classification of severity of errors which ranged from mild to severe (see Appendix A).

The correlations between severity of clock drawing errors and performance on the Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975) and the Short Mental Status Questionnaire (Robertson, Rockwood, & Stolee, 1982), which are considered to be valid and reliable measures of cognitive functioning, were computed. Participants with low scores on the measures of cognitive functioning also had clock drawing errors of a more severe nature. When the participants were given a measure of depression (Geriatric Depression Scale; Yesavage et al., 1983), their level of depression was not related to cognitive functioning, as measured by the cognitive test battery and the clock drawing task. These findings contradict past research which were conducted using other cognitive measures (e.g., Cummings & Benson, 1983; Gallagher, 1986; Miller, 1975; Weingartner, 1986) in which the direct relationship between depression and cognitive impairment was evident.

Conflicting research has been reported on the relationship between depression and cognitive impairment on clock drawing performance. It has been hypothesized (Dastoor, Schwartz, & Kurzman, 1990; Sinoff, 1991) that since clock drawing is a higher cortical function it would not be affected by the presence of depression. This relationship is important since up to half of the patients diagnosed with early dementia are also suffering from depression (Blazer, 1989). In addition, cognitive impairment due to depression is in most cases reversible (Cummings & Benson, 1983). Dastoor et al. (1991) argued against the conclusions forwarded by Shulman et al. (1986) on the basis that the Geriatric Depression Scale becomes less valid as the degree of cognitive impairment increases (Brink, 1984; Spreen & Strauss, 1991), and the Short Mental Status Questionnaire is not sensitive to mild forms of cognitive dysfunction (Shulman et al., 1986). Based on the conclusions of Shulman et al. (1986) the relationship between clock drawing and depression was inconclusive and required further investigation.

Two studies, used the scoring criteria devised by Shulman et al. (1986) in order to further examine whether clock drawing can differentiate between demented and depressed individuals. In the first study, Dastoor et al. (1991) examined clock drawing abilities in participants diagnosed with dementia of the Alzheimer type, multi infarct dementia, or depression. Using a modified version of Shulman's scoring criteria, different patterns of clock drawing errors were observed for the demented and depressed participants. Depressed participants not only made fewer errors than both groups of demented participants, but the types of errors that they made were less severe. In addition, the depressed participants had higher scores (thus reflecting less cognitive impairment) than both groups of demented participants on a cognitive test battery. It was concluded that when an individual performs poorly on a cognitive screen, but draws a perfect or near-perfect clock, s/he is more likely to be suffering from depression than dementia.

In the second study, Sinoff (1991), also used Shulman et al.'s (1986) classification of clock errors to examine the relationship between depression, dementia, and clock drawing and found conflicting results. Similar to previous studies, the author found that participants who drew a normal clock were less likely to be demented. Sinoff (1991) concluded, however, that with demented and non demented participants, one is unable to distinguish between depressed and demented participants based on clock drawing scores. Presently, the results are inconclusive as to whether clock drawing scores can be used in order to differentiate between demented and depressed individuals.

Shulman et al. (1986) developed the clock drawing test as a measure of gross cognitive dysfunction with the suggestion that it could be used as a screening tool for cognitive impairment in the elderly. They did acknowledge, however, that a new scoring system should be devised that would both provide a standardized method of quantifying clock drawing, and maintain the qualitative assessment of the process of the task. Although not researched at the time, these criteria were considered important since different types of errors may reveal different underlying sources of deficits. Taken together, a valid and reliable indicator of cognitive impairment would emerge.

Following Shulman's (1986) recommendations, several new methods for clock drawing as a means of assessing cerebral status have been proposed. In addition, clock drawing has become of sufficient interest to merit more precise evaluation. Various test procedures are presently being used. These methods differ in five major considerations. First is whether the participant is given a predrawn circle (e.g., Friedman, 1991; Tuokko et al., 1990a), is required to construct a clock on a blank sheet of paper (e.g., Anderson et al., 1990), or is simply required to place the minute and hour hands to indicate a specific time (e.g., Albert & Kaplan, 1980; Farver, 1975). Second, the times to which the clock must be set. Third, whether or not there exists a center point for the placement of the hands (e.g., Sunderland et al., 1989; Tuokko et al., 1990a). Fourth, whether or not anchor points are included (i.e., the number 12) which aid and help guide the participants to place the remaining numerals (e.g., Doyon, Bouchard, Morin, Bourgeois, & Côté, 1990). Fifth, whether of not the placement of the hands is required (Friedman, 1991). In addition to the methods used by researchers to present clock drawing, a variety of scoring systems have been proposed. These systems include: (a) the rating of clock drawings from best to worst (e.g., Sunderland et al., 1989; Wolf-Klein, Silverstone, Levy, Brod, & Breur, 1989), (b) assigning clock drawings to discrete error categories based on clinical experience and rough guidelines (e.g., Wolf-Klein et al., 1989), (c) using personal and subjective beliefs as to what constitutes a good representation of a clock in order to rate clock drawings (e.g., Sunderland et al., 1989), (d) grading clocks on a pass-fail dichotomy (Friedman, 1991), and (e) scoring clocks based on both the qualitative and quantitative aspects of the clock drawings in order to better appraise it (e.g., Tuokko et al., 1990a). The following sections describe in greater detail the procedures used by various researchers to present the clock drawing task, and the methods used to score them.

Methods of Clock Presentation and Scoring

The Alzheimer Clinic - University of British Columbia. Research from this clinic uses a dual scoring system (presented as Appendix B). This system assesses both the general level of efficiency of performance (quantitative appraisal) and the specific types of errors made (qualitative appraisal). Briefly, the system includes seven dimensions, or factors, each composed of a number of elements. These factors are Omissions, Perseverations, Rotations, Misplacements, Distortion, Substitutions, and Additions (these factors are described in greater detail in a subsequent chapter; See also Appendix B). In addition to Clock Drawing, two tasks were added: Clock Setting and Clock Reading. These tasks were included since high level conceptual abilities, which may be particularly sensitive to cognitive change, are required to perform these tasks. In addition, with these additional tasks, a participant's concept of time is more thoroughly examined.

In the first of a series of studies, Tuokko, Olson, Peters, Horton, Beattie, and Crockett (1988) examined the relationship between the three parts of the clock test (i.e., drawing, setting, and reading) which were given to normal elderly adults, and referrals with a suspected memory disorder (SMD). The latter group consisted of a heterogeneous sample of participants differing with respect to cognitive impairments and level of severity. The performance of the normal adults was significantly better than that of the SMD participants on all three tasks. When the participants were classified to their respective groups based on the test scores, 85% of the participants were correctly classified. Specifically, 95% of the normal, but only 71% of the impaired participants were correctly assigned to their appropriate group.

In subsequent analyses, the SMD group was divided into subgroups reflecting level of functional impairment (based on the Functional Rating Scale (FRS; Crockett, Tuokko, Koch, & Parks, 1989; Tuokko & Crockett, 1989; Tuokko, Crockett, Beattie, Horton, & Wong, 1986). The FRS (to be discussed in further detail in the Method section) was designed to assess important skills which are evaluated when making a diagnosis of dementia. Using the three scores of the Clock Test to assign individuals to groups reflecting one's level of impairment resulted in an accuracy rate of only 59.7%. This indicates that although the Clock Test can differentiate between normal and impaired participants, it does not reliably discriminate between levels of impairment. Further research is required to address the issue of classifying individuals into groups based on the level of impairment.

In a second study, Tuokko, Beattie, Crockett, and Horton (1989) compared normal elderly participants, patients with Probable Alzheimer's disease, and patients with other memory disorders, on each of the three components of the Clock Test. As would be expected, normal elderly participants performed significantly better than both memory impaired groups on all three components of the clock test. Tuokko et al. (1989) then categorized the memory impaired participants, based on the FRS, into groups reflecting level of functional impairment. Across all three tasks participants with the least amount of functional impairment performed at the highest level. Further analyses revealed that Clock Drawing scores remained relatively constant over questionable to mild levels of impairment, but decreased significantly with moderate to severe impairment. Clock Setting and Clock Reading, on the other hand, seemed to demonstrate relatively consistent rates of decline as the level of functional impairment increase. The authors conclude that Clock Setting and Clock Reading may be measures of the abstract conceptualization of time, whereas Clock Drawing may, to a greater extent, be dependent on constructional or visuospatial skills. Thus, disorientation to time as determined through Clock Setting and Reading may be functionally different from that of Clock Drawing.

In a follow-up study, Tuokko, Hadjistavropoulos, and Beattie (1990b) examined whether patients with a diagnosis of dementia simply make more errors than normal individuals or whether their errors are of a distinct and characteristic nature. This issue is of interest because if certain types of errors may be characteristic of specific types (and levels) of dementia, then it may be possible to developed a refined scoring system, based on error types, with a powerful ability to discriminate.

With the exception of Rotation errors (see Appendix B), all of the error categories significantly discriminated between the normal and demented elderly. Rotation errors include such errors as the rotation of the clock face or specific numbers, as well as such errors as numbers being written backwards or the time being written in its mirror image (see Appendix B). The fact that these errors (which include both number and hand rotations) do not differentiate between the groups is not surprising since it has been shown that this type of error frequently occurs in the clock drawings of normal elderly adults (e.g., Albert & Kaplan, 1980; Kaplan, 1989). Tuokko et al. (1990b) state that although no specific type of error was representative of patients with dementia, there was a tendency for the demented participants to produce more Omission errors (i.e., neglect of part of the clock face, omission of numbers or hands) than any other type of error. The authors concluded that the total number of errors yields the highest overall correct rate of classification into dementia versus normal groups.

In the most recent study (Tuokko et al., in 1992), cut-off scores were derived for the three parts of the clock test in order to maximize the separation between normal and demented groups. For Clock Drawing, a cut-off score of two or more errors was suggested to reflect cognitive impairment. Using this criterion to classify participants into either normal or demented groups, 92% of the normal and 86% of the demented patients were correctly classified. For Clock Setting, a cut-off score of 13 correct responses resulted in the correct classification of 87% and 97% of the normal and demented individuals, respectively. Finally, for the Clock Reading task, a cut-off of less than 13 correct resulted in the correct classification of 92% and 85% of the normal and demented participants, respectively. In addition, for both the Clock Setting and Clock Reading tasks, the two groups differed significantly on all items, with the performance of the normal adults superior to that of the demented participants. In a final analysis, the authors used an impairment criterion of two or more of the three test components to classify individuals to their respective groups. Using this method, 94% of the normal, and 93% of the demented patients were correctly classified.

Longitudinal Adult Memory Project. Although used primarily as a research tool, Clock Drawing and Clock Setting (discussed in a subsequent section) were included in the assessment by the Longitudinal Adult Memory Project (LAMP) at the University of Alberta. Data were collected from over 500 normal participants ranging in age from 30 to 100 years of age (Dobbs, personal communication, March 1991). In the Clock Drawing task, the participant was given a blank sheet of paper and was required to draw a clock which included the clock outline, the 12 numerals and the two hands, set to 3 o'clock.

In the LAMP project, clock drawings are scored according to the following factors: (a) Overall judgement, or a subjective rating scored on a 5-point Likert scale ranging from 1 (inadequate) to 5 (adequate), (b) Clock outline, (c) Numbering, and (d) Hands. Dimensions B through D are scored on a 3-point Likert scale, from 1 (adequate) to 3 (extreme errors) and each consist of a number of elements. A final category, Other, includes such factors as the presence of irrelevant figures, words, or embellishments (which are rated according to the degree of irrelevancy), and the use of anchor points (i.e., 3, 6, 9, 12).

The clock outline score includes such elements as: (a) Location of clock relative to the four quadrants of the page, (b) Symmetry, or the degree to which parts of the clock outline correspond, (c) Diameter of the clock outline, (d) Closure, or the degree to which the end point of the outline meets the beginning point, and (f) Perseveration, or continuous drawing of the clock outline after the first outline is completed. Numbering includes the following elements: (a) The presence of the numbers one through twelve in sequence in either arabic or roman numeral notation, (b) The correct orientation of the clock face, (c) Spacing, or the placement of the numbers at approximately equivalent intervals, (d) The consistency in the numbering systems used, and (e) Such perseverations as the repetition of numbers or continuation of numbers beyond 12. Spikes may be substituted for the numbers if they are placed in the proper number locations. In addition, the use of the four anchor points is noted if they aid a participant in setting the correct time. Hands includes the following elements: (a) The differentiation between the minute and hour hand on the basis of length, (b) The degree to which the hands meet at the designated center of the clock face, and (c) Hand perseverations, or whether or not more than two hands are drawn.

Presently, only preliminary results for this scoring system are available. Anderson et al. (1990) matched normal elderly adults and Alzheimer's patients on age and level of education. When the scores were added across all the factors, the performance of the normal participants was significantly better than that of the Alzheimer patients. When each factor was examined separately, the Numbering factor was the only element to differentiate between the two groups, with Alzheimer patients performing worse than the normal participants. Within this factor, the spacing element was the most differentiating measure. In addition, those Alzheimer patients who used anchor points to aid in the placement of numbers made fewer spacing errors than those who did not use this technique. Anderson et al. (1990) stated that visuospatial skills are required more for correctly spacing numbers than for some of the other measures. The authors concluded that although clock drawing may involve verbally mediated cognitive processes, it appears to be more of a visuospatial than verbal memory task.

Other Scoring Systems. Additional scoring systems have been reported in the literature. The first three are similar in that they are both based on a 10-point hierarchical system for the evaluation of clocks, whereas the fourth is similar to previous scoring systems which score errors by type. The hierarchy of Sunderland et al. (1989) was based on the qualitative nature of clocks which were drawn (i.e., clockface, numbers, and hands) and set to 2:45. The hierarchy ranged from 1 (worst) to 10 (best). The scoring system contains several methodological problems. First, the scale is vague since there are neither objective nor specific defining criteria for the points. Second, the defining criteria for the points are relative and subjective (e.g., point 9 "slight errors in placements of hands" or point 8 "more noticeable errors in the placement of hour and minute hand" do not specify defining criteria). Third, there is no support that the criteria truly represent a best to worst hierarchy. For example, no data are presented to support the hypothesis that, for example, point 6 (Inappropriate use of clock hands) actually represents a clock that is better than one which receives a 4 (Further distortion of number sequence). Fourth, the cut-off point of 6 to represent a normal clock is arbitrary. Fifth, it is assumed through the scoring criteria that the representation of the hands are affected first and additional errors in the representation of numbers and the clockface occur later. Therefore, some drawings will receive low scores on the scale because of minor errors in the representation of the numbers even though the hands were correctly set (Rouleau, Salmon, Butters, Kennedy, & McGuire, 1992). In addition, although the inter-rater reliabilities are high (ranging from .86 to .97 for clinician and nonclinician groups), the authors note that when the clocks of normal controls and Alzheimer participants were scored together, the Alzheimer clocks worsened by a mean of .5 when compared to ratings obtained when both groups were scored independently. Inversely, the ratings from controls improved by a mean of .3 when rated together with Alzheimer patients. Although these differences are not large, they illustrate the problems which can arise when a scoring system is not specifically operationalized.

Sunderland et al. (1989) did not claim that the clock test was either sensitive or specific to dementia. They claim, however, that it can be performed by patients with dementia and it can be reliably rated by both trained clinicians and nonclinicians. Furthermore, it is consistent with previous research in that Alzheimer's patients performed significantly poorer than normal controls, and there were high correlations between clock drawing and standardized measures of dementia severity. In addition, the authors reported non-significant correlations between clock score and age in both the normal and demented participants, and non-significant correlations for duration of symptoms in Alzheimer's patients and clock drawing. Overall, since the present system is based on a qualitative hierarchy which prohibits a quantitative analysis of errors, it is inferior to other scoring systems as a sensitive measure of dementia. It does however have the advantage of being simple to score while still providing some important clinical information.

Based on clinical experience, Wolf-Klein et al. (1989) identified 10 clock patterns and linked them to mental status. They presented participants with a four inch predrawn circle and instructed them to draw a clock. No other instructions were given. This method is different from those previously described in that the participants are not given a time to which the hands should be placed. The 10 clock patterns range from 1, "Irrelevant Figures", representing the worst possible clock to 10, "normal", representing the best or perfect clock pattern. Wolf-Klein et al. (1989) did not score the placement of the hands, but based the patterns on the presence, placement, and sequence of the numbers. The authors concluded that clock patterns seven through ten indicate a clock drawn by a "mentally normal" individual, whereas patterns one through six are indicative of Alzheimer's disease. These classifications are based on the clock patterns drawn by two groups of participants and require further testing in order to be validated.

Wolf-Klein et al. (1989) reported a sensitivity (i.e., number of Alzheimer patients with abnormal clock patterns divided by the number of Alzheimer patients) of 75%, and a specificity (i.e., number of normal participants with normal clock patterns divided by the number of normal adults) of 94%. Although the sensitivity and specificity were high, the authors failed to show that the clock test was sensitive to mild cognitive impairments. In fact, they excluded clock pattern 8 "almost normal except for spacing" because of its "indeterminate relation to mental status" (p.733). This clock pattern was produced regularly by both normal and demented participants, and after normal clocks, it was the most frequent pattern. Like the previous scoring system, data were not provided to support the order and range of clock drawing errors.

Using clock drawing, Rouleau et al. (1992) compared the constructional and visuoperceptive impairments of Alzheimer's patients and patients with Huntington's disease who were matched for overall level of dementia. Participants were asked to draw a clock (including the circle, numerals, and hands) and set it to 11:10. In addition, they were asked to copy, as accurately as possible, a clock from a model. The method developed by Sunderland et al. (1989), and a new 10-point scoring system, were used to evaluate the clocks. The new scoring system was designed to assess independently the accuracy of the representation of the clockface (a maximum of 2 points), the layout of the numbers (a maximum of 4 points), and the position of the hands (a maximum of 4 points). In addition, in order to examine the underlying mechanisms involved in clock drawing and clock copying, the following qualitative dimensions were examined: (a) clock size, or whether the clock was too large or too small; (b) graphic difficulties, or whether the lines were precise, the hands were straight and connected in the middle, the clockface was distorted, or the numbers were difficult to read; (c) stimulus-bound response, or the tendency of a drawing to be dominated or guided by a single stimulus (e.g., the time is written in, the time was set at 10 to 11 instead of 10 after 11); (d) conceptual deficits, or the misrepresentation of the clock or the hands; (e) spatial/planning deficit, or the neglect of the left hemisphere of the clock, or a deficit in the spatial layout of the numbers (e.g., large space between numbers, numbers written outside the clockface, or numbers written counterclockwise); and (f) perseverations, or the presence of more than two hands, the extension of numbers beyond 12, or the inappropriate recurrence of the same numbers.

Normal controls outperformed the Alzheimer's and Huntington's groups on both clock drawing and clock copying, although the demented groups did not differ from one another on either task. The Alzheimer participants, however, performed significantly better under the clock copying condition than under the clock drawing condition, whereas the normal and Huntington disease groups showed no difference between the two conditions. The errors of Alzheimer patients included errors in the spatial layout of the numbers, the size of the clocks (i.e., they were very large), and perseveration of the numbers and hands and other stimulus bound responses. The authors state that, when Alzheimer patients lack a model to guide the drawings, they appear unable to retrieve the specific attributes that define a clock and are therefore unable to draw an accurate representation of a clock. They add that the improvement under the copy condition suggests that the drawing deficiencies are not due primarily to graphic, motor, or visuospatial difficulties, rather the deficit is due to a loss of the semantic association usually evoked by a clock. The performance of Huntington patients was characterized by the presence of graphic deficits. They drew smaller clocks, and there were planning deficits in the spatial layout of the numbers. The authors state that the planning deficit may result from an inability to write simultaneously the numbers in the right order and space them correctly around the clock. This hypothesis is further supported by the fact that Huntington patients often self-corrected their clocks.

Nussbaum, Fields, and Starratt (1992) compared the clock drawing scoring systems of Sunderland at al. (1989), Wolf-Klein et al. (1989), and Rouleau et al. (1992) in a heterogeneous sample of 100 patients. The three scoring systems had high inter-rater reliabilities (ranging between .79 and .93) and were significantly correlated with other standardized screening instruments of dementia. Although similar results were obtained across the three scoring systems, they differ with respect to the rating criteria used. For example, two of the scoring systems require the appropriate placement of clock hands (Rouleau et al., 1992; Sunderland et al., 1989), and the other does not. In addition, whereas one of the procedures presents the participant with a clock face (Wolf-Klein et al., 1989), the other two include the clock face in the overall score.

The fourth scoring system is that of Doyon et al. (1990). They presented participants with a predrawn circle, which contained the number 12, and were asked to write the remaining numbers and set the time to 10 past 11. Scoring was based on the following criteria: (a) numbers, or a half point for the presence of each of the remaining 11 numerals; (b) number location, or a half point for the correct placement of each of the numerals; (c) hands, or two points for the presence of each hand; (d) hand location, or two points for the correct placement of each of the two hands; and (e) hand length, or one point if the hour hand is shorter than the minute hand (which must be pointing to the correct time). No penalty is given for such perseverations as extra numbers or hands. An individual can receive a maximum of 20 points. Significant differences were observed between normal controls and patients with Alzheimer's disease. In addition, using a cut-off score of 14.5 to indicate normal performance resulted in a sensitivity rating of 88.5% and a specificity of 94.9%.

Summary. This section described various reported systems that are used in order to score clock drawing. It has been shown that all of these systems can be used to differentiate between normal controls and participants with Alzheimer's disease. Although some of these systems have weaknesses, further research would be required in order to determine which of these scoring systems is the best. Simplified scoring systems cannot address the subtleties of disturbance which are likely to be of most importance in terms of diagnostic sensitivity (Tuokko et al., 1992). The UBC method has not yet been empirically contrasted to any of the other scoring systems. It is the contention of this author, however, that the hierarchy approach to scoring clocks is inferior to the present approach which consists of both a quantitative appraisal of performance and a qualitative appraisal of types of errors made. Although beyond the scope of the present study, which is concerned with the validity of the Clock Test, future research would be required in order to confirm whether or not the present method is, in fact, better than any other.

Normal Aging and Clock Drawing

Researchers (e.g., Borod, Goodglass, & Kaplan, 1980; Farver, 1975; Goodglass & Kaplan, 1983) have reported that normal adults show age related changes, with older adults having more difficulty than do younger adults on clock drawing and clock setting. Borod et al. (1980) reported a consistent trend in clock drawing favouring younger adults. That is, performance consistently decreased as a function of age. In addition, they reported a consistent trend for education, with performance increasing as a function of years of education. Farver (1975) used clock drawing, clock setting, and clock copying as part of a test battery which examined the functioning of the parietal lobes in old age. Across age, 84% of the participants were able to draw a satisfactory clock and received full credit for it. The most common errors included a failure to write the numbers symmetrically around the clock. Incorrect numerals were a less frequent error, but nonetheless occurred in the drawings of some of the older participants.

As part of a normative study on older aging adults, Albert and Kaplan (1983; see also Borod et al., 1980) included clock drawing. The participants were required to draw a clock which included the circle, the numerals, and the hands (set to 11:10). Clock errors for individuals with high scores on a neuropsychological screening battery included such errors as inappropriately drawn hand lengths, a misplaced center point (i.e., the center of the clock was pulled away from the middle towards the 11), and incorrect hand setting (i.e., one hand was set on the 10 and the other on the 11). Individuals with low scores on the neuropsychological screening battery were more severely impaired. Their errors were similar to those of high scorers and included production errors (i.e., the point at which the hands meet is pulled to the 11), and hand setting errors. There was a tendency, however, for low scorers to focus on the most salient bits of information (i.e, the 10 and the 11). This is seen in examples where the participant draws a line directly from the 11 to the 10, or when the participant writes the words "clock", or "10 past 11", rather than performing the task. This shows that the participants' errors are not based on a failure to comprehend the command, but result from a focus on the instructions themselves. Albert and Kaplan (1980) stated that the "errors seem to be the result of a pull to the most prominent bits of information and a failure to appreciate the relationships between the parts and the parts to the whole" (p. 421). They concluded that the quality of degraded drawings of low scoring elderly adults were consistent with the productions of brain damaged patients with lesions lateralized to the right hemisphere and focalized to the frontal system. In addition, Poon (1985) stated that degraded performance in old adults may be due to a low level of arousal and attention which is mediated by the frontal system of the brain.

Farver (1975) tested healthy nonhospitalized elderly participants (M age = 55.39 years; range = 40 - 88 years) on a number of tasks associated with right and left parietal lobe function. She found age-related changes on clock setting and other tasks with a strong visuospatial component, most of which are associated with right hemisphere functions. She stated that some of the worst drawings of the elderly participants were comparable to those of brain damaged adults. There were examples of simplified drawings, drawings reduced or enlarged in size, loss of perspective, rotations, and errors of spatial relations. Almost all of the participants' drawings improved when they were asked to copy a picture of a clock.

There were, however, no difference in scores between Probable Alzheimer's and other memory disorders. It is important to note that normal elderly participants also make a number of errors of various types on all three tasks, particularly with respect to clock drawing. This is not surprising since it is well-documented that visuospatial skills tend to decline with age (Gaylord & Marsh, 1975).

Clock Setting

As previously mentioned, Goodglass and Kaplan (1983) used clock setting as a nonverbal subtest in the Boston Diagnostic Aphasia Examination. In the Clock Setting task, participants were given a sheet with a clock face and short lines marking the positions of the 12 numbers. The participants were required to draw in the minute and hour hands for four times (i.e., 1:00, 3:00, 9:15, and 7:30). The participants received one point for the correct placement of each hand and an additional point for correctly indicating the relative lengths of the hour and minute hands. In a sample of normal elderly adults, no age differences were found until age 60, when performance decreased. In addition, adults with an education level over 13 years, outperformed those adults who had less than eight years of formal education. Farver (1975), however, found no significant age differences until the age of 70. In addition, less than half of the individuals tested received full credit on this task, indicating that errors are common regardless of age or cognitive standing. The participants were also given the task with a prenumbered clock face. Using this method, no age differences were observed. Findings based on level of education, however, were the same as those previously mentioned and performance did not improve with the aid of the numbers. Borod et al. (1980) reported significant age and education effects, with performance decreasing as a function of age, but increasing as a function of education.

As previously described, Anderson et al. (1990) used Clock Drawing and Clock Setting as part of the LAMP study. In the clock setting part of the test, the participant was given two test sheets which contained a predrawn clockface (i.e., it contained a circle, a center point for the placement of hands, and the 12 numerals along the inner circumference of the circle). The participant was asked to set the first clock to 10 after 11 o'clock and the second to 20 after 8 o'clock. The clocks were scored on a 3-point Likert scale, from 1 (adequate) to 3 (extreme) and included the following factors: (a) differentiation of hands, or whether the relative lengths of the two hands are unambiguous; (b) closure, or the degree to which the hands meet or fail to meet at the designated center; (c) accuracy of time, or the degree of accuracy in placement of the hour and minute hands in designating the specified time; and (d) perseverations, or the number of hands that are drawn. Regardless of the number point to which the hands are directed, the shorter hand is designated as the hour hand. In addition, a final category is scored and includes irrelevant figures and embellishments (which are rated according to their degree of irrelevancy), whether a number is circled and/or whether additional numbers are written in order to designate the time, and whether words are included. Anderson et al. (1990) found no differences between the scores on the two predrawn clocks. The performance of Alzheimer's disease patients and normal elderly adults matched on age and education was examined. The cumulative score across all of the factors indicated that the performance of normal participants was significantly better than that of the Alzheimer's patients. The accuracy of time factor reliably differentiated between the two groups, with normal controls making significantly fewer of these errors than the Alzheimer's patients.

Goodglass and Kaplan (1983) hypothesized that two skills are involved in clock setting: (a) The comprehension of number relations, and (b) geometric representation. Thus, clock setting seems to be sensitive to spatial-quantitative impairments. Farver (1975) reported that a number of errors made by normal adults in Clock Setting were of a visuospatial nature. Common errors which were similar to those presented for Clock Drawing include: (a) Incorrect placement of the hands, (b) placement of the hands towards the center of the clock, (c) placement of the hands in positions that made no reference to the center of the clock, and (d) reversal of the long and short hands (i.e., 3:45 instead of 9:15). The most common error, however, was of a more subtle nature and consisted of a neglect in the relative lengths of the two hands. Farver (1975) concluded that participants who made these errors "showed an appreciation of the spatial relationships of the clock face, but failed to demonstrate the significance of the long and short hands for the designation of hours and minutes" (p. 99). It can therefore be concluded that in order to successfully complete the clock setting task, one requires visuospatial skills, although not to the extent required by Clock Drawing.

Clock Reading

There is relatively little reported research on Clock Reading. In one report, Lowenstein et al. (1989) used Clock Reading, in addition to orientation to date, as part of the Direct Assessment of Functional Status scale (DAFS). These two dimensions comprise the Time Orientation subscale. Normal controls, individuals diagnosed with Alzheimer's disease, and depressed participants were asked to tell the time at four progressively more difficult clock settings (i.e., 3:00, 8:00, 10:30, and 12:15). In addition, they were asked to state the day of the week, the month, and the year. For each question, the participant was able to receive a maximum of 2 points for a correct response.

The time orientation subscale revealed a high interrater reliability and high test-retest reliabilities for both the demented and normal participants. The demented participants' performance did not differ from the performance of the depressed or normal participants on Clock Reading. However when Clock Reading and the Orientation to Date subscale scores were combined, the demented participants' performance was inferior to that of the normal and depressed adults. The authors concluded that one's ability to read time from a clock may not be affected until later stages of the disease process. In addition, clock reading does not appear to be subject to cultural biases when compared to common measures of cognition and language, since it has been translated and administered (along with the other subscales of the DAFS) to non-English speaking individuals from various cultural backgrounds. In addition to a visuospatial component, Clock Reading may be a verbally mediated task.

Johansson, Zarit, and Berg (1992) included Clock Drawing, Setting, and Reading in a Swedish study which examined the cognitive changes during a two-year interval in a representative sample of the oldest old (i.e., M age = 87 years; range = 84 - 90 years). The study was designed to use familiar stimuli and procedures which would not push the participants to their limits. The procedure was similar to that used by Goodglass and Kaplan (1983), with the addition of Clock Reading. The participant was shown a wooden clock which was set to different times and was asked to respond to the question "What's the time?." Longitudinal changes were observed with a decline in performance over the two-year period. In addition to visuospatial skills, Clock Reading required more verbal skills than both Clock Drawing and Clock Setting.

The Clock Test as a Test of Dementia

For bedside or office evaluation, a clinician should use an examination that quickly probes wide areas of patient functioning (Cummings & Benson, 1983). The purpose of the Clock Test is not to replace currently available scales, rather it is to introduce an easily administrable test of visuospatial skills that would supplement other measures and add important information that would improve the screening ability of other measures. The Clock Test, however, is a clinical screening instrument which may be useful as a quick and easy screening device to identify gross cerebral dysfunction.

Possible advantage of the Clock Test over longer and more complex diagnostic devices could be that it: (a) is relatively free of cultural and educational biases, (b) is easy and quick to administer, (c) is non-threatening to the patient (Wolf-Klein et al., 1989), (d) can be quickly scored, reliably measured, and easily interpreted by both trained clinicians and untrained observers (Sunderland et al., 1989), and (e) has minimal practice effect (Shulman et al., 1986). Although these possible advantages are not addressed empirically in the present study, they will be discussed further in a subsequent chapter. It can be argued that, as digital clocks become increasingly popular, the ability to tell time in the "traditional way" will be compromised, but until that time, diagnosis of patients in the present generation of middle-aged and elderly people may benefit from using the clock test. This potential limitation of the Clock Test will be discussed further in a subsequent chapter. The present study will explore the construct validity of the Clock Test in demented and normal adults.

Hypotheses

Structure of the Clock Test. The hypotheses of the present study can be divided into three parts. The first set encompasses the issue of the internal structure of the Clock Test and the effect of age, education, and diagnostic group on the performance of the three Clock Test components. Specifically, this part will address the following questions:

1. Is the Clock Test a reliable measure? I hypothesized that the three components of the Clock Test will demonstrate high reliability coefficients. Specifically, the internal consistency (i.e., Cronbach's alpha) of Clock Setting and Clock Reading will be high for both normal and demented groups of subjects. For Clock Drawing, the inter-correlations between the seven dimensions will correlate highly with the total Clock Drawing score.

2. What are the effects of age, education, and diagnostic group on the performance on the Clock Test? Similar to previous research (Albert & Kaplan, 1980; Borod et al., 1980; Goodglass & Kaplan, 1983) in which adults over 60 performed more poorly on clock drawing and clock setting than younger adults, I hypothesized that performance will decrease across all three tasks as a function of age. In other words, the older the participant, the lower the test scores will be. With respect to education, I hypothesized that performance will increase as a function of education. That is, the more education an individual has, the higher the test scores will be. Similar to previous research (i.e., Albert & Moss, 1984; Anderson et al., 1990; Sunderland et al., 1989; Tuokko et al., 1990a,b; Wolf-Klein et al., 1989), I hypothesized that demented participants' performance will be inferior to that of the normal controls. Finally, an Age by Diagnostic group interaction is hypothesized. Whereas, the performance of the not demented adults will decrease as a function of age, the performance of the demented adults will be similar across age since they performing at a lower

level due to their cognitive status.

3. Do participants with different diagnoses make different types of Clock Drawing errors? It has been argued (Delis, Freeland, Kramer, & Kaplan, 1988; Kaplan, 1989) that total test scores are unsuitable as the basic unit for the analysis of test performance. For example, Delis et al. (1988) demonstrated that brain damaged patients with markedly different cognitive deficits can obtain the same overall score on intellectual and neuropsychological tests. They stressed that global scores are insufficient for characterizing patients' spared and impaired cognitive functions. It is feasible that cognitively impaired individuals receiving the same scores on Clock Drawing may in fact have different diagnoses. For example, a normal adult may have an error score of three representing three misplaced numbers, while a mildly demented adult may receive a score of three which represents two points for distortions and one point for additions. Using total scores, both participants had equal performance when in fact the demented subjects performance was worse. When one examines the global clock drawing score there is no means to determine the types of errors that occur. Subjects with different diagnoses may make different errors which are related to their global test score. I hypothesized that the patterns of errors will differ according to the cognitive status of the individuals.

4. Can one classify individuals into groups based on Clock Test scores? Using the three components of the Clock Test, Tuokko et al. (1986) classified individuals into groups reflecting either normal or impaired cognitive status. Specifically, 95% of the normal and 71% of the impaired participants were correctly assigned to their appropriate group. Furthermore, Tuokko et al. (1992) suggested that cut-off scores of two or more errors indicates cognitive impairment. Using this criterion, individuals were classified into their respective groups. With respect to Clock Drawing, 92% of the normal and 86% of the

demented were classified correctly. For Clock Setting, 87% of the normal and 97% of the demented individuals were classified correctly. Lastly, for Clock Reading, 92% of the normal and 85% of the demented were classified correctly. In addition, when two or more errors on at least two of the three components was imposed as a criterion of impairment, 94% of the normal and 93% of the demented individuals were correctly classified into their respective groups. The present study will attempt to replicate these findings.

Construct Validity. The second set of hypotheses address the construct validity of the three components of the Clock Test. There is no single method or summary index to establish construct validity. Rather, evidence for construct validity can be accumulated through a variety of methods. To determine the construct validity of a test, the entire body of evidence surrounding the particular test must be examined. By evaluating the evidence of a test, a clearer definition of the constructs measured by the test emerges. The method used in the present study, to determine construct validity, involves comparing the Clock Test components to other validated neuropsychological measures, through a series of correlations. This method indicates whether tests measure the same construct and/or what features the tests do, and do not, share in common.

Although constructs are themselves hypothetical abstractions, all constructs are related, directly or indirectly to behaviours or experiences. Since psychological measurement is based upon concrete and observable behaviours, a psychological test can be described as nothing more than a measure of a sample of behaviours. In order to determine whether a test provides a good measure of a specific construct, behaviours that are related to a given construct, as well as those which are unrelated to the construct, must be identified. This system of relationships, referred to as a nomological network, provides a way to systematically describe a particular construct (Cronbach &

Meehl, 1955). The more behaviours (or in this case tests), that are described by a nomological network, the more precision there will be in describing the constructs (Murphy & Davidshofer, 1988). In the present study, I hypothesized that the three Clock Test components are in fact measures of the construct of visuospatial and constructional abilities. As such, the test components should therefore be related to other tests which assess this construct (e.g., WAIS-R Block Design and WAIS-R Object Assembly). The three components, however should be unrelated to other constructs from which they are theoretically different (e.g., Mood and Hand Strength).

The goal of investigating construct validity, therefore, is to determine whether or not test scores provide a good measure of a specific construct. The nomological network provides a definition of the construct, and the relationship between the constructs, in terms of concrete behaviours (which are assessed via test scores). Construct validity, therefore, depends on a detailed description of the relationship between the construct and a number of different behaviours (or tests). The most basic method to assess construct validity is to correlate scores on the test in question with scores on a number of other tests.

Campbell and Fiske (1959; see also Campbell, 1960) stated that, in order to demonstrate construct validity, a test must: (a) correlate highly with other tests with which it is theoretically related, and (b) not correlate significantly with variables from which it differs theoretically. The former is referred to as convergent validity, or the confirmation of a theoretical dimension through independent measurement procedures. The latter is referred to as discriminant validity, and it is important since tests can be discredited if they are too highly correlated with other tests from which they are intended to differ. Because many psychological tests are multifactorial and multidimensional, those tests which are used in order to demonstrate discriminant validity may, in fact, be correlated to some degree with the test of interest. In addition correlations may ensue due to Method Bias (or method variance), which indicates the relationship between various traits measured using the same method. In other words, method variance is the part of the score which is attributable to the common method of measurement of the test battery (Murphy & Davidshofer, 1988). In the present study, this evidence is not obtained because each trait (or construct) is measured by only one test method.

When compared to the convergent tests, the discriminant validity tests should not be as highly correlated with the target task. A trait should correlate higher with another measure of a given construct than with a measure of a different construct. For example, McCann (1991) examined the convergent and discriminant validities of the Minnesota Multiphasic Personality Inventory (MMPI) and the Millon Clinical Multiaxial Inventory (MCMI) in a sample of patients with a personality disorder. They were able to demonstrate, through a series of correlations, that both scales manifested convergent and discriminant validity. In addition, the convergent validity correlations, were in most cases, significantly greater than the discriminant validities. Another example is provided Lachman, Baltes, Nesselroade, and Willis (1982), in which they examined whether context specific measures of personality were more strongly related to older adults' performance on intelligence tests than would be true for more general personality scales. Specifically, the Personality in Intellectual-Aging Contexts (PIC) Inventory was used in order to determine the relationship between personality and intellectual abilities in advanced adulthood and old age. The pattern of correlations were generally in line with the hypothesized pattern of convergent and discriminant validities, with the former correlations in most cases being higher than the latter. A final example is provided by Hertzog, Hultsch, and Dixon (1989). In this study, the convergent validity of the Metamemory in Adulthood instrument (MIA) and the Memory Functioning Questionnaire (MFQ) were examined, through confirmatory factor analysis and structural regression models. Convergent validity was evaluated on the basis of the pattern of relationships among the entire set of metamemory scales. The authors demonstrated that the observed correlations between measures of memory self-efficacy reflected a near-perfect relationship between the underlying latent variables. These findings provide evidence that the measures converge to measure a construct that the authors referred to as Memory Self-Efficacy.

Specifically, the following research questions will be examined:

1. Are the Clock Test components inter-correlated? I hypothesized that the correlations between Clock Setting and Clock Reading will be higher than either's correlation to Clock Drawing. Although all three tasks are similar, Clock Drawing may be more dependent than Clock Setting and Clock Reading on constructional and visuospatial skills (Tuokko et al., 1989).

2. Are the Clock Test components highly correlated to other tests with which they are theoretically related (i.e., convergent validity)? I hypothesized that the Clock Test components will correlate significantly with convergent validity tests. It is expected that the integrity of the correlations will hold even when the analyses are conducted separately for the normal and demented groups.

3. Are the Clock Test components significantly less correlated to tests from which they theoretically differ (i.e., discriminant validity)? I hypothesized that the Clock Test components will be significantly less correlated with discriminant validity tests. Once again it is expected that these hypotheses will be supported regardless of diagnosis.

Confirmatory Factor Analysis. The final set of hypotheses examine the factor structure of the Clock Test relative to the other neuropsychological measures using a maximum likelihood Confirmatory Factor Analysis (Jöreskog & Sörbom, 1989). A detailed description of the hypotheses and model will be presented in a subsequent chapter.

Chapter 3
Method

Participants

The present study used data collected by the University of British Columbia's Clinic for Alzheimer's Disease and Related Disorders. Participants (N = 456) were characterized as belonging to one of the following groups: (a) Unlikely Alzheimer's Dementia (n = 21), (b) Possible Alzheimer's Disease (n = 105), (c) Probable Alzheimer's Disease (n = 154), (d) Not Demented (n = 111), and (e) Normal control (n = 65). Group A (Unlikely AD) consisted of a heterogenous group of patients diagnosed with a dementia other than that of the Alzheimer type (e.g., Multi-Infarct dementia). Participants within group D (Not Demented) were referred to the clinic for memory impairment problems. As such they are treated as a distinct group even though they did not show any significant cognitive impairment that would justify a diagnosis of dementia. Because these participants were referred to the clinic with possible memory problems, they may be of further interest due to the possible susceptibility for further cognitive decline. All diagnostic participants (i.e., all participants except for the normal controls) were referred to the clinic. In the present study, no attempt was made to limit the diagnostic groups to a predetermined size, and all participants who met the minimum requirements of the study were included. Since visual acuity can have an effect on performance, participants with visual defects (e.g., legal blindness), were not included in the present study. Visual defects were determined through reports by the participant and family or through the participant's medical charts.

Diagnostic Groups. Within the patient groups, each patient was referred to the clinic. The patient, and the patient's primary caregiver, were interviewed by a multidisciplinary team consisting of a geriatrician, psychiatrist, neuropsychologist, neurologist, and social worker. All diagnoses were determined according to the criteria of the Diagnostic and Statistical Manual - Third edition revised (DSM-III-R; American Psychiatric Association, 1987) and the National Institute for Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Diseases Association Work Group (NINCDS-ADRDA; McKhann et al., 1984). In addition, on the basis of the information obtained from (a) the source of referral, (b) structured interviews with the patient and a collateral informant, (c) the clinical examinations, and (d) laboratory findings, each team member rated each participant on the eight dimensions of the Functional Rating Scale (FRS; Crockett et al., 1989; Tuokko et al., 1986; Tuokko & Crockett, 1989), which is designed to assess psychosocial functioning in the elderly. The FRS is a multidimensional scale which allows one to evaluate various abilities that are required for successful performance on a number of cognitive and behavioral skills. These skills are affected by the onset of dementia, and are thereby important to assess in order to establish a diagnosis of dementia (Crockett et al., 1989). In addition, the FRS allows one to differentiate between levels of functioning in adults with dementia by comparing their performance on each of the skills. As expected, a one-way ANOVA conducted on the total scores of the four patient groups, indicated that there was a significant difference between the groups ($F(3,387) = 129.84$, $p < .001$). Further analyses (Scheffé post hoc tests, $p < .05$) indicated that the Not Demented group ($M = 16.56$) had a significantly lower total score than the three demented groups (M range = 25.75 - 26.99) which did not significantly differ from one another.

Normal Controls. In addition to the patient groups, 65 normal male and female participants were recruited from senior citizen's activity groups in Greater Vancouver. All of the individuals met the following criteria at the time of testing: (a) they were not receiving institutional care or services, (b) they had no history of a neurological or psychiatric disturbance, and (c) they were over 50 years of age. Initially, each participant was interviewed and asked to complete a medical questionnaire.

Demographic Information. As described in Table 1, when summed across the five diagnostic groups, 189 males (41.4% of the sample) and 267 females (58.6% of the sample) participated in the present study. In each of the groups (except for Unlikely Demented) the proportion of males to females ranged between .48 (Probable AD) and .91 (Possible AD). In the Unlikely AD group, however, the proportion of females to males was reversed, with males outnumbering females by a ratio of 1.62 to 1. Since AD is 2-3 times more prevalent in females than in males (Cummings & Benson, 1983), the greater proportion of females to males for the Probable AD group is consistent with the literature.

The participants ranged in age from 36 to 91 years of age (M age = 70.27 yrs; SD = 9.12 yrs). A one-way Analysis of Variance (ANOVA) indicated that the groups differed in age ($F(4,455) = 10.73$, $p < .001$). Furthermore, Scheffé post hoc tests ($p < .05$) determined that the Not Demented group (M = 65.62 years), while similar in age to the Unlikely AD group (M = 70.43 years), was significantly younger than the normal participants (M = 70.64) and both AD groups (M = 72.37, 71.88 years; Possible AD, Probable AD, respectively). The latter groups did not significantly differ from each other. These differences may be due to the fact that AD generally occurs in older adults, whereas those adults who were not demented may have been demonstrating cognitive loss associated with normal aging. In addition, the Unlikely AD group consisted of participants with a diagnosis of dementia other than that of AD. As such some of its members are younger in age (range: 36 - 85 yrs.). Controlling for the group differences in age will be discussed in a subsequent chapter.

Insert Table 1 about here

According to participants' self-reports, 94.1% percent of them were right handed while the remaining 5.9% were left handed, ambidextrous, or of unknown handedness. Based on self-reports (or reported by the primary care giver), the mother tongue for 84.6% of the participants was English. All participants, however, reported a working knowledge of the English language and were therefore tested in English. Socio-Economic Status (SES) was based on the categorization developed by Weschler (1955). Using this method, participants were classified into groups based on one's own, or one's spouse's past or current occupation. A 10-point SES scale ranging from 1 (Professional Occupations; Highest SES) to 10 (Chronically Unemployed; Lowest SES) was used. An ANOVA on SES indicated group differences, $F(4,434) = 5.71$, $p < .001$. Furthermore, Scheffé post hoc tests ($p < .05$) indicated that the normal adults ($M = 3.02$) had a significantly higher status than both AD groups ($M = 4.99, 4.82$; Possible AD, Probable AD; respectively). The Unlikely AD ($M = 3.60$) and the Not Demented ($M = 4.15$) groups did not significantly differ from one another. Similar findings were observed for years of education ($F(4,449) = 8.80$, $p < .001$), with normal adults ($M = 13.33$ years) reporting the highest number of years completed. This group differed significantly from both AD groups ($M = 10.69, 11.03$ years; Possible AD, Probable AD, respectively) and the Not Demented Group (M = 11.61 years). Unlikely AD (M = 11.95 years) participants did not significantly differ from the other groups. These findings are similar to those observed for SES, since they are based on occupation which is in fact highly correlated with years of education, $r = -.55$, $p < .001$ (the negative correlation refers to a lower SES categorization indicating a higher status).

Although many of the demented participants have been tested on multiple occasions, the present study restricts the data to the participants' first assessment. Although of great interest, the effect of intraindividual change in the performance of adults with a cognitive impairment, is beyond the scope of the present study.

Procedure

The present study received ethical approval by the ethics committee at the University of Victoria. In addition, the Alzheimer's Clinic has received ethical approval for both the collection of data for research purposes, and for the assessment of dementia and related disorders in individuals with a suspected memory disorder. The purpose of the neuropsychological assessment, as well as the fact that the data would be used for research purposes, was explained to both the participant and the participant's primary caregiver. In addition, informed consent statements authorizing the evaluation of one's memory problems and the use of the data for research purposes was signed by the participant or by a relative or caregiver with the proper authority to sign on behalf of the participant. An individual was assessed at the clinic based on the referrals of doctors, or when an individual asked for an assessment because of complaints of memory problems. The neuropsychological assessment took approximately two hours to complete, and breaks were given to the participant, when the clinician felt they were necessary or when the participant indicated that they were tired.

As part of a broad neuropsychological assessment of the participant's cognitive abilities and skills, the participant received the Clock Test (Tuokko et al., 1990a; presented as Appendix C) which includes the following subtests; (a) Clock Drawing, (b) Clock Setting, and (c) Clock Reading. Although there is no time limit, the full Clock Test usually requires no more than 7-10 minutes to complete. Instructions are written at the bottom of each test sheet and are given orally when necessary.

In addition to the Clock Test, a number of intellectual and language based tests were given to all of the participants. These tests are discussed in a subsequent section. Table 2 lists the tasks given in the order in which they were presented to the participants.

Insert Table 2 about here

Measures

The Clock Test. The Clock Test consists of three subtests which are given in the following order: (a) Clock Drawing, (b) Clock Setting, and (c) Clock Reading. Clock Reading and Clock Setting were originally appended as supplemental tasks to Clock Drawing, in order to examine the facility with time concepts at varying levels of complexity. In the present study, the three tests are considered to be of equal importance. There was no time limit for any of the Clock Test components and the time to complete each test was not recorded.

Clock Drawing. In the clock drawing task the participant is given a sheet of paper with a predrawn circle 7 cm in diameter. The words "Draw a Clock" are typewritten at the bottom of the sheet. The participant is told to "imagine that the circle is the face of a clock" and s/he is then told "to place the numbers on the clock face." When the participant has indicated that s/he is finished or when a reasonable time has lapsed (2 - 3 minutes) s/he is then told to "place the hands on the clock to indicate 10 past 11." These instructions are repeated when required. No other instructions are given. When the participant completes the task, or after an appropriate time has lapsed (2 - 3 minutes), the experimenter takes the test sheet and proceeds to the second subtest.

Scoring. The score for Clock Drawing is based on the error score system for the evaluation of clocks (Tuokko et al., 1990a). This system (a) classifies errors by type, (b) determines the frequency of specific types of errors made by the participant, and (c) provides a total error score. Since errors are classified by number and type, one can examine both the quantitative performance by examining the number of errors made, and the qualitative characteristics of the participant's performance by examining the types of errors that occur.

Seven dimensions or specific categories, which are described in detail in Appendix B, are scored:

(1) Omissions. This dimension includes such errors as neglect of part of the clock face, and exclusion of numbers or hands.

(2) Perseverations. This factor includes repetition of numbers, hands, or the progression of numbers following 12 (e.g., 13, 14, 15).

(3) Rotations. This error type includes such flaws as the rotation of the clock face or specific numbers, and such errors as numbers being written backwards or the hands being placed in its mirror image (i.e., 5 minutes to 2 o'clock).

(4) Misplacements. This factor includes numbers or hands being placed

in an incorrect location, and numbers not being written in the correct sequence.

(5) Distortions. These errors arise when one uses the clockface to draw pictures or when the placement of numbers is not guided by the clockface.

(6) Substitutions. These errors occur when numbers are replaced by letters, scribbles, or words, when there is no difference in hand length, and when additional numbers are present.

(7) Additions. This dimension occurs when irrelevant words, scribbles, or figures are present.

Errors are mutually exclusive in that a specific error is not scored on more than one dimension, although a clock may have errors in a number of categories. If an error could potentially be scored in more than one category, precedence is given to the error category which is earliest on the list since the dimensions are presented in a hierarchical and logical fashion for scoring the clocks. For example, if a clock is missing the minute hand, that hand will be scored as missing rather than misplaced, or if a number is omitted, it does not receive an error for a misplaced number. Finally all types of numeric systems (Roman numerals, Arabic numerals, or combinations of the two) are considered correct.

Clock Setting. In this task, five test sheets are individually placed in front of the participant. Drawn on each sheet is a circle 7 cm in diameter with twelve spokes placed along the inner circumference of the clockface to mark the hour units (i.e., to indicate the location of the numbers). No numbers are present on the test sheets. The participant is told to "place the hands of this clock to make it read 1:00". In addition, the requested time is written in words at the bottom of the test sheet (e.g., one o'clock). The five times (in the single order presented) include (a) 1:00 (b) 11:10, (c) 3:00, (d) 9:15, and (e) 7:30. These times were chosen to reflect those which are most often used in clinical practice and for research purposes. In addition, the minute and hour hands are located in each of the four quarters of the clock face, thus controlling for any possible neglect to a part of the clockface. This task usually requires no more than one to two minutes per clock.

Scoring. The scoring for clock setting was adopted from Goodglass and Kaplan (1972) and is obtained by summing the scores for the five items that make up this test. Up to three points are given per clock for a possible score ranging from 0 to 15 for this subtest. One point is given for correctly placing the minute hand, one point for correctly placing the hour hand, and an additional point for correctly indicating the relative lengths of the hour and minute hands. In order to get the additional point: (a) the hour hand must be at least half a centimetre longer than the minute hand, and (b) both the minute and the hour hands must be placed in the correct locations to reflect the desired time. The hands are scores as correct if they fall within half a centimetre to the left or right of the correct position on the clock face.

Clock Reading. In this final subtest, the participant once again receives five sheets of paper, given one at a time. The test sheets are similar to those described for the Clock Setting task, however, in the present task the minute and hour hands are set to a specific time. The participant is asked to verbally state the time at which the two hands are set. The words "What time is it?" are written at the bottom of each test sheet. The five times that the clocks are set to are identical to those given in the Clock Setting task but they are presented in the following single order: (a) 3:00, (b) 7:30, (c) 1:00, (d) 9:15, and (e) 11:10. The same times were used for both Clock Setting and Clock Reading, although the order in which the times were presented were different across the two tasks. Each participant therefore experienced each task with a similar degree of practice since the items were presented in a predetermined order. The effect that practice may have on performance will be discussed further in a subsequent chapter. This task requires no more than one or two minutes per clock.

Scoring. The scoring for the Clock Reading subtest is obtained by summing the scores for the five items. Up to three points are allocated per clock for a possible total of 15 points for this subtest. One point is given for correctly reading the hour hand, one point for correctly reading the minute hand, and an additional point if both are read correctly (e.g., nine fifteen, a quarter past nine, or fifteen minutes after nine o'clock). The third point is included since one may be able to correctly read both the minute and hour hand while incorrectly associating them with respect to the construct of time. For example, one may correctly state the numbers "ten and "eleven." In order to ensure that the participant comprehends the meaning with respect to clock reading, however, s/he must say "ten past eleven" (or some correct representation of the time). If the participant says, for example, eleven ten, then s/he is given credit since this is an acceptable form of telling the time.

Upon completing the Clock Test, the participant continued onto the next part of the neuropsychological assessment battery. This battery will be discussed in more detail in a later section.

Inter-Rater Reliability and Test-Retest Reliability

Upon completion of an assessment, the psychologist who administered the neuropsychological assessment was responsible to score the Clock Test (in addition to the other tests) according to the criteria presented in Appendix B. As such, the tests were scored by a number of different psychologists, all of whom were trained in the administration and evaluation of neuropsychological assessments. In order to determine the inter-rater reliability of the Clock Test, a newly trained individual randomly selected and scored 40 clocks. These clocks were completed by participants from each of the five diagnostic categories. Inter-rater reliabilities between the newly trained individual and the mixed clinician subsample, for the seven clock drawing dimensions scored were the following: (a) Omissions, r = 1.00, (b) Perseverations, r = .96, (c) Rotations, r = .98, (d) Misplacements, r = .93, (e) Distortions, r = 1.00, (f) Substitutions, r = .82, and (g) Additions, r = .73. The inter-rater reliability for the total clock drawing scores was r = .98. Inter-rater reliability scores for each of the five clock setting trials ranged between r = .95 and .98, with a correlation of r = .99 for the total scores. Finally, inter-rater reliabilities for clock reading were consistent at r = .99. This implies that the scoring scheme used can be easily administered by individuals with modest amounts of training.

There is little test-retest reliability information available for the Clock Test. In one pilot study, Tuokko et al. (1992) computed the test-retest reliability for the clock drawing task on a sample of 32 participants who were referred to the clinic for an evaluation of dementia. Between one and four days after completing the initial Clock Test, the participants were asked to complete the same test which was given in the same form, in the same order, and with the same instructions as that of the original testing session. This short interval was chosen in order to minimize the important effects of progressive deterioration in functioning in a demented population. Tuokko et al. obtained a test-retest reliability coefficient of .70, indicating that the scores are moderately stable. Future research will be required in order to more accurately examine the test-retest reliability at various time intervals. Test-retest reliabilities for Clock Setting and Clock Reading have not yet been determined.

Supplementary Tests

The following is a detailed description of the test battery given to all of the subjects that participated in the present study (see Table 2). In addition, such information concerning the issues of reliability, stability, validity, and normative data when applicable are presented. Table 3 describes the theoretical dimensions measured by each of the tasks.

Insert Table 3 about Here

The Weschler Adult Intelligence Scale - Revised (WAIS-R)

The WAIS-R (Wechsler, 1981) is one of the most common measures used in neuropsychological test batteries given to adults between 16 and 74 years of age (Lezak, 1983; Spreen & Strauss, 1991). The WAIS-R supplies information regarding one's overall level of intellectual functioning and the presence or absence of a significant intellectual disability. In addition, it provides clues to altered functions (Lezak, 1983).

The WAIS-R is composed of 11 individually administered composite tests in the form of a test battery. These tests have been classified into Verbal and Performance oriented subtests. The Verbal subtests include the following six tests: (a) Information, (b) Digit Span, (c) Vocabulary, (d) Arithmetic, (e) Comprehension, and (f) Similarities. There are five Performance subtests: (a) Picture Completion, (b) Picture Arrangement, (c) Block Design, (d) Object Assembly, and (e) Digit Symbol.

Factor analytic studies varying across subjects' age (Parker, 1983), populations sampled (Atkinson & Cyr, 1984; Beck, Horwitz, Seidenberg, Parker, & Frank, 1985) and statistical techniques have identified three basic factors: (a) A large Verbal Comprehension factor, (b) a strong Perceptual Organization factor, and (c) a weaker Memory/Freedom from Distractibility factor (Leckliter, Matarazzo, & Silverstein, 1986; Sattler, 1988). The Verbal Comprehension factor is assumed to measure verbal knowledge and comprehension, knowledge obtained partially from formal education, and knowledge reflecting the application of verbal skills to novel situations. Subtests which make up this factor (in rank order from highest to lowest) include Vocabulary, Information, Similarities, and Comprehension (Leckliter et al., 1986). The Perceptual Organization factor is assumed to measure perceptual and organizational dimensions, and reflects the ability to interpret and organize visually perceived material within a time limit. This factor appears to measure a variable common to the Performance subtests and includes (factor loadings in order from highest to lowest) Object Assembly, Block Design, Picture Completion, and Picture Arrangement (Leckliter et al.,1986; Spreen & Strauss, 1991). The third factor, Memory/Freedom from Distractibility is assumed to measure processes related to attention, concentration, and memory. It has also been suggested that this factor reflects executive and short term memory processes involved in the planning, monitoring and evaluation of task performance (Cohen, 1957a, 1957b, cited in Anastasi, 1988). It also examines one's ability to concentrate and to resist distraction. The major subtests of this factor include Digit Span, Arithmetic, and Digit Symbol tasks.

The present study uses seven of the 11 WAIS-R subtests. It is important to note that items within each subtest are arranged in order of increasing difficulty and complexity. The seven subtests are as follows.

(1) Information

The information subtest measures general knowledge that is normally available to persons growing up in the United States. Although the original questions are primarily geared towards an American milieu, a Canadian version of the test has been developed. Psychometric studies of the Canadian version of the information subtest have concluded that there is no evidence that Canadians are penalized by the American items (Crawford & Boer, 1985). Changes have been made, however, in order to increase the face validity of the test for Canadian samples (Violato, 1986).

Subjects are verbally given 29 questions which vary in content as well as in difficulty. Subjects are asked to give a single best answer and if they give more than one, they are told to choose a best answer. In addition, when an answer to a question is unclear or incomplete, the subject is asked to elaborate further and to explain what they mean. They receive one point for every correct answer for a maximum possible score of 29. The items are arranged in increasing difficulty such that the first 4 questions may be answered by all but the highly impaired, while few adults correctly answer the most difficult questions (Lezak, 1983). This subtest is discontinued after 5 consecutive failures.

In addition to measuring general knowledge, the Information subtest examines verbal skills, breadth of knowledge, and remote memory (Lezak, 1983). Information tends to reflect formal education and motivation for academic achievement (Saunders, 1960). In brain injured subjects, Information tends to appear among the least affected subtests and because of its resiliency to change it serves as the best estimate of original ability (Lezak, 1983).

(2) Vocabulary

In the Vocabulary subtest, subjects are verbally presented with 35 words (one word at a time) arranged in order of difficulty. The examiner reads the following question "What does ________ mean?". The examiner places a given vocabulary word in the blank space. If it is difficult to determine whether or not the subject knows the meaning of the word, s/he is asked to further elaborate on what s/he has said. For each vocabulary word, a subject can receive from 0 to 2 points, where 0 refers to an incorrect definition, 1 is an acceptable although incomplete definition, and 2 refers to an accurate and precise definition. A subject can get a maximum of 70 point on this test. The subjects score thereby reflects both the extent of one's recall and the effectiveness of one's speaking vocabulary (Spreen & Strauss, 1991). This task is discontinued if the subject fails five consecutive words. This task normally takes about 15 to 20 minutes to administer, making it the most time consuming subtest.

Within the WAIS-R, the Information and Vocabulary subtests are the best measures of general ability, the test factor that appears to be the statistical counterpart of learning capacity plus mental alertness, speed and efficiency (Lezak, 1983). The Vocabulary subtest has been identified as the single best measure of both verbal and general mental ability, although Information serves equally well as a measure of general ability. One's Vocabulary score is likely to reflect one's SES status and cultural origins, and is less likely than Information to be affected by academic motivation or achievement.

(3) Similarities

The Similarities subtest is a measure of verbal concept formation. The subject is asked to explain what each pair of words has in common (i.e., "In what way are these two words alike?"). The subject is given a total of 14 word pairs that increase in difficulty from the simplest concrete word pair (orange - banana), which only the most impaired adults fail to answer correctly, to the most difficult abstract concept (praise - punishment). The test begins with the first item for all subjects and is discontinued after four failures. A subject can receive a score ranging between 0 and 2, with 0 being indicative of an incorrect response, 1 reflecting a specific concrete response which has a likeness to the correct answer, or 2 which refers to an abstract generalization of the word pair. If the subject gives an unclear or ambiguous answer, s/he is asked to elaborate on his/her response.

Similarities is a good test of general intellectual ability but through middle age it reflects the verbal factor only to a moderate degree. Since similarities is virtually independent of any memory component, it becomes the best test of verbal ability for older people whose memory assumes much more importance on other verbal tests. Of all the verbal tests, similarities is the least affected by the subject's background and experiences. In addition, unlike Information and Arithmetic, Similarities does not depend on academic skills. It is also relatively independent of social, cultural, or educational background.

(4) Digit Span

The Digit Span subtest consists of two distinct and separate tasks, Digits Forward and Digits Backward. Both tests consist of 14 sets of random number sequences that are read aloud by the examiner at an approximate rate of one number per second.

Digits Forward, a measure of attention (i.e., efficiency of attention; Spitz, 1972, cited in Lezak, 1983), is given first. In this subtest, the subject must repeat, in the same order, the number sequences that the experimenter reads.

The subject is given two trials at each level of difficulty for a total of seven levels. When the subject completes two trials at a given length, the experimenter reads the next, longer number sequence. Subjects are given trials ranging from three to nine digits. The normal range for digits forward is 6 +/- 1 digits (Spitz, 1972). This task is discontinued when a subject fails two trials at a given level of difficulty, or completes all items correctly.

Even in cases where a subject receives a score of zero on Digits Forward, they are given the Digits Backward subtest. Digits Backward sequences are two to eight digits long. Upon hearing the sequence, the subject's task is to repeat them in the reversed order. Like Digits Forward, this test continues until the subject fails a pair of sequences or correctly recalls all of the reversals. The Digits Backward requirement of storing numbers while juggling them around mentally is an effortful activity that calls upon working memory. It is therefore more of a memory test than digits forward. The task involves mental double tracking, in that both the memory and the reversing operations must proceed simultaneously (Lezak, 1983).

A subject gets one point for every correct trial, resulting in a maximum possible score of 14 points for each subtest. The Digits Forward and Digits Backward scores are then added in order to get a total Digit Span score. In combining the two digit span tasks to obtain one score, the two tests are treated as if they measure the same, or highly similar behaviours. This is true for most normal adults through middle age: however, with advancing age, Digits Forward span remains rather stable, whereas Digits Backward span typically shrinks (Botwinick & Storandt, 1974, cited in Lezak, 1983). The normal raw score difference between digits forward and digits backward tend to range around 1.0 (Costa, 1975) with a spread of reported differences running as low as .59 (Mueller & Overcast, 1976) and as high as 2.0 (Black & Strub, 1978).

It should be noted that Digit Span scores do not correlate very highly with other measures of cognition. In addition, education appears to have an effect on how well one will perform on the task (Weinberg, Diller, Gerstman, & Schulman, 1972). Overall, this test is sensitive to brain damage and to the kind of diffuse damage that occurs with many types of dementing illnesses (Lezak, 1983).

(5) Block Design

The Block Design subtest is a construction test in which the subject is presented with either four or nine, red and white blocks. It is the best measure of visuospatial conceptualization and organization in the WAIS-R. It reflects general ability to a moderate degree so that intellectually capable, but academically or culturally limited persons, frequently obtain their highest score on this task (Lezak, 1983).

In this test, the subject is required to use blocks to construct a replica of a design that is presented to the subject by the experimenter. All of the subjects begin with the first item, which is presented and demonstrated as a block copying rather than a design copying test. A subject can repeat the first two items if s/he fails to produce a correct design. Subjects are scored for both accuracy and speed. On the first two items, the subject gets two points if s/he successfully completes it within the time allotted. If the subject fails to complete the item on the first attempt, s/he can get one point if s/he correctly completes the design on a second attempt. On subsequent designs, a subject can get up to four bonus points (depending on the design) for completing the puzzle quickly. There is a total of nine designs to be copied (five of which contain four blocks, four of which contain nine blocks). Subjects are given one minute to complete the four block designs and two minutes for the others. The test is discontinued after three successive failures.

Individuals with Alzheimer's disease are likely to perform poorly on this test (Lezak, 1983). In the early stages of the disease, individuals will understand the task and may be able to copy one or two of the designs. However, they soon get so confused between one block and another or between their construction and the examiner's model that they are unable to imitate the placement of more than one or two blocks.

(6) Object Assembly

In the Object Assembly subtest, the subject is required to assemble cut-up cardboard puzzles of familiar objects. There are a total of four items which are given in order of increasing difficulty. All of the items are administered to every subject. The objects (in the order given) include: (a) Manakin, (b) Human Profile, (c) Hand, and (d) Elephant. All responses are scored for both time and accuracy. Although each item has a time limit (two minutes for the first two items, three minutes for the others), partially complete responses receive credit as well. Twelve of the 41 possible points on this task are rewarded for performance speed.

Object assembly has the lowest association with general intellectual ability of all of the performance scale subtests and is second only to digit span in weakness on this factor. In normal adults, performance on this subtest tends to vary relatively independently of performance on other subtests. Like block design, it is a relatively pure measure of visuospatial organization ability, as well as a measure of constructional ability. Rather than abstract thinking, the ability to form visual concepts quickly, and to translate them into rapid hand responses is essential for a high score. This test is therefore as much a test of speed of visual organization and motor response as it is of the capacity for visual organization itself.

Object assembly and block design correlate more highly with one another than with any other subtest (since both measure constructional ability). This reflects their similarity in requiring the subject to synthesize a construction from discrete parts, and probably reflects the speed component as well.

(7) Digit Symbol

The Digit Symbol test is a test of psychomotor performance that is relatively unaffected by intellectual competence, memory, or learning (Erber, Botwinick, & Storandt, 1981). The task consists of four rows, each containing 25 small blank squares paired with a randomly assigned number from one to nine. Above these rows is a printed key that pairs each number to a different nonsense symbol. Following a practice trial on the first seven squares, the subject's task is to fill in the blank spaces, as quickly as possible, with the symbol that is paired to the number. The score is the number of squares correctly filled in during 90 seconds. The importance that motor speed plays in the scoring creates validity problems for subjects whose motor responses tend to be slow. It is also particularly difficult for elderly subjects whose vision or visuomotor coordination is impaired (Savage, Britton, Bolton, & Hall, 1973, cited in Lezak, 1983).

Motor persistence, sustained attention, response speed, and visuomotor coordination play important roles in the normal person's performance. Although some perceptual organization components do show up in the performance of older adults, the natural response slowing that comes with age seems to be the most important variable contributing to the age differences in this test. This subtest is more sensitive to brain damage than any other subtest (Lezak, 1983). This is not surprising since it can be affected by so many factors that are related to performance.

The Controlled Oral Word Association Test

This test, also known as the Word Fluency Test and the FAS Test, was developed by Benton and Hamsher (1978) as part of the Multilingual Aphasia Examination. It is a sensitive measure of verbal associative fluency (Spreen & Strauss, 1991) which is often impaired in patients with diffuse or focal brain disease. Factor analytic studies (e.g., DesRosiers & Kavanagh, 1987) have showed that this test loads mainly on a Verbal Knowledge factor along with tests of vocabulary. The subject is given three word naming trials and is asked to produce, in a limited period of time, as many words as possible that begin with the letters F,A, and S. They are timed via a stopwatch and are given one minute for each letter. A subjects raw score is the sum of all acceptable words for all three letters. Inadmissible words (i.e., proper nouns, repetitions, variations, wrong words, and numbers) are not counted as correct.

In general, the raw score is modified for the influence of educational level and age. Three points are applied to subjects over 55 years of age. In addition, nine additional points are given to adults with an education level below 8 years, four points are given to adults with an education level between 8 and 11 years, and 0 points given to adults with an education level above 11 years. Benton and Hamsher (1981) maintained that when the scores are corrected for age and education, the test shows little decline up to 80 years of age, and can therefore be used as a good measure of cognitive functioning while controlling for age and education effects. Therefore, a low score would not be attributed to differences in age or education, rather a different reasons would be explored. In the present study, raw uncorrected Word Fluency scores were used.

One year test-retest reliability coefficients of .70 have been reported in normal elderly adults, and a test-retest reliability of .88 after a period of 19 - 42 days (desRosiers & Kavanagh, 1987). Correlations with age ranges between

.14 (Mittenberg, Seidenberg, O'Leary, DiGiulio, 1989) and -.19 (Yeudall, Fromm, Reddon, & Stefanyk, 1986). Yeudall et al. (1986) showed correlations of .21 with education, .14 with the verbal IQ score of the WAIS, and .29 with the performance IQ score of the WAIS. Overall, this test is a sensitive measure of individual word production under a restricted search condition.

The Finger Tapping Test

The Finger Tapping Test, which is also known as the Finger Oscillation Test (Halstead, 1947), is the most widely used test of manual dexterity (Lezak, 1983). It is a measures of motor speed of the index finger of each hand (Spreen & Strauss, 1991). Using a specially adapted tapper, the subject is instructed to tap as fast as possible first with the index finger of the preferred hand, and then with the non-preferred hand. The subject is told to move only the index finger while keeping the whole hand and arm at rest, and is provided with a practice trial. Five 10 second trials are given for each hand, with a brief rest period (i.e., between 20 and 30 seconds) given after each trial. The finger tapping score is computed separately for each hand and is the mean of the five consecutive trials within a range of five taps. A maximum of 10 trials with each hand is allowed and if the above criterion is not met, the score is the mean of the best five trials.

Finger tapping measures are often included in neuropsychological examinations to assess subtle motor impairments. Typically the performance of the preferred and nonpreferred hands are compared and the usual guideline is that the preferred hand should perform about 10% better than the nonpreferred hand (Reitan & Wolfson, 1985, in Spreen & Strauss, 1991).

Performance with each hand is quite stable over time, even with test-retest intervals of two years (Dodrill & Troupin, 1975; Spreen & Strauss, 1991).

Reliability coefficients ranging from .58 to .93 have been reported with both normal and neurologically impaired adults. Slowing occurs with age, tending to show up in the fifth to sixth decades and increasing significantly thereafter (Bak & Greene, 1980). In general, better performance is associated with the preferred hand, male gender, younger age, and more years of education (Bornstein, 1985).

Hand Dynamometer

The Hand Dynamometer, also referred to as the Grip Strength Test, measures the strength or intensity of the voluntary movements of each hand (Spreen & Strauss, 1991). The subject is required to hold the upper part of the dynamometer in the palm of the hand and squeeze the stirrup with the fingers as tightly as possible, taking as long as necessary in order to get maximum intensity. The subject's arm must be placed down at the side and away from the body. Subjects are given one practice trial and two recorded trials with each hand, alternating between preferred and nonpreferred hands. Ten second pauses are given between each trial in order to avoid fatigue. A third trial is provided if the subject's hand slips. The amount registered at each trial is recorded (in kilograms) and the mean is calculated separately for each hand. This task takes approximately five minutes to administer.

Grip strength measures are included in neuropsychological test batteries in order to assess subtle motor impairments and to permit inferences about the cerebral speech pattern and the functional integrity of the two cerebral hemispheres (Spreen & Strauss, 1991). The measure has proven useful in discriminating epileptic patients with left hemisphere speech from those with right hemisphere speech (Strauss & Wada, 1988), and in differentiating brain damaged from normal people (Dodrill, 1978). In general, performance with the preferred hand is superior to that of the nonpreferred hand (Bornstein, 1985, 1986).

Performance tends to be better with the preferred hand, and males tend to be stronger than females (Spreen & Strauss, 1991). There is also a positive correlation between grip strength and both height and weight, as well as between grip strength and education. Strength is also negatively related to age, although the magnitude of intermanual differences are not affected by age or level of education (Bornstein, 1986).

The Multi-Focus Assessment Scale (MAS)

The Multi-Focus Assessment Scale (MAS; Koch, Coval, Crockett, & Anderson, 1984) was designed "for the evaluation of cognitive, mood, and language functioning in the frail elderly" (Coval, Crockett, Holliday, & Koch, 1985, p. 2). The MAS consists of eight performance and rating scales that are scored on the basis of an interview with the subject. The time required for the administration of the MAS is between 15 and 25 minutes. The subscale scores are not designed to be summated to yield a composite score or a global measure of psychological functioning, but are intended to help the clinician generate a profile based on the strengths and weaknesses of a deteriorating elderly person. Test-retest reliability after a period of four months ranges between .67 and .92 with only one subtest falling below .72. Inter-rater reliabilities are quite high, not falling below .92 for any of the subscales. The scales of the MAS include the following.

(1) Social Behaviour Skills.

This is an adapted version of the Minimal Social Behaviour Scale

(Lawton, 1971). This scale assesses one's basic behavioral skills for sustaining social interactions. Occurrences of appropriate social skills are checked off as they occur throughout the interview. This subscale consists of 11 items which are rated by the interviewer.

(2) Receptive Language Skills.

These are defined as the ability to follow simple commands accurately. These commands are presented in both visual and auditory modalities. The first set of performance tasks consists of eight verbally administered requests to perform simple tasks, and the second consists of nine visually presented stimuli to perform specific actions. All but the most debilitated subjects can perform some of the items.

(3) Expressive Language Skills

These examine the appropriateness and understandability of the patients speech during the assessment. The test consists of three ratings carried out by the interviewer following the completion of the assessment.

(4) Cognitive Status

This is assessed through two subscales which are embedded within a set of 28 items. The first is a Canadian version of the Mental Status Questionnaire (Kahn, Goldfarb, Pollack, & Peck, 1960) which contains ten items that examine orientation to time, place, and person. The second, entitled Orientation to Extended Care Unit (ECU) consists of 18 items measuring the subject's awareness to the immediate environment. All items are scored as correct or incorrect and are given in sequential order.

(5) Mood

This subscale consists of 14 of the 24 original items from the Memorial University of Newfoundland Scale of Happiness (MUNSH; Kozma & Stones, 1980). Items are presented verbally to the subject and scored as either yes, no, don't know, or no response. The scale verifies the presence or absence of general demoralization (i.e., psychological well being among the elderly). The test was shortened for the sake of brevity with four positive and negative affects, and three each of the positive and negative general experience items. Test-retest reliability has been reported to be between .70 and .79 with an internal consistency coefficient of .85 (Kozma & Stones, 1980).

(6) Accessibility and Sensory Abilities

This three-item subscale examines the capacity to attend to conversations. It includes the interviewer's rating of the patient's interference from sensory deficits, distractibility, and mental alertness.

The Present Functioning Questionnaire (PFQ)

The PFQ (Crockett et al., 1989) consists of 60 items which reflect five potential problem areas: (a) personality, (b) everyday functioning, (c) language skills, (d) memory, and (e) self-care. Individual items of the PFQ are scored on a presence or absence basis. The PFQ is given to an informant, typically a spouse or child of the subject, in the form of a structured interview.

Luria Nebraska Item 227

The present study uses a modified version of Item 227 of the Luria Nebraska Test Battery (Golden, Purisch, & Hammeke, 1976). In the Design Recall part, a picture of five geometric shapes (i.e., square, circle, triangle, cross, and diamond) is presented. The subject is asked to examine the pictures for a period of 7 seconds and is then required to draw as much as s/he can recall. In the Design Reproduction task, the subject is presented with the picture of the shapes and is asked to make a copy of each figure. The Design Reproduction task is not timed. After a 5 minute delay (during which the subject is engaged in other verbal and motor tasks) the subject is given a Delayed Recall task in which s/he must draw the shapes from memory. Across the three tasks, the number of shapes correctly reproduced is scored, for a maximum possible score of five points per task. This test involves both visual memory and visuospatial skills (King, Caine, Conwell, & Cox, 1991) and according to Golden et al. (1976) both recall trials load on a memory factor. In addition, normal controls performed significantly better than brain injured patients.

Buschke Cued Recall

The Buschke Cued Recall test of memory (Buschke, 1984) allows for the manipulation of both the initial encoding of material, and its retrieval through the use of specific cues. A modified version of the test was used in the present study (Tuokko & Crockett, 1989; Tuokko, Vernon-Wilkinson, Weir, & Beattie, 1991). Subjects are presented with a random array of 12 pictures of common items belonging to different semantic categories. The subject is asked to search for and name each of the target items when the category label for that item is presented verbally by the examiner (e.g., tool, fruit) resulting in repeated searches of the entire array. The successful identification of the items demonstrates that the subject has processed the stimuli, can understand and make use of category labels to name the items, and allows the experimenter to use the category labels for effective cued recall tests. All but the most debilitated or severely aphasic subjects can correctly search and identify the items by category label.

A distractor task (i.e., counting backwards from 100 for 60 seconds) is employed before the first free recall trial begins. For those items that were not recalled, the experimenter verbally presents the category labels one at a time, in order to elicit verbal cued recall. Cued recall is only tested for those items not retrieved by free recall since items retrieved by free recall should also be retrieved by cued recall (Buschke, 1984). This procedure is a kind of selected reminding task (Buschke, 1973) in which the subject is reminded only of items that were not recalled, so that learning can be shown by recall without further presentation of the original stimuli.

If any item has not been recalled during either free or cued recall, the entire array of pictures is once again presented and the subject is asked to locate and name only those items that were missed. However, if all of the items were retrieved there is no further presentation and the subject proceeds to the next trial of free followed by cued recall. A delayed recall trial is administered after a fifteen minute delay (during which the subject is engaged in non-memory tasks). The delayed recall trial has been described as a remote memory trial (Tuokko & Crockett, 1989).

Each trial results in two scores: (a) The number of items correctly identified through free recall, and (b) The number of items correctly recalled through both free and cued recall (i.e., total recall). Total recall provides a minimal estimate of the number of items encoded and available for recall

(Buschke, 1984). Although there is no time limit, this task takes approximately 15 minutes to complete.

Free recall learning has been shown to be severely impaired in demented adults. Total recall, however, is similar for both groups since mildly to moderately demented adults are able to retrieve essentially all of the remaining items by cued recall (Buschke, 1984; Tuokko & Crockett, 1989).

Chapter 4
Results

The aim of the present study was to examine the construct validity of the Clock Test as a diagnostic tool for the assessment of dementia. The construct validity of a test is the extent to which a test measures a theoretical construct or trait (Anastasi, 1988; Cronbach, 1990). In the first part of this chapter, the structure of the Clock Test was examined. Specifically, the issue of reliability was investigated for the three components of the Clock Test. In addition, the effects of age and diagnostic group as predictors of Clock Test performance were assessed. The seven dimensions that constitute Clock Drawing were specifically examined with respect to diagnostic group. The second part of the chapter examined the relationship of the clock test with other validated neuropsychological measures. In the final part of this chapter, a confirmatory factor analysis was used to examine the factor structure of the Clock Test relative to the other neuropsychological measures.

The goal of this study was to clarify what the clock test actually measures (i.e., visuospatial abilities vs. abstract conceptualization of time), which factors or dimensions are related to it, and whether the clock test, used as an assessment tool of dementia, can add unique information to the neuropsychological test battery. Unless otherwise stated, all statistical procedures were computed using the Statistical Packages for the Social Sciences (SPSSX, 1988).

Restricted Age

Since the five groups (i.e., UAD, Possible AD, Probable AD, ND, and Normal) significantly differ with respect to age, age and diagnostic group are confounded. The analyses were therefore conducted on an age-restricted sample. By restricting age to the point to which age is not significantly different between diagnostic groups, it becomes possible to use age as an independent variable without the problems of a confound. Specifically, the data from the 25 of the youngest participants (1.75 SD from the mean) and 13 of the oldest participants (1.5 SD from the mean) were discarded. The difference in the number of participants discarded from each extreme reflects the desire to discard as few subjects as possible. This resulted in non-significant differences between the groups with respect to age, ($F(4,413) = 2.28$, $p > .05$). This criterion reduced the sample to 418 participants ranging in age from 54 to 84 years of age ($M = 71.20$ years, $SD = 6.97$) compared to the original sample size of 456 participants. Analyses on the demographic variables using the restricted age sample were consistent with those presented in the method section.

The Structure of the Clock Test

Reliability

Overall across test items, average scores ranged from 1.53 to 1.80 ($M = 1.64$; $N = 414$) for Clock Setting, and from 1.69 to 2.42 ($M = 2.05$; $N = 409$) for Clock Reading. The maximum possible score per item was three points. Cronbach's alpha (Cronbach, 1951) was computed in order to examine the internal consistency of the Clock Setting and Clock Reading components. The alphas were computed across the five items that comprise each of the subtests. The internal consistency was high for both Clock Setting ($\alpha = .95$) and for Clock Reading ($\alpha = .86$) thereby demonstrating reliability within the two tasks. As further evidence of reliability, the intercorrelations between test items ranged from r = .75 to r = .84 for Clock Setting, and from r = .43 to r = .63 for Clock Reading.

The reliability was also examined separately for: (a) the Normal and Not Demented groups together, and (b) the three demented groups together. The categorization of participants into Relatively Normal (i.e., cognitively intact) versus Demented (i.e., cognitively impaired) was used in order to reflect cognitive ability, rather than diagnostic groups. This was therefore conducted to examine whether the reliabilities change with cognitive status. For ease, these groups will be referred to as Relatively Normal versus Demented for the remainder of the study.

With respect to Clock Setting, average item scores ranged from 2.46 to 2.62 (M = 2.57; N = 151) for relatively normal participants, and from .94 to 1.33 (M = 1.10; N = 263) for the demented participants. In addition, intercorrelations between test items ranged from r = .37 to r = .66 for the relatively normal participants, and from r = .68 to r = .78 for the demented participants. With respect to Clock Reading, average item scores ranged from 2.53 to 2.93 (M = 2.76; N = 151) for the normal participants, and from 1.67 to 1.85 (M = 1.75; N = 258) for the demented participants. In addition, intercorrelations between test items ranged from r = .14 to r = .34 for normal participants, and from r = .36 to r = .63 for the demented participants. The internal consistency for Clock Setting was high (α = .84, .93; normal and demented participants, respectively). The internal consistency for Clock Reading, however, was low for the normal participants (α = .57), whereas the converse was true for the demented participants (α = .83). Further analyses were conducted in order to determine why the reliability of Clock Reading was low for the relatively normal participants. Item to total scale score correlations were computed. These correlations ranged between r = .75 and r = .83 overall, and between r = .74 and r = .80 for the demented participants. These correlations are reflective of the high internal consistency when Cronbach's alpha was computed. However, item to total correlations for the normal participants ranged between r = .40 and r = .75, with the lowest correlation between the first item and the total score. However, even after removing the first item, Cronbach's alpha was still only equal to .55 for the normal participants. The low reliability for Clock Reading can be explained in part by the fact that normal adults performed close to the ceiling. This will be discussed further in a subsequent section.

Cronbach's alpha was not performed for the Clock Drawing task for the following reasons. First, because Clock Drawing is hierarchical in nature, scores on former dimensions are linked to, and thereby constrained by, the latter dimensions. This violates the requirement that part scores be experimentally independent (Cronbach, 1990). Second, some of the dimensions, have no predefined maximum score. Intercorrelations between Clock Drawing dimensions ranged between r = -.08 and r = .29. Further analyses revealed that the intercorrelations ranged between r = -.09 and r = .16 for the normal participants, and between r = -.17 and r = .30 for the demented participants. Three sets of correlations were computed and included: (a) item to total correlations between the Clock Drawing variables and the dimension to which it is associated, (b) item to total correlations between the Clock Drawing variables and the total Clock Drawing score, and (c) correlations between each dimension and the total Clock Drawing score. These correlations, which are presented in Table 4, revealed that most of the variables which make up each dimension correlate highly with the total score for its associated dimension. In addition, except for Distortion errors, the dimensions correlated significantly with the total Clock Drawing score. When correlations were examined by cognitive status similar findings were obtained with the exception that both Distortion and Substitution errors were not significantly correlated with the total score for the Normal participants. These dimensions, however, had relatively few errors associated with them (M errors = .04, .23; Distortion and Substitution errors, respectively), thereby explaining to some degree the low correlations.

Insert Table 4 about here

The Effect of Age, Education and Diagnostic Group

First, a blockwise hierarchical multivariate multiple regression was conducted in order to determine whether a multivariate effect was present between the three dependent variables (Clock Drawing, Setting, and Reading) and the independent variables (Diagnostic Group, Age, Education, as well as their interactions). This analysis is similar to a multivariate analysis of variance (MANOVA) in which one examines the influence of multiple predictors by conducting an omnibus test of all of the criterion variables at once, before going on to test the criterion variables individually (Cliff, 1987). This analysis revealed a significant effect (Wilks' $F(36,1155.98) = 9.15$, $p < .001$; $\eta^2 = .52$) indicating that these variables act as multivariate predictors of Clock Test performance. Subsequent blockwise analyses were conducted separately on each of the dependent variables.

The following analyses were conducted in order to examine the prediction of age, education, and diagnostic group on performance, within each component of the Clock Test. Age is a continuous variable, ranging from 54 to 84 years ($M = 71.20$, $SD = 6.97$ years). Education is also a continuous variable, ranging from a reported 0 to 22 years ($M = 11.38$, $SD = 3.19$). Diagnostic group, however, is a categorical variable, in which members are dummy coded according to group membership. The codings themselves are nothing more than symbols used to categorize individuals with identical codings as equals on a particular variable (Pedhazur, 1982). In addition, these variables can be used to differentiate participants based on group membership. The following dummy codes were used in order to differentiate the diagnostic groups: (a) 1 represented individuals with a diagnosis of dementia other than that of AD (Unlikely AD), (b) 2 represented individuals diagnosed with Possible AD, (c) 3 represented individuals diagnosed with Probable AD, (d) 4 represented individuals diagnosed as Not Demented, and (e) 5 represented Normal controls. Codings within a regression model are required in order to examine categorical variables (Cohen & Cohen, 1983; Ferguson, 1981; Pedhazur, 1982).

Multiple regression analyses in which a dependent variable is regressed on coded vectors (that represent a categorical variable with more than two categories) yield results that are identical to those obtained from the application of an analysis of variance (see Pedhazur, 1982, for a description of this technique). In order to examine specific planned comparisons, a set of k - 1 coded variables were generated, where k equals the number of diagnostic groups. Planned comparisons (also referred to as a priori comparisons) refer to comparisons between groups that have been planned in advance, and are based on the theory guiding the study. The coded variables (or contrasts) represent the planned comparisons which were examined within the regression analyses. The result is a set of four coded variables each corresponding to one degree of freedom. When these variables are combined (or entered together in a regression as a block), all of the information of a given construct is maintained. In other words, a variable (i.e., diagnostic group) is partitioned into k - 1 additive parts (Ferguson, 1981). The planned comparisons, which are presented in Table 5, were constructed to examine the following hypotheses:

(1) The Normal and Not Demented adults' performance on the dependent variables would not significantly differ from one another (see Contrast 1).

(2) The three demented groups would not differ from one another. This was conducted by first comparing the Possible and Probable AD groups to each other (see Contrast 2), followed by comparing their combined performance to that of the Unlikely AD participants (see Contrast 3).

(3) The Normal and Not Demented participants' combined performance would significantly differ from the three demented groups, with the former groups outperforming the latter (see Contrast 4). It was expected that the hypothesized differences between groups would be similar across the three Clock Test components.

Insert Table 5 about here

Separate blockwise hierarchical multiple regression analyses were performed on the total scores of each component of the Clock Test in order to examine the effect of age, education, and diagnostic group and their subsequent interactions. The interactions, or product variables, were computed by multiplying the three possible pairs of predictor variables. The product of two variables carries the desired interaction information (Cohen & Cohen, 1983). In the case where one variable is a dummy coded contrast, the interaction term is computed by multiplying the continuous variable by the dummy coded contrast. The resulting term represents the amount of variation of the continuous variable within the group defined by the categorical variable. For example, in order to represent the Age x Diagnostic Group Interaction, one interaction term was computed for each group, resulting in a total of four interactions. Each interaction represented the variation of age within the particular Diagnostic Group. When all four interaction terms were entered together as a block, all of the information for the Age x Diagnostic Group Interaction was represented (Pedhazur, 1982). The three interaction terms (Age x Education, Age x Diagnostic Group, Education x Diagnostic Group) were therefore entered together as a separate block, for a total of 9 interaction terms. Within regression, interaction terms are included after the main effects since two variables are said to interact in their accounting for variance when they have a joint effect over and above any additive combinations of their separate effects (Cohen & Cohen, 1983).

Analyses. The analyses consisted of a series of hierarchical regression models in which the predictors, or independent variables (Diagnostic Group, Age, Education, as well as their interactions) were entered into the equation in a priori specified orders. The following regression models were used for each component of the Clock Test.

(1) The first block of predictors consisted of years of education, age of participant, and diagnostic group (i.e., the four sets of contrasts were entered together). The second block included the interactions between all of the previous variables. This model determined the proportion of variance predicted by the independent variables prior to subsequent analyses in which education is used as a covariate. In other words, it is a global regression in which all of the predictors are entered together followed by all of the interactions.

(2) Education was entered first, as a covariate, followed by Age and the contrasts which reflect Diagnostic Group as the second block. The Age x Diagnostic Group interactions were entered last. Education was entered as a covariate in order to remove any effect that it may contribute to the criterion variable. Although it has been shown that performance increases as a function

of education (Goodglass & Kaplan, 1983, Farver, 1975), the variable was used as a covariate since the five diagnostic groups differed significantly with respect to years of education. Specifically, reliable conclusions based on education are confounded with the fact that the diagnostic groups significantly differ in years of education. For example, participants within the normal groups have significantly more years of education than the Probable AD participants. If the effect of education was significant, it becomes difficult to determine whether performance increases as a function of years of education, or whether it is due to the difference in years of education between diagnostic groups.

(3) In order to determine whether the Diagnostic Group contrasts are significant, a separate regression analysis was conducted for each of the contrasts. In this model, both the effects of age and education were entered first as covariates, followed by the contrast of interest. The analysis of contrasts within regression is analogous to an ANOVA. A significance test which allows one to determine whether the groups being compared within a particular contrast differs significantly from one another was conducted for each of the contrasts. This F test equals the mean squared regression (MS_{reg}) when a single contrast is entered into the equation divided by the mean squared residual (MS_{res}) from the regression analysis when all of the contrasts are entered into the equation. MS_{res} is equivalent to the Mean Squared Within (MS_{w}) used in the ANOVA model, and equals the error term to which each contrast will be compared. The MS_{reg}, on the other hand, is equivalent to the Mean Squared Between (MS_{b}), or the variance attributable to groups. This analysis is therefore conducted in order to get the MS_{reg} for each of the contrasts which will then be used with the MS_{res} from the previous analysis to determine whether the contrasts are significant.

Clock Drawing

In the first analysis, the first block of predictors (or main effects), consisting of years of education, age of participant, and diagnostic group (i.e., the four sets of contrasts entered together) were significant ($R^2 = .23$, $F(6,409) = 20.82$, $p < .001$). The unstandardized regression coefficients (or partial regression coefficients (beta weights); see Cohen & Cohen, 1983, p.83) revealed that neither age ($t(409) = 1.12$, $p > .05$) nor education ($t(409) = 1.4$, $p > .05$) was a significant predictor of Clock Drawing performance. The results for the contrasts which represented diagnostic group will be described in the following analysis. The second block consisting of the interactions between all of the previous variables, did not significantly predict variance in Clock Drawing performance above that predicted by the main effects (R^2 change = .02, $F(9,400) = 1.34$, $p > .05$). In fact, the beta weights revealed that none of the interaction terms were significant predictors of performance.

In the second analysis education was entered in the first block, followed by age and the four contrasts representing diagnostic group in the second block. The Age x Diagnostic Group interactions were entered in the third block. Education was not a significant predictor of Clock Drawing performance, accounting for less than 1% of the variance ($R^2 = .00$, $F(1,414) = .71$, $p > .05$). Age and Diagnostic Group however, accounted for 23.2% of the variance ($R^2 = .23$, $F(5,409) = 24.8$, $p < .001$). Contrary to the hypothesis that an age effect would be observed, the beta weights (see Table 6) revealed that age was not a significant predictor of performance ($t(409) = 1.12$, $p > .05$). The correlations between age and the diagnostic comparisons ranged between $r = -.08$ to $r = .13$.

Insert Table 6 about here

As previously described, separate regression analyses were conducted on each of the contrasts in order to determine whether they are significant. The MS_{reg} was then divided by the MS_{res} from the previous analysis ($MS_{res} = 36.56$). The model consisted of age and education entered as the first block followed by the contrast of interest in the second block.

The first contrast compared the normal participants to the Not Demented individuals. This contrast was not significant ($MS_{reg} = 83.81$, $F(1,414) = 2.29$, $p > .05$) indicating that the two groups did not differ from one another with respect to Clock Drawing performance ($M = 2.67$, 1.18 errors; Not demented and Normals, respectively). The second contrast compared the Possible to Probable Alzheimer's disease groups. This contrast was not significant ($MS_{reg} = 84.84$, $F(1,414) = 2.32$, $p > .05$), indicating that the participants did not significantly differ from one other ($M = 9.16$, 8.69 errors; Possible and Probable AD, respectively). The third contrast compared the two previous groups (Possible and Probable AD) to the Unlikely AD group. This analysis was significant ($MS_{reg} = 496.27$, $F(1,414) = 13.57$, $p < .01$), indicating that individuals with a diagnosis other than that of AD ($M = 8.21$ errors) outperformed those with a diagnosis of Possible or Probable AD ($M = 8.93$ errors). In the final comparison, Normal and Not Demented participants were compared to participants with any diagnosis of dementia. This analysis was significant ($MS_{reg} = 1475.55$, $F(1,414) = 40.36$, $p < .01$) indicating that relatively normal participants ($M = 2.07$ errors) had fewer errors than demented participants ($M = 8.84$ errors).

The Age x Diagnostic group interactions were entered into the equation as the third block. The Age x Diagnostic group interactions did not predict a significant proportion of the variance. They accounted for less than 1% of the variance ($R^2_{change} = .003$, $F(4,405) = .35$, $p > .05$). It was hypothesized, however, that an age effect would be observed for those participants who are relatively normal (i.e., Normal and Not Demented), with performance declining with age. For individuals who are demented, it was hypothesized that the effect of age would not be observed. This would result from the demented participant's poor (i.e., at floor level) performance at any age. Therefore, age would not account for any additional variance over that which was accounted for by the diagnosis. In other words, the regression of each component on age would vary if the participants were normal or demented.

In order to examine this hypothesis, two simple main effect regression analyses were conducted. Simple main effects, which are more sensitive to differences within groups than are interactions, can be defined as the analysis of a given variable at one level of a second variable (Pedhazur, 1982). This analysis examined the effect of age separately for the normal and demented participants. The model consisted of education (as a covariate) as the first block followed by age. In the first regression analysis, the Normal and Not Demented groups were collapsed (N = 150). Education was not a significant predictor of Clock Drawing performance, accounting for less than 1% of the variance ($R^2 = .004$, $F(1.148) = .60$, $p > .05$). The participants' age accounted for 3.3% of the variance ($R^2_{change} = .033$, $F(1,147) = 5.01$, $p < .05$).

The three groups representing participants with a diagnosis of dementia were also collapsed (N = 266). Education was not significant ($R^2 = .007$, $F(1,264) = .16$, $p > .05$). In addition, age did not predict performance, accounting for less than .05% of the variance ($R^2_{change} = .00$, $F(1,263) = .14$, $p > .05$). These findings, which can be seen in Figure 1, support the hypothesis that the effect of age varies according to cognitive status.

Insert Figure 1 about here

Summary. These analyses demonstrated that the best predictor of Clock Drawing performance was a individual's diagnostic group. Furthermore, the regression analyses revealed that the Normal and Not Demented participants did not differ from each other on Clock Drawing performance. Whereas the Clock Drawing performance of individuals diagnosed with Possible or Probable AD was inferior to that of the Unlikely AD group, their performance did not differ from one another. Lastly, the relatively normal groups (i.e., Normal and Not Demented) outperformed the demented groups on Clock Drawing. Whereas age was a significant predictor of performance for the relatively normal groups, age was not a significant predictor of Clock Drawing performance for the demented groups.

Clock Setting

In the first analysis, the first block of predictors consisted of years of education, age of participant, and diagnostic group. These main effects were significant, accounting for 47.1% of the variance ($R^2 = .471$, $F(6,405) = 60.22$, $p < .001$). The beta weights revealed that neither age ($t(405) = -1.41$, $p > .05$) nor education ($t(405) = 1.50$, $p < .05$) were significant predictors of Clock setting performance. The results for the contrasts which represented the diagnostics groups will be described in a subsequent analysis. The Age x Diagnostic Group interactions, which were entered as the second block, did not predict Clock Setting performance (R^2 change= .02, $F(9,396) = 1.60$, $p > .05$). In fact, none of the interaction terms were significant predictors of performance.

In the second analysis, education was entered (as a covariate) in the first block. The main effects of age, and the four contrasts representing diagnostic groups were entered in the second block, followed by the Age x Diagnostic Group interactions which were entered in the third block. Unlike Clock Drawing, education was a significant predictor of Clock Setting performance, accounting for approximately 4.2% of the variance ($R^2 = .04$, $F(1,410) = 17.83$, $p < .001$). Since the groups differed significantly with respect to number of years of education, one can not make reliable conclusions based on this finding. The main effects for Age and Diagnostic Group accounted for 43.0% of the variance ($R^2 = .430$, $F(5,405) = 65.88$, $p < .001$). When the unstandardized regression coefficients were examined (see Table 7), age was not a significant predictor of performance ($t(405) = -1.41$, $p > .05$). The correlations between age and diagnostic comparison ranged between $r = -.07$ and $r = .13$.

Insert Table 7 about here

Separate regression analyses were conducted for each of the contrasts in order to determine whether the contrast was significant. In this regression, both age and education were entered first. The MS_{res} to which the MS_{reg} was compared was equal to 15.1.

The first contrast compared the Normal participants to those who were diagnosed as Not Demented. This contrast was significant ($MS_{reg} = 226.92$, $F(1,410) = 15.02$, $p < .01$) indicating that the Normal participants ($M = 14.00$) performed significantly better on Clock Setting than the Not Demented participants ($M = 12.11$). The second contrast compared those participants who were diagnosed with Possible AD to those diagnosed with Probable AD.

Possible AD participants ($\underline{M}$ = 6.01) significantly outperformed Probable AD participants ($\underline{M}$ = 4.91; MS_{reg} = 292.15, $\underline{F}(1,410) = 19.34$, $\underline{p} < .01$), indicating that these groups differ from each other with respect to Clock Setting performance. The third contrast compared the two previous groups (Possible and Probable AD) to the Unlikely AD group. This contrast was significant (MS_{reg} = 825.11, $\underline{F}(1,410) = 54.63$, $\underline{p} < .01$), indicating that those individuals with a diagnosis other than that of AD ($\underline{M}$ = 7.83) outperformed those with a diagnosis of AD ($\underline{M}$ = 5.46). In the final comparison Normal and Not Demented participants were compared to participants with a diagnosis of dementia. This analysis was significant (MS_{reg} = 1737.57, $\underline{F}(1,410) = 115.04$, $\underline{p} < .01$) indicating that the relatively normal participants ($\underline{M}$ = 12.87) had fewer errors than individuals with a diagnosis of dementia ($\underline{M}$ = 5.53).

The Age x Diagnostic Group interactions were entered on the last block of the equation. These interactions did not predict a significant proportion of the variance. They accounted for less than 1% of the variance ($R^2_{change} = .003$, $\underline{F}(4,401) = .63$, $\underline{p} > .05$). As previously described, it was hypothesized that Age x Diagnostic group interactions would be observed. Whereas Clock Setting performance was expected to decrease as a function of age for the normal participants, no age effect was expected for the Demented participants. In other words, age was expected to be a significant predictor of Clock Setting performance for those participants who are relatively normal (i.e., Normal and Not Demented), with adults performing worse as they get older. However, for adults who are demented, it was hypothesized that age would not predict performance; rather, it was expected that the participants would perform at floor levels due to the dementia.

As previously described, two simple main effect regression analyses were conducted. The model consisted of education as the first block, followed by age. The Normal and Not Demented groups (N = 150) were collapsed and a regression analysis was performed. Education was a significant predictor of Clock Setting performance, accounting for 2.9% of the variance ($R^2 = .029$, $F(1,148) = 4.48$, $p < .05$). Since the groups significantly differed in number of years of education, one can not make any reliable conclusions based on this finding. The participants' age, which was entered in the second block, accounted for an additional 4.7% of the variance ($R^2_{change} = .047$, $F(1,147) = 7.46$, $p < .01$).

The three demented groups (N = 262) were also collapsed. Education was not significant ($R^2 = .011$, $F(1,260) = 2.93$, $p > .05$). In addition, age did not predict performance, accounting for less than .01% of the variance ($R^2_{change} = .000$, $F(1,159) = .00$, $p > .05$). The hypothesis that the predictive power of age would vary according to cognitive status was therefore confirmed, and can be seen in Figure 2.

Insert Figure 2 about here

Summary. These regression analyses demonstrated that one's diagnostic grouping was the best predictor of Clock Setting performance. Furthermore, the analyses indicated that the Normal participants outperformed the Not Demented participants, and the Possible AD participants outperformed the Probable AD participants. In addition, the Unlikely AD participants outperformed the Possible and Probable AD participants. Lastly, the relatively normal participants outperformed the demented participants. Whereas age was a significant predictor of Clock Setting performance for the relatively normal groups, age was not a significant predictor of Clock Setting performance for the demented groups.

Clock Reading

In the first analysis, years of education, age of participant, and diagnostic group were entered first, followed by their subsequent interactions. The main effects were significant, accounting for 30.7% of the variance ($R^2 = .307$, $F(6,400) = 29.50$, $p < .001$). The beta weights revealed that neither age ($t(400) = -.22$, $p > .05$) nor education ($t(400) = .60$, $p > .05$) were significant predictors of Clock Reading performance. The results for the diagnostic group contrasts will be described in the following analysis. The second block consisted of the interactions between all of the previous variables. The interactions did not predict Clock Setting performance (R^2 change = .03, $F(9,391) = 1.77$, $p > .05$). In fact, the beta weights revealed that none of the interactions were significant predictors of performance.

In the second analysis, education was entered (as a covariate) in the first block. The main effects of age, and the four contrasts representing diagnostic groups were entered in the second block, followed by the Age x Diagnostic Group interactions which were entered in the third block. Similar to Clock Setting, education was a significant predictor of Clock Reading performance. It accounted for approximately 2.0% of the variance ($R^2 = .02$, $F(1,405) = 8.29$, $p < .01$). Age and diagnostic group were significant predictors of Clock Reading performance, accounting for 28.7% of the variance ($R^2 = .287$, $F(5,400) = 33.08$, $p < .001$. However, when the unstandardized regression coefficients were examined (see Table 8), age was not a significant predictor of performance ($t(400) = -.22$, $p > .05$). The correlation between age and diagnostic comparisons ranged between $r = -.08$ and $r = .14$.

Insert Table 8 about here

As previously described, separate regression analyses were conducted for each of the contrasts. In this regression, both the effects of age and education were entered first. The MS_{res} to which the MS_{reg} was compared was equal to 17.76.

The first contrast compared the Normal participants to those who were Not Demented. This was significant (MS_{reg} = 94.10, $F(1,405) = 5.30$, $p < .05$) indicating that the Normal participants (M = 14.39) performed significantly better on Clock Reading than the Not Demented participants (M = 13.38). The second contrast revealed that Possible AD participants (M = 9.23) significantly outperformed Probable AD participants (M = 7.38; MS_{reg} = 194.17, $F(1,405)$ = 10.92, $p < .01$). This indicates that the two groups differed from each other with respect to Clock Reading performance. The third contrast compared the two previous groups (Possible and Probable AD) to the Unlikely AD group. This analysis was also significant (MS_{reg} = 371.36, $F(1,405) = 20.91$, $p < .01$), indicating that those individuals with a diagnosis other than that of AD (M = 9.11) outperformed those with a diagnosis of AD (M = 8.31). In the final comparison, Normal and Not Demented participants were compared to participants with a diagnosis of dementia. This analysis was significant (MS_{reg} = 973.85, $F(1,405) = 54.82$, $p < .01$) indicating that the relatively normal participants (M = 13.79) had fewer errors than the demented participants (M = 8.22).

The Age x Diagnostic Group interactions did not predict a significant proportion of the variance. They accounted for less than 1% of the variance (R^2_{change} = .008, $F(4,396)$ = .1.14, $p > .05$). As previously described, it was hypothesized that the regression of age on Clock Reading would vary according to cognitive status. Two simple main effect regression analyses were conducted. The models consisted of education entered as the first block, followed by age in the second.

The Normal and Not Demented groups (N = 150) were collapsed and a regression analysis was performed. Education was not a significant predictor of Clock Setting performance, accounting for less than 1% of the variance (R^2 = .009, $F(148) = 1.49$, $p > .05$). The participants' age was entered in the second block, accounting for 3.2% of the variance ($R^2_{change} = .032$, $F(1,147) = 5.01$, $p < .05$).

A second regression analysis, using the model described above, was also conducted on the three demented groups (N = 257) which were collapsed. Education was not significant ($R^2 = .003$, $F(1,255) = .66$, $p > .05$). In addition, age did not predict performance, accounting for less than 1% of the variance ($R^2_{change} = .002$, $F(1,254) = .49$, $p > .05$). The hypothesis that the effect of age varies according to group was therefore once again confirmed, and can be seen in Figure 3.

Insert Figure 3 about here

Summary. These analyses revealed similar findings to those observed for Clock Setting. Specifically, one's diagnostic group was the best predictor of Clock Reading performance. Furthermore, the regression analyses revealed that the Normal participants' performance was superior to that of the Not Demented participants, and that the Possible AD participants outperformed the Probable AD participants. In addition, the Unlikely AD participants outperformed the Possible and Probable AD participants. Lastly, the relatively normal participants outperformed the demented participants. Whereas age was a significant predictor of Clock Reading performance for the relatively normal participants, age was not a significant predictor of Clock Reading performance for the demented participants.

The Effect of Education. In order to examine the effect of education on the Clock Test components, two further analyses were conducted. A restricted education range was implemented such that no significant differences were observed between the groups on this variable (range = 7 - 17 years, M = 11.38 years, *SD* = 3.19, N = 384). The five groups were therefore similar on both age and number of years of formal education.

In the first analysis the main effects of age and years of education were used in order to predict performance on each of the Clock Test Components. Neither age nor education was a significant predictor of Clock Drawing (R^2 = .01, $F(2,381) = 1.67$, $p > .05$) or Clock Reading (R^2 = .01, $F(2,373) = 1.48$, $p > .05$. These findings, however, were not replicated for Clock Setting. Age and education were significant predictors of Clock Setting performance (R^2 = .02, $F(2,378) = 4.11$, $p < .05$. However, when the unstandardized regression coefficients were examined, age was significant, $t(378) = -2.21$, $p < .05$, whereas education approached significance $t(378) = 1.80$, $p = .07$.

In the second analysis, the first block of predictors consisted of the main effects of age, years of education, and diagnostic group, followed by their interactions in the second block. Across Clock Test components, upon examination of the unstandardized regression coefficients, neither age nor education were significant. The results were consistent with those previously described.

Summary. These analyses revealed that unlike previous research, the present study can not support the hypothesis that education is a significant predictor of Clock Drawing and Clock Reading. Although not significant, education approached significance for Clock Setting.

Group Differences in the Dimensions of Clock Drawing

In this set of analyses, I examined whether the number and type of Clock Drawing errors varied according to diagnostic group. A Multivariate Analysis of Variance (MANOVA) comparing the performance of the five diagnostic groups on the seven Clock Drawing dimensions was performed. A significant diagnostic group effect, Wilks' $F(28,1468.88) = 4.97$, $p < .001$, $\eta^2 = .53$, was observed. Follow-up ANOVAs (and Scheffé post hoc tests) were performed separately for each of the seven dimensions. In addition, the analyses were repeated using for each of the seven dimension through a series of Analysis of Covariance (ANCOVAs). These ANCOVAs were performed in order to observe whether group differences emerged when age and education were used as covariates (i.e., the effect of age and education were controlled). An alpha level of 5% (i.e., $p < .05$), as a criterion for significance, was required for all post hoc tests.

An ANOVA on the first dimension, number of Omissions, resulted in a significant effect, $F(4,413) = 15.30$, $p < .001$. Follow-up analyses indicated that Normal ($M = .08$ errors) and Not Demented participants ($M = .49$ errors) had significantly fewer errors of this type than both the Possible and Probable AD groups ($M = 3.44$, 4.08 errors, respectively). In addition, the Normal participants had significantly fewer errors than the Unlikely AD group ($M = 3.74$ errors). The Normal and Not Demented groups did not differ from one another. Significant group differences were also observed when age and education were used as covariates, $F(4,409) = 15.22$, $p < .001$. Possible AD and Probable AD participants (adjusted $M = 3.51$, 4.17 errors, respectively) made significantly more errors than both the Not Demented and Normal Participants (adjusted M = .45, -.01 errors, respectively). The Unlikely AD participants did not significantly differ from any of the other groups (adjusted $M = 3.72$). Table 9 presents the means for each diagnostic group with and without controlling for the effect of age and education.

Insert Table 9 about here

Significant findings were observed for the number of Perseveration errors, $F(4,413) = 17.87$, $p < .001$. Follow-up analyses indicated that Normal ($M = .03$ errors) and Not Demented ($M = .12$ errors) participants had significantly fewer errors than participants with a diagnosis of Possible AD (M = 1.04 errors). The Unlikely AD participants ($M = .32$ errors) and Probable AD participants ($M = .86$ errors) did not significantly differ from any of the other diagnostic groups. When age and education were used as covariates, significant group differences were still observed, $F(4,409) = 5.23$, $p < .001$. Post hoc tests revealed that Normal participants (adjusted $M = .00$ errors) made significantly fewer errors than the Possible AD participants (adjusted $M = 1.05$ errors) on Perseveration errors. The Probable AD (adjusted $M = .88$ errors), Unlikely AD (adjusted $M = .30$ errors), and Not Demented (adjusted $M = .15$ errors) groups did not significantly differ from one another or from any of the other groups (See Table 9).

Significant group differences were observed for the third dimension, Rotation errors, $F(4,413) = 4.83$, $p < .001$. Follow-up analyses indicated that

Normal (M = .15 errors) and Not Demented (M = .39 errors) had significantly fewer errors than participants with a diagnosis of Possible AD (M = 1.26 errors). The Unlikely AD participants (M = 1.42 errors) and Probable AD participants (M = .62 errors) did not significantly differ from any of the other diagnostic groups. When age and education were used as covariates, significant group differences were observed, $F(4,409) = 4.01$, $p < .01$, however no significant differences were revealed when Scheffé post hoc tests were conducted (see Table 9). This may be due to the fact that the Scheffé test is more rigorous than other multiple comparison methods with regard to Type I errors (i.e., asserting that a difference exists when no such difference exists), leading to fewer significant differences (Ferguson, 1981). The pattern of errors, however, was similar to those described when covariates were not used.

An ANOVA on the fourth dimension, number of Misplacement errors, revealed significant differences between the groups, $F(4,413) = 4.67$, $p = .001$. Normal participants (M = .62 errors) significantly outperformed the Possible AD participants (M = 2.30 errors). The Not Demented (M = 1.14 errors), Unlikely AD (M = 1.74 errors), and Probable AD groups (M = 1.86 errors) did not significantly differ from any of the other groups. When age and education were used as covariates significant group differences were observed, $F(4,409) = 5.22$, $p < .001$. Follow-up analyses revealed that both the Possible and Probable AD participants (adjusted M = 2.35, 1.94 errors, respectively) made significantly more errors than the Normal participants (adjusted M = .48 errors). The Unlikely AD (adjusted M = 1.67 errors) and the Not Demented (adjusted M = 1.22 errors) participants did not significantly differ from any of the other groups (see Table 9).

The fifth dimension, number of Distortion errors revealed a significant group effect, $F(4,413) = 4.67$, $p = .001$. Follow-up analyses indicated that the Normal participants (M = .00 errors) significantly outperformed the Possible AD participants (M = .14 errors). The Not Demented (M = .07 errors), Unlikely AD (M = .00 errors), and Probable AD (M = .16 errors) did not significantly differ from any of the other groups. When age and education were used as covariates, group differences emerged, $F(4,409) = 3.78$, $p < .01$. Scheffé post hoc tests, however did not reveal any significant differences between the groups with respect to this type of error. The pattern of errors, however, was similar to those described when covariates were not used (see Table 9).

An ANOVA on the sixth dimension, number of Substitutions errors, also revealed significant group differences, ($F(4,413) = 2.84$, $p < .05$). Although significant group differences were obtained, Scheffé post hoc tests did not reveal any significant differences between the groups. Even when age and education were used as covariates, significance was achieved ($F(4,409) = 2.92$, $p < .05$). The post hoc tests, however, did not reveal any significant differences between the groups. The pattern of errors, however, was similar to the observed pattern when covariates were not used (see Table 9).

Lastly, an ANOVA on the number of Addition errors revealed significant group differences, $F(4,413) = 3.19$, $p < .05$). Once again Scheffé post hoc tests did not reveal any significant differences between the groups. When age and education were controlled for, significance was achieved ($F(4,409) = 3.00$, $p < .05$). Once again, post hoc tests, did not reveal any significant differences between the groups. The pattern of errors, however, was similar to the observed pattern when covariates were not used (see Table 9).

The previous analyses examined whether the number and types of Clock Drawing errors varied according to diagnostic group. These analyses were conducted in order to determine whether individuals with differential diagnoses make different types of errors. In the next set of analyses, I examined the issue of group differences in Clock Drawing errors when: (a) the Normal and Not Demented groups were collapsed, and (b) the three demented groups were collapsed. These analyses were conducted in order to examine differences between cognitively intact versus cognitively impaired individuals on the Clock Drawing dimensions.

A MANOVA comparing the performance of the two groups on the seven Clock Drawing dimensions was performed. A significant diagnostic group effect, Wilks' $F(7,410) = 18.23$, $p < .001$, $\eta^2 = .87$), was observed. Follow-up ANOVAs were performed separately for each of the seven dimensions. The normal participants consistently had significantly fewer errors than the demented participants. These findings were reproduced even when one's age and level of education were controlled for. The F values and means are presented in Table 10.

Insert Table 10 about here

Summary. These results support the hypothesis that the number and type of Clock Drawing errors varied by diagnostic group. The performance of the Normal and Not Demented participants and between the Possible and Probable AD participants were similar across each of the Clock Drawing dimensions. Whereas, the Possible AD participants' performance was significantly worse than that of the Not Demented participants for the first three dimensions (Omission, Perseveration, and Rotation errors), when compared to the Normal group, their performance was worse for the first five dimensions (Omission, Perseveration, Rotation, Misplacement, and Distortion errors). With respect to Distortion and Substitution errors, although not significant, there was a trend for the performance of the Not Demented and Normal participants to be superior to that of the three demented groups. Finally, when the demented groups were collapsed and the not demented groups were collapsed, the Demented participants consistently performed worse than the Not Demented groups.

Classification in Clock Drawing Dimensions

In the next set of analyses the following question was addressed: Can one correctly classify an individual into his/her respective diagnostic group based on Clock Drawing scores? The Clock Drawing scoring scheme is based on scoring the seven dimensions in a hierarchical fashion. Therefore this set of analyses examined whether all seven dimensions are required to correctly classify individuals into their respective diagnostic groups, or whether a subset of these dimensions would suffice. In other words, what is the optimum number of dimensions needed to classify individuals into their respective diagnostic groups? These questions were examined by conducting a Hierarchical Discriminant Function Analysis (DFA). This method evaluates the contribution to the prediction of group membership by predictors as they enter an equation in a pre-determined order (Tabachnick & Fidell, 1989).

When a discriminant function analysis was conducted using all five diagnostic groups across all seven dimensions of Clock Drawing, only 41.63% of the participants were correctly classified into their respective groups. In order to examine the differential effect of a predictor to discriminate between the groups, a series of DFA's were performed. Each predictor was added into the DFA, one at a time, until all seven dimensions were entered into the analysis. When each predictor was entered into the equation in a serial fashion, the classification results changed. These results are presented in Table 11.

Insert Table 11 about here

When scoring Clock Drawings, Omission errors are examined first. Omission errors (Lambda = .871) resulted in an averaged correct classification of 34.9% of the participants. With the addition of Perseveration errors to the DFA, the correct classification rose to 41.63%. In order to evaluate the improvement in classification as a new predictor is added to the analysis, a test of the differences between two lambdas was conducted (See Appendix D). This procedure involved comparing the lambda from the step with the larger number of predictors, to the lambda from the step with fewer predictors (Tabachnick & Fidell, 1989; see p. 546-547). Lambda (Λ) is an estimate of error variance. This analysis revealed that with the addition of Perseverations, classification improved significantly over that of Omission errors alone (Λ = .830; $F(8,832) = 2.54$ $p < .05$).

When Rotation errors were included in the DFA with Omissions and Perseverations, the correct classification increased significantly by .48%, to 42.11%, (Λ = .792; $F(12,1098.278) = 2.18$, $p < .05$). With the addition of Misplacement errors to the equation, the correct classification increased significantly to 42.34% (Λ = .758; $F(16,1265.429) = 1.75$, $p < .05$). When Distortion errors were included into the equation, the correct classification decreased by 1.67% to 40.67%. This decrease was not significant (Λ = .730; $F(20,1370.716) = 1.27$, $p > .05$). The addition of Substitution errors to the equation decreased the correct classification by an additional .24% to 40.43%. This decrease was not significant (Λ = .725; $F(24,1438.506) = -.21$, $p > .05$). With the addition of the final predictor, Additions, the correct classification of Clock Drawings into the five diagnostic groups increased to 41.63%. This increase was not significant (Λ = .721; $F(28,1483.304)$ = .14, p > .05). The DFA did not classify any individuals into either the Unlikely AD group, or into the Normal control group at any step.

These analyses suggest that Omission, Perseveration, and Rotation errors are the best predictors for the classification of individuals into groups. One should not expect however, to categorize with certainty, since only 41.63% of the individuals were correctly classified into their respective groups.

The previous analyses were repeated in order to determine whether the classification of individuals would increase if cognitive status (i.e., Demented vs. Normal) was used rather than actual diagnostic group. The three demented groups were collapsed, as were the Normal and Not Demented groups. These results are presented in Table 12.

Insert Table 12 about here

The first predictor, Omissions (Λ = .875), resulted in a correct classification of 63.88% of the participants. With the addition of Perseveration errors (Λ = .839) to the equation, the overall classification of the participants into the groups remained at 63.88%. With the addition of Rotation errors to the equation, correct classifications increased by 8.61% to 72.49%. This increase, however, was not significant (Λ = .815; $F(3,415)$ = 1.99, p > .05. Misplacement errors increased the correct classification to 76.32%. This increase, however, was not significant (Λ = .785; $F(4,414)$ = 1.98, p > .05. The addition of Distortion errors increased the correct classification marginally to 76.79%, (Λ = .768; $F(5,413)$ = .88, p > .05), while the addition of Substitution errors had no effect on the overall classification of errors (Λ = .763; $F(6,412)$ = .22, p > .05).

Finally, Addition errors resulted in a decrease in correct classification to 75.36%, although this decrease was not significant ($\Lambda = .762$; $F(7,411) = .03$, $p < .05$).

Summary. From these analyses one can conclude that in order to classify individuals into groups that represent normal versus cognitively impaired, one needs to only examine errors of Omission. However, classification was correct only 63.9% of the time, reflecting the inadequacy of simply using this scale to classify groups. Using the remaining Clock Drawing dimensions to classify individuals increased the overall percentage of correct classification to 75.36%, however this increase was not significant at the 5 percent alpha level.

Classification in Clock Test Components

The next set of analyses examined the degree to which each Clock Test component could classify individuals into groups reflecting either demented or relatively normal cognitive status. As previously described, the three demented groups were collapsed, as were the Normal and Not Demented groups. Separate Linear Discriminant Function Analyses (DFA) were conducted for each of the Clock Test components.

The results for Clock Drawing were described in the previous section. To review, using the seven dimensions of Clock Drawing resulted in an overall correct classification rate of 75.4%. Specifically, 75.7% of the demented, and 74.8% of the relatively normal individuals were classified correctly. For Clock Setting, 69.7% of the demented and 94% of the relatively normal were classified correctly. The total correct classification was 78.5%. For Clock Reading, 59.7% of the demented, and 94.7% of the relatively normal individuals were classified correctly into their respective groups. A total correct classification of 69.1% was observed.

A cut-off score of two or more errors (as described by Tuokko et al., 1992) was used in order to signify a cognitive impairment. A cut-off of greater than two errors on Clock Drawing yielded a total correct classification rate of 78%. Specifically, 90.3% of the demented, and 56.3% of the relatively normal individuals were classified correctly. A cut-off score of less than 13 correct responses for both Clock Setting and Clock Reading produced a total correct classification rate of 79% and 77.8%, respectively. For Clock Setting, 95.5% of the demented were classified correctly, whereas only 51% of the relatively normal individuals were classified correctly. For Clock Reading, 82.4% of the demented and 69.5% of the relatively normal individuals were classified correctly.

Using a criteria of impairment of 2 or more Clock Test components yielded an overall correct classification rate of 81%. Specifically, 93.3% of the demented and 59.6% of the relatively normal individuals were classified correctly. With respect to Clock Drawing, it was observed in a previous section that only omission errors were required in order to classify individuals into groups reflecting cognitive status. The following analysis used a criterion of impairment of two or more components of the Clock Test, where an impairment on Clock Drawing was defined as two or more errors of omission. This analysis produced an overall correct classification rate of 81%. Specifically, 83.5% of the demented and 76.8% of the relatively normal individuals were classified correctly. The last analysis used a criteria of impairment of all three components of the Clock Test. This analysis produced an overall correct classification rate of 79.7%. Specifically, 77.2% of the demented and 84.1% of the relatively normal individuals were classified correctly.

Summary. These analyses demonstrated that whereas between 54.7% and 75.7% of the demented individuals can be classified by the Clock Test components, these test components can correctly classify between 74.8% and 94.7% of the relatively normal individuals. When cut-off scores of two or more errors was used in order to signify an impairment, between 82.4% and 95.5% of the demented, but only between 51% and 69.5% of the relatively normal individuals were classified correctly. Finally, when an impairment on two or more Clock Test components was used to signify impairment, the highest overall correct classification rate was observed.

Part II: The Relationship of the Clock Test to other Neuropsychological Tests

In the following analyses, the components of the Clock Test were compared to a number of neuropsychological tests which were administered to the participants. These tests, which are presented in Table 2, were described in detail in the previous chapter.

Missing Data

There were a number of instances for which data were missing. Certain data were missing due to experimental error in scoring, the participants were unable or unwilling to complete a given test, or the experimenter could not give the full test battery because of a lack of time. In addition, the test battery has been shortened such that the Object Assembly subtest of the WAIS-R, Vocabulary subtest of the WAIS-R, and the Hand Dynamometer test were not given to all of the participants.

There are a number of procedures available to handle the problem of missing data. First, one can simply drop all cases with missing data. Since missing values are scattered throughout cases and variables, this procedure was not considered as a viable option. A list-wise deletion of all participants with missing data would result in a substantial loss of data with a resultant sample size of only 269 participants (N = 115, 154; Normal, Demented participants, respectively). Second, missing variables can be estimated during data analysis using mean values. Means are calculated from the available data and are then used to replace the missing values. In the absence of any other information, the mean is the best estimate about the value of a variable (Tabachnick & Fidell, 1989). This method is conservative since the mean of a distribution does not change and the variance of the variable is reduced. The correlation a variable will have with other variables (because of the reduction in variance) however, would be affected. The extent of loss in variance depends on the amount of missing data. A less conservative approach is to insert the group mean as a missing value rather than the global mean. The final option is to use other available data in order to estimate the missing values. Cases with complete data generate a regression equation which is then used to predict missing values for incomplete cases. The advantage of this method is that it is more objective than simply using the mean.

The BMDP (Dixon, 1988) statistical software package was used for data estimation. In addition to estimating missing data, this program describes the pattern of missing data (i.e., whether the missing values are missing at random, where the values are located, and whether cases with missing values are outliers). Table 13 reports the percentages of missing data for each variable used in the estimation. Any variable for which 15% or more of its data were missing was neither estimated nor used to estimate other variables. The following variables met this criterion and were therefore not estimated: (a) Hand Dynamometer, for which 30.1% of the participants had missing data, (b) WAIS-R Vocabulary, for which 53.6% of the participants had missing data, and (c) WAIS-R Object Assembly, for which 54.1% of the participants had missing data. All other variables used in the estimation procedure had less than 15% of the total cases missing. In addition, since the following sections deal with the relationship between the Clock Test and other neuropsychological tests, the Clock Test was not used in the estimation of missing data. Finally, the data estimation procedure was conducted twice using different methods. The first method was based on estimating missing data by regressing all of the available data for all of the participants. The second method consisted of estimating missing data separately for the normal (i.e., Normal and Not Demented) and demented participants (Unlikely AD, Possible AD, and Probable AD) using the same regression technique as the one used by the first method.

Insert Table 13 about here

Both methods resulted in almost identical estimations for the missing data. These data were compared and contrasted to the original data set (which contained missing data). For each variable, the difference between the means for the original data set and both estimation procedures were computed. The means for the estimation procedures were similar to each other and to the means of the original non-estimated data set, ranging between -.076 and .467 units from the mean. In addition, the standard deviations ranged between -.70 and .18 units from those of the original data set. Finally, each estimated data point was examined in order to determine whether (a) the estimated data points fit within the parameters of possible scores for a given test, and (b) the estimated data were consistent with scores achieved by other members of the group. The participants with and without missing data were compared in order to determine whether these participants differed with respect to education, age, gender, SES, language, and marital status. Significant differences were not observed between the groups for any of the variables thereby lending support to the notion that the pattern of missing data was random.

The Relationship between Clock Test Components

In the first analysis, the relationship between the Clock Test components was examined. A MANOVA was conducted in which the dependent variable was the three Clock Test components. The independent variable was group (relatively normal vs. demented). The MANOVA revealed significant differences between the three tasks across the two groups (Wilks' $F(3,414) = 53.73$, $p < .001$, $\eta^2 = .28$). Follow-up univariate ANOVAs revealed that Normal participants ($M = 2.07$ errors, *SD* = 2.86, range = 0-15) had significantly fewer Clock Drawing errors than the Demented participants ($M = 8.84$ errors, *SD* = 7.26, range = 0-31; $F(1,416) = 120.51$, $p < .001$). In addition, Normal participants ($M = 12.87$, *SD* = 2.7, range = 0-15; $M = 13.79$, *SD* = 2.18, range = 2-15, Clock Setting and Reading, respectively) significantly outperformed Demented participants ($M = 5.53$, *SD* = 4.55, range = 0-15; $M = 8.22$, *SD* = 5.11, range = 0-15) on both Clock Setting and Clock Reading ($F(1,412) = 327.47$, $p < .001$; $F(1,407) = 162.13$, $p < .001$).

When the inter-correlations of the three Clock Test Components were computed, each component of the Clock Test correlated significantly with one another. Specifically, the following correlations were observed: (a) Clock Drawing and Clock Setting, $r = -.64$, (b) Clock Drawing and Clock Reading, $r = -.60$, and (c) Clock Setting and Clock Reading, $r = .79$. All correlations were significant at $p < .001$. The negative associated with Clock Drawing reflects the fact that higher scores (i.e., more errors) reflect greater impairment in this task, whereas the opposite is true for the remaining tasks. The significance of the difference between the correlations was computed. This analysis revealed that when the correlation between Clock Setting and Clock Reading was compared to Clock Setting and Clock Drawing ($t(407) = 4.47$, $p < .001$) and to Clock Reading and Clock Drawing ($t(407) = 5.4$, $p < .001$), the inter-correlation between Clock Setting and Clock Reading was significantly higher than either task's correlation with Clock Drawing. The previous analyses were conducted without concern for the direction of the effect. That is to say, whereas the correlation with Clock Drawing involves a negative, the t-tests were computed without regard to the direction. This high correlation reflects an underlying similarity that exists between these two tasks, whereas Clock Drawing, although still highly correlated may be tapping into an additional ability, in addition to that of Clock Setting and Clock Reading.

When the inter-correlations were computed separately for the normal and demented participants, the following results were observed: (a) Clock Drawing and Clock Setting, $r = -.35, -.52$ (Normal and Demented, respectively), (b) Clock Drawing and Clock Reading, $r = -.17$ ($p < .05$), $-.50$ (Normal and Demented, respectively), and (c) Clock Setting and Clock Reading, $r = .62, .71$ (Normal and Demented, respectively). Unless otherwise noted, all correlations were significant at $p < .001$. When the significance differences between the correlations were computed, the results for both the normal and demented groups mirrored those described above. These results, however, indicate that for the demented participants, the tasks are correlated to a greater degree than they are for the normal participants, possibly due to ceiling performance on the part of the normal participants. The ceiling effect that was observed for the normal participants will be described in a later section.

Discriminant and Convergent Validity

The correlations between the three Clock Test components and several other neuropsychological tests were computed. The tests were divided into eight discriminant validity tests and 20 convergent validity tests. As previously described, tests that demonstrate convergent validity should correlate highly with other tests with which it is theoretically related. Conversely, tests that demonstrate discriminant validity should not correlate with variables from which they theoretically differ. Although missing data were estimated as described in the previous section, four test scores (i.e., Dynamometer Left and Right, WAIS-R Vocabulary, and WAIS-R Object Assembly) still contained substantial missing data. As noted earlier, they surpassed the criterion of missing data that was established in order for estimation to occur (i.e., a maximum of 15% missing). These variables, however were included in the following analyses with the caveat that the sample size associated with these variables is less than that of the remaining variables. They were included so as to not lose potentially important information, even if this information is only based on a subsample.

The correlations were computed separately for both the Normal and Demented participants, as well as for the combined groups as a whole. In addition, two significance tests were conducted. The first significance test involved the significance of the difference between dependent correlations through the use of McNemar's (1969; see also Steiger, 1980) formula for testing the significance between correlated correlations. The formula, which is presented in Appendix D, takes into account the fact that the correlations which are being compared are based on the same sample, and are therefore not independent. The convergent validity coefficients were compared to the discriminant validity coefficients in order to test the hypothesis that the convergent correlations would be significantly greater than the discriminant correlations. Since the direction of the significance was hypothesized a priori, unless otherwise noted, a one-tailed test was used at an alpha level of $p < .05$ (Ferguson, 1981; McNemar, 1955). The second significance test examined the significance of the difference between two correlation coefficients when the correlations were obtained from independent samples or groups. The purpose of this analysis was to determine whether the correlations between a pair of tasks significantly differed for the Normal and Demented groups. The test (see Appendix D) involved a Fisher r-to-z transformation for each pair of observed correlations (Steiger, 1980). The correlations were transformed into z-scores and were then tested via a t-distribution (see Steiger, 1980). A two-tailed alpha level of $p < .05$ was used as an indicator of significance.

As shown in Tables 14 through 16, the pattern of correlations for the full sample was generally consistent with the expected pattern of convergent and discriminant validity. With the exception of Remote Memory (Buschke Free and Total Recall: trial 7) discriminant validities, as reflected by Hand Speed and Motor Performance, as well as by Mood and Social Behaviour ranged between $r = -.04$ and $r = -.23$ for Clock Drawing, $r = .04$ and $r = .27$ for Clock Setting, and $r = .05$ and $r = .23$ for Clock Reading. (The negative correlations for Clock Drawing reflect the fact that higher scores refer to inferior performance.) When Buschke Free and Total Recall (trial 7) were included, the discriminant validity coefficients ranged between $r = -.04$ and $r = -.42$ for Clock Drawing, $r = .04$ and $r = .61$ for Clock Setting, and $r = .05$ and $r = .53$ for Clock Reading. In contrast, the convergent validity correlations, as reflected through the remaining tests, ranged between $r = -.27$ and $r = -.56$ for Clock Drawing, $r = .35$ and $r = .77$ for Clock Setting, and $r = .36$ and $r = .60$ for Clock Reading (see Tables 14 to 16). Overall, and across all three tasks, Remote Memory (i.e., Buschke Free Recall trial 7) went contrary to the hypothesis since it's correlation to the Clock Test components was within the range of many of the convergent correlations.

Significance of a correlations was based on an alpha level of $p < .002$ which was used in accordance with the Bonferroni procedure for multiple comparisons (Ferguson, 1981).

Insert Tables 14 to 16 about here

Although the convergent validity coefficients were generally greater than the discriminant validity coefficients, McNemar's (1969) formula for testing the significance between correlations from the same sample was used in order to statistically test the observed pattern. Each of the discriminant validity coefficients were individually compared to each of the convergent validity coefficients. This resulted in a total of 160 comparisons for each of the three Clock Test components. Across the three tasks, 88.3% (424 from a total of 480) of the comparisons were significant in the expected direction. Specifically, 85.6% (137 of 160 comparisons) of the Clock Drawing comparisons, 90.6% (145 of 160 comparisons) of the Clock Setting, and 88.7% (142 of 160 comparisons) of the Clock Reading comparisons were significant in the hypothesized direction. Overall, with the exception of Buschke Free and Total Recall (trial 7), all of the discriminant validity coefficients were lower than the convergent validity coefficients, and were thereby in the expected direction. McNemar's (1969) test indicated that when an alpha level of $p < .01$ was used, the convergent validities were in almost all of the cases significantly greater than the discriminant validities. Briefly, the comparisons which were not significantly different (although they were in the expected direction), are described below. First, across all three components of the Clock Test, the correlations between the Clock Test components and MAS Auditory Receptive

Skills did not significantly differ from the correlations between the given component and Finger Tapping Left, Tapping Right, Hand Dynamometer Left and Dynamometer Right. Second, the correlations between Clock Drawing and Digits Forward (r = -.32), Design Reproduction (r = -.34), and MAS Expressive Language Skills (r = -.32) did not significantly differ from the correlation between Clock Drawing and Finger Tapping Right (r = -.23). Third, the correlation between Clock Setting and Design Reproduction (r = .38) did not significantly differ from the correlations between Clock setting and Finger Tapping Left (r = .27), Tapping Right (r = .27), and from Hand Dynamometer Right (r = .22). As previously stated, with the exception of Buschke Free and Total Recall (trial 7), the correlations were in the expected direction.

When the Normal and Demented groups were examined separately, a different pattern of correlation emerged.

Demented Participants

With respect to the demented participants, the convergent validity correlation coefficients were generally greater than the discriminant validity correlations, across the three Clock Test Components. The discriminant validity coefficients ranged between r = -.17 and r = .03 for Clock Drawing, r = .02 and r = .29 for Clock Setting, and r = .05 to r = .33 for Clock Reading. In contrast, the convergent validity coefficients ranged between r = -.44 and r = -.15 for Clock Drawing, r = .22 and r = .66 for Clock Setting, and r = .25 to r = .58 for Clock Reading.

McNemar's (1969) formula for testing the significance between correlations from the same sample was used in order to determine whether the convergent validity coefficients were in fact significantly greater than the discriminant validity coefficients. For the Demented participants, convergent validity coefficients were generally significantly higher than the discriminant validity coefficients (at $p < .05$). It is important to note that with the exception for Buschke Total Recall (trial 7) and Free Recall (trial 7), all of the convergent validity coefficients were higher than the discriminant validity coefficients, as hypothesized.

With respect to McNemar's significance test, each of the discriminant validity coefficients were individually compared to each of the convergent validity coefficients. This resulted in a total of 160 comparisons for each of the three Clock Test components. Across the three tasks, 84.6% (406 from a total of 480) of the comparisons were significant in the expected direction. Specifically, 75% (120 of 160 comparisons) of the Clock Drawing comparisons, 88.1% (141 of 160 comparisons) of the Clock Setting, and 90.6% (145 of 160 comparisons) of the Clock Reading comparisons were significant in the hypothesized direction. Briefly, the nonsignificant comparisons are described below. With the exception of Buschke Free and Total Recall (trial 7), the correlations were in the expected direction.

With respect to Clock Drawing: (a) the correlation between Clock Drawing and WAIS-R Similarities (r = -.15), MAS Auditory Receptive Skills (r = -.16), and Buschke Free Recall (trial 1; r = -.15) did not significantly differ from the correlations between Clock Drawing and any of the discriminant coefficients, (b) the correlation between Clock Drawing and WAIS-R Information (r = -.20) did not significantly differ from the correlation between Clock Drawing and Tapping Left (r = -.11), Tapping Right (r = -.17), and MAS Social Behaviour (r = -.12), and (c) the correlations between Clock Drawing MAS Expressive Language (r = -.23), Buschke Total Recall (trial 1; r = -.23), WAIS-R Digits Forward (r = -.19), and WAIS-R Vocabulary (r = -.27) did not significantly differ from the correlations between Clock Drawing and Tapping Left (r = -.11), Tapping Right (r = -.17), and MAS Social Behaviour (r = -12).

With respect to Clock Setting: (a) the correlation between Clock Setting and MAS Auditory Receptive Skills (r = .22) did not significantly differ from the correlations between Clock Setting and Tapping Left (r = .17), Tapping Right (r = .18), Dynamometer Left (r = .11), Dynamometer Right (r = .10), and Buschke Free (r = .17) and Total Recall (trial 7; r = .29), (b) the correlation between Clock Setting and WAIS-R Digits Forward (r = .23) did not significantly differ from the correlations between Clock Setting and Tapping Right (r = .18), and Tapping Left (r = .17), and (c) the correlation between Clock Setting and Buschke Free Recall (trial 1; r = .25) did not significantly differ from the correlations between Clock Setting and Tapping Right (r = .18) and from Clock Setting and Dynamometer Right (r = .10).

With respect to Clock Reading: (a) the correlation between Clock Reading and MAS Auditory Receptive Skills (r = .25) did not significantly differ from the correlation between Clock Reading and Dynamometer Left (r = .12), Dynamometer Right (r = .07), and Buschke Free (r = .16) and Total Recall (trial 7; r = .33), and (b) the correlation between Clock Reading and Buschke Free Recall (trial 1; r = .27) did not significantly differ from the correlations between Clock Reading and Dynamometer Left (r = .12) and between Clock Reading and Buschke Total Recall (trial 7; r = .33). Although the previous correlations were not significantly different, all convergent validity coefficients were higher than the discriminant validity coefficients (and therefore in the hypothesized direction) with the exception of Buschke Total Recall (trial 7). Although Buschke Total Recall (trial 7; a discriminant test) had a greater correlation to the Clock Test Components than did various convergent tests, these variables were never significantly greater than the convergent tests.

Normal Participants

With respect to the Normal participants, across the three Clock Test components the results lent only limited support to the hypothesis that the convergent validity correlation coefficients would be greater than the discriminant validity correlation coefficients. The discriminant validity coefficients ranged between r = -.33 and r = -.04 for Clock Drawing, r = .11 and r = .44 for Clock Setting, and r = .05 to r = .37 for Clock Reading. In contrast, the convergent validity coefficients ranged between r = -.37 and r = .01 for Clock Drawing, r = .13 and r = .54 for Clock Setting, and r = -.07 and r = .38 for Clock Reading (See Tables 14 to 16). When the correlations with Buschke Free and Total Recall (trial 7) were excluded, the following pattern was observed. First, for Clock Drawing, 16.7% (20 of 120) of the convergent validity coefficients were lower than a discriminant validity coefficient. Second, for Clock Setting, 12.5% (15 of 120) of the convergent validity coefficients were lower than a discriminant validity coefficient. Lastly, 22.5% (27 of 120) of the convergent validity coefficients were lower than a discriminant validity coefficient. Buschke Free and Total Recall (trial 7) were excluded from the previous description because of their high correlations with the Clock Test components.

McNemar's (1969) formula for testing the significance between correlations from the same sample was used in order to determine whether the convergent validity coefficients were in fact significantly greater than the discriminant validity coefficients. As previously described, each of the discriminant validity coefficients were individually compared to each of the convergent validity coefficients. This resulted in a total of 160 comparisons for each of the three Clock Test components. Across the three tasks, only 41% (197 from a total of 480) of the comparisons were significant in the expected direction. Specifically, 36.9% (59 of 160 comparisons) of the Clock Drawing comparisons, 50.6% (81 of 160 comparisons) of the Clock Setting, and 35.6% (57 of 160 comparisons) of the Clock Reading comparisons were significant in the hypothesized direction.

Briefly, the results are summarized below. Across the three components of the Clock Test, Buschke Free and Total Recall (trial 7) were either not significantly different from any of the convergent validity coefficients or were significantly different in the direction opposite to the prediction. These variables were therefore not included in any further description of the pattern of correlations for any of the Clock Test components.

With respect to Clock Drawing, the following results were obtained. First, as previously described, a number of convergent validity coefficients were equal to, or less than, at least one of the discriminant validity coefficients. These include the correlations between Clock Drawing and the following variables: (a) Design Reproduction (r = -.01), (b) WAIS-R Similarities (r = -.16), (c) Word Fluency (r = -.20), (d) MAS Auditory Receptive Skills (r = -.19), (e) MAS Visual Receptive Skills (r = -.17), (f) MAS Expressive Language (r = -.19), (g) WAIS-R Digits Forward (r = .01), (h) WAIS-R Digits Backward (r = -.14), and (i) WAIS-R Vocabulary (r = -.11). The previous tests were not significantly greater (i.e., in the hypothesized direction) than any of the discriminant validity correlation coefficients.

Second, the correlations for Clock Drawing revealed the following: (a) the correlation between Clock Drawing and Design Recall (r = -.25), WAIS-R Digit Symbol (r = -.24), MAS Orientation (r = -.26), and WAIS-R Object Assembly (r = -.35) were not significantly different from the correlations between Clock Drawing and Dynamometer Left (r = -.04), Dynamometer Right (r = -.08), MAS Social Behaviour (r = -.23), and MAS MUNSH (r = -.12), (b) the correlation between Clock Drawing and WAIS-R Information (r = -.23), and between Clock

Drawing and MAS Accessibility (r = -.27) were not significantly different from the correlations between Clock Drawing and MAS Social Behaviour (r = -.23), and MAS MUNSH (r = -.12), (c) the correlation between Clock Drawing and MAS Mental Status (r = -.32) was not significantly different from the correlations between Clock Drawing and Dynamometer Left (r = -.04), Dynamometer Right (r = -.08), and MAS Social Behaviour (r = -.23), (d) the correlation between Clock Drawing and Buschke Free Recall (trial 1; r = -.37) was not significantly different from the correlations between Clock Drawing and MAS Social Behaviour (r = -.23), (e) the correlation between Clock Drawing and Buschke Total Recall (trial 1; r = -.30) was not significantly different from the correlation between Clock Drawing and MAS MUNSH (r = -.12), (f) the correlation between Clock Drawing and Digits Forward (r = .01) was not significantly different from any of the discriminant validity correlation coefficients with the exception of the correlation between Clock Drawing and MAS Social Behaviour (r = -.23), and (g) the correlation between Clock Drawing and WAIS-R Digits Backward (r = -.14), and the correlation between Clock Drawing and WAIS-R Vocabulary (r = -.11), were not significantly different from any of the discriminant validity correlation coefficients. All of these correlations, although not significant, were in the hypothesized direction.

With respect to Clock Setting, 12.5% of the convergent validity coefficients were equal to, or less than at least one of the discriminant validity coefficients (not including Buschke Free Recall 7). These include the correlations between Clock Setting and the following variables: (a) Design Reproduction (r = .13), (b) MAS Expressive Language (r = .22), (c) MAS Accessibility (r = .22), (d) WAIS-R Digits Forward (r = .27), and (e) MAS Auditory Receptive Language (r = .29). The correlation between Clock Setting and MAS Auditory Receptive Language (r = .29) was significantly greater than its correlation with MAS MUNSH (r = .11). The remaining correlation coefficients were not significantly greater (i.e., in the hypothesized direction) than any of the discriminant validity correlation coefficients.

Although many of the remaining convergent validity tests, did not significantly differ from the discriminant validity coefficients, these correlations were in the hypothesized direction. The following discriminant validity coefficients, were in the hypothesized direction, however, they were not significantly different from the convergent validity coefficients. These include: (a) the correlations between Clock Setting and Digits Backward (r = .36), Word Fluency (r = .37), and Buschke Free (r = .34) and Total Recall (trial 1; r = .33) did not significantly differ from the correlations between Clock setting and Tapping Left (r = .30), Tapping Right (r = .22), and Dynamometer Left (r = .25), (b) the correlation between Clock Setting and WAIS-R Vocabulary (r = .41) was not significantly different from the correlations between Clock Setting and Tapping Left (r = .30), Dynamometer Left (r = .25), MAS Social Behaviour (r = .22), and MAS MUNSH (r = .11), (c) the correlation between Clock Setting and WAIS-R Information (r = .40), and the correlation between Clock Setting and WAIS-R Similarities (r = .41) did not significantly differ from Tapping Left (r = .30), (d) the correlation between Clock Setting and Design Delayed Recall (r = .47), and the correlation between Clock Setting and MAS Mental Status (r = .48), were not significantly greater than the correlation between Clock Setting and Dynamometer Left (r = .25), and (e) the correlation between Clock Setting and MAS Orientation (r = .34) was not significantly different from any of the discriminant validity coefficients with the exception of the correlation between Clock Setting and MAS MUNSH (r = .11). The remaining correlation coefficients were significantly greater than the discriminant validity coefficients.

With respect to Clock Reading, 22.5% of the convergent validity coefficients were equal to, or less than at least one of the discriminant validity coefficients (not including Buschke Free Recall (trial 7). These convergent validity coefficients were less than, or equal to, at least one discriminant validity coefficient (and were therefore not significantly different from the discriminant validity coefficients in the hypothesized direction). These include the correlations between Clock Reading and the following variables: (a) Design Reproduction (r = -.07), (b) WAIS-R Information (r = .19), (c) WAIS-R Similarities (r = .17), (d) Word Fluency (r = .16), (e) MAS Orientation (r = .19), (f) MAS Expressive Language (r = -.02), (g) MAS Accessibility (r = .07), (h) WAIS-R Digits Forward (r = .14), (i) MAS Visual Receptive Skills (r = .25), (j) Buschke Free Recall (trial 1; r = .23), and (k) WAIS-R Vocabulary (r = .12). The correlation between Clock Reading and WAIS-R Digits Forward (r = .14), Buschke Free Recall (trial 1; r = .23), and between Clock Reading and MAS Visual Receptive Skills (r = .25) were significantly greater than the correlation between Clock Reading and Tapping Right (r = .06).

The remaining convergent correlation coefficients, were in the hypothesized direction. The following coefficients however, were not significantly different from the discriminant coefficients (although they were in the hypothesized direction). These include: (a) the correlation between Clock Reading and MAS Auditory Receptive Skills (r = .28), WAIS-R Digits Backward (r = .26), Design Recall (r = .29), and Buschke Total Recall (trial 1; r = .26) did not significantly differ from the correlations between Clock Reading and Tapping Left (r = .14), Dynamometer Left (r = .19), and Dynamometer Right (r = .17), and (b) the correlation between Clock Reading and Design Delayed Recall (r = .37), MAS Mental Status (r = .35), WAIS-R Digit Symbol (r = .30), and WAIS-R Object Assembly (r = .38) did not significantly differ from the correlations between Clock Reading and Dynamometer Left (r = .19), and Dynamometer Right (r = .17).

<u>Group Differences</u>

In order to determine why the profound differences between correlation were observed for the two groups, Fisher's r-to-z transformation was used to test the difference between correlations between the Normal and Demented groups. As previously mentioned, this test is used in order to determine whether two independent correlation coefficients are significantly different from each other. Essentially, this analysis involves transforming the correlations into z-scores which are then compared by means of a t-test. This was conducted in order to determine whether the correlations between a pair of tasks differed significantly for the two groups.

As seen in Tables 14 through 16, the Demented and Normal groups significantly differed on a number of correlations. Across the three Clock Test components, the correlation between a given component and Design Reproduction and Digit Symbol was significantly higher for the Demented group than for the Normal Group. The correlations between Clock Drawing and Buschke Total Recall (trial 1), and Clock Drawing and MAS Visual Receptive Skills were significantly different by group, with higher correlations for the Demented participants than for the relatively Normal participants. Finally, the correlations between Clock Reading and most of the tests which embody the General Knowledge and Verbal skills factor, and the correlation between Clock Reading and the Digit Symbol subtest of the WAIS-R were significantly different for the two groups, with higher correlations for the demented participants than for the Normal participants. In addition, across the three Clock Test components, the correlation between a given Clock Test component and Buschke Free Recall (trial 7) was significantly higher for the Normal group than for the Demented group.

These differences may be explained in part by the fact that the size of a correlation is reduced when there is a high degree of selection or restriction in the range of a variable (Ferguson, 1981). As previously described, the descriptive statistics of the Clock Test components differ for the two groups (see Table 17.) The range and standard deviation of the Clock Test components was somewhat restricted for the Normal group, reflecting a possible floor effect for Clock Drawing and a possible ceiling effect for both Clock Setting and Reading. In fact for the Normal participants, 91.4% of the Clock Drawing errors ranged between 0 and 4 errors, 84.1% of the Clock Setting and 92.3% of the Clock Reading scores ranged between 11-15. The ranges were not restricted for the demented participants. For example, for the Demented participants, only 40% of the Clock Drawing Scores ranged between 0 and 4 errors, and only 13.5% of Clock Setting and 37.1% of the Clock Reading ranged between 11 and 15 correct. In fact, whereas the distribution of scores for the relatively normal participants were restricted, the demented participants demonstrated a fairly consistent distribution of scores. Furthermore, relative to the demented participants, the range of scores for the Normal participants was restricted across many of the tests given to the participants (See Table 17). These differences, which will be discussed further in a subsequent chapter, may explain to some degree the differences in the observed correlations between the Normal and Demented participants.

Insert Table 17 about here

Part III: Confirmatory Factor Analysis

In the final part of the results section, a Confirmatory Maximum Likelihood Factor Analysis (CFA) on a proposed model was conducted. Whereas exploratory factor analytic procedures are unrestricted in the way they allow a data set to derive a solution, a CFA uses a priori theoretical information to derive the solution (Gorsuch, 1983). A CFA refers to a process in which one tests the validity of a proposed model that is assumed to account for a body of data in terms of relatively few parameters. In addition, exploratory factor analyses are limited in the way they estimate communalities, (i.e., the squared multiple correlation of a variable as predicted from the factors), and rotate axes, (i.e., the way it maximizes high correlations and minimizes low ones). This often results in a lack of uniformity and replicability of the dimensionality of a given set of correlated variables (Khattab, Michael, & Hocevar, 1982). CFA, on the other hand, permits a relatively objective approach to testing a theoretically derived model. Specifically, CFA provides a means of testing how consistent an observed pattern of interrelationships among measured variables is in the configuration of interrelationships predicted from a set of hypothesized latent variables, or factors, which are related to one another in a manner consistent with theory (Byrne, 1988). Once the theoretical constructs (i.e., latent variables) and their indicators (i.e., measured variables) have been specified, the interrelationships among each of the constructs can be indicated, and the relationship of each construct to a measurement variable can be designated (Khattab et al., 1982).

The measurement model relates observed (i.e., measured) variables to unmeasured constructs (i.e., factors) which are specified and estimated. LISREL VII (Linear Structural Relations; Jöreskog & Sörbom, 1989) fits the observed matrix to the structure implied by the model and provides indices of how well the hypothesized model fits the data. The LISREL program attempts to minimize a maximum-likelihood loss function that is based upon differences between the original and reproduced correlation matrices, and provides an overall chi-square test of the goodness-of-fit of the proposed model (for a more detailed review of the LISREL program see Jöreskog & Sörbom, 1989.) LISREL also provides an indication of model identification, asymptotically efficient estimates of each free parameter in the proposed model under the assumptions of multivariate normality, estimates of the standard error of each fitted parameter, and additional information that is helpful in determining what changes in the proposed model would provide a better fit of the data (Jöreskog & Sörbom, 1989). A chi-square value indicates how much variance in the data is accounted for by the model and how well the model fits the data. Unlike most conventional tests of significance, within the domain of CFA, the goal is to get a non significant chi-square. This inverted goal stems from the fact that a non significant chi-square indicates a good fit between the theoretical model and the model that is derived from the data. Smaller nonsignificant values indicate a better fit. However, as the sample size increases, even minor deviations in the data can result in a statistically significant chi-square (Bentler & Bonett, 1980; Bollen, 1990). In order to compensate for this problem a number of indices independent of sample size (although still sample specific) have been devised (see Bollen, 1990). These goodness-of-fit indices indicate how well a model fits the data even when statistically it may not be perfect (Mulaik, James, Van Alstine, Bennet, Lind, & Stilwell, 1989).

A LISREL (measurement) model requires the specification of three different matrices. A full structural equation model may require the specification of up to 8 matrices. (This, however, is beyond the scope of the present study; for a detailed discussion of this issue see Anderson & Gerbing, 1988; Jöreskog & Sörböm, 1989.) The matrices include the Lambda X (Λ_x) matrix which contains the hypothesized factor loadings, the Phi (Φ) matrix which contains the correlations between the factors, and the Theta Delta (Θ_δ) matrix which contains the error/uniqueness of each measured variable. In CFA, various parameters may be constrained in the different matrices to test alternative models. On the basis of these three matrices, a reproduced correlation matrix is derived within the constraints that are imposed by the proposed model to provide the closest possible fit to the original correlation matrix (which is derived from the data).

The conceptual model used for the present analysis is represented in Figure 4. Specifically, the 4 latent constructs are depicted in ovals and include: (a) Mental Status and Memory, (b) Verbal Abilities, (c) Performance and Visuospatial abilities, and (d) Coordination and Hand Strength. The measured variables include 28 test scores which are depicted in rectangles. These measured variables were hypothesized to load on to the given latent variables based on theoretical justification and support from past and present research. Chapter 3 detailed the tests used in the model including the theoretical constructs measured by each of the tests. Although a fifth factor (Mood/Social Behaviour) was hypothesized, it was not included into the model because there were only two measured variables to indicate the construct. Although two measured variables can be used to indicate a construct, Byrne (1989) suggested that a latent factor should include a minimum of three measured variables in order to be identified. Finally, measurement error or uniqueness (depicted as ε within circles) is associated with each of the measured variables. Uniqueness refers to a composite of random measurement error and specific measurement error associated with a particular measuring instrument. In a cross-sectional study, the two can not be separated (Gerbing & Anderson, 1984).

Insert Figure 4 about here

Testing the Correlation Matrices

Green (1992; see also Werts, Rock, Linn, & Jöreskog, 1976) described a statistical technique which uses LISREL in order to evaluate whether two correlation matrices differ from one another. One reason to test for a significant difference between two correlation matrices is to evaluate whether the relations among the same variables differ across different groups. Specifically, one can evaluate whether a set of measures show the same or different patterns of relations for different groups of subjects. Before combining measures across two different (or independent) groups, it is important to know whether the measures have different patterns of intercorrelations. Thus, a multivariate test of the equality (or similarity) of correlation matrices can help determine whether an entire set of measures show a different pattern of interrelations across different groups. This matrix approach, which provides an omnibus test of differences in correlation, is similar to an ANOVA which provides an omnibus test of group mean differences (Green, 1992).

Knowing whether two correlation matrices are significantly different can act as a preliminary test in order to determine whether a single model should apply to two independent groups. Different correlation matrices imply different factor models, whereas equal (or similar) correlation matrices imply similar factor structures. According to Browne (1978; cited in Green, 1992) one would require a strong theoretical basis to factor analyze separately groups whose correlation matrices are not significantly different. In addition, if two correlation matrices are not statistically different, the two underlying factor models can be assumed to be similar.

For all of the models in this study, the analyses were conducted on the observed correlation matrices, rather than the covariance matrix, since either kind of matrix will yield the same statistical indices of fit (Cudeck, 1989). The null hypothesis for this analysis is that the two samples (normal vs. demented) have equivalent correlation matrices. In terms of the sample correlations, the two correlation matrices presented in Figure 4 were compared simultaneously. This resulted in a correlation matrix of 28 measured variables (loading on 4 latent factors) or 378 correlations in total which were simultaneously compared. The demented sample consisted of 267 participants and the normal sample was based on 151 participants. As described in the previous section, although missing data were estimated, three measured variables still contained missing data. These variables were included in the present analysis since they consisted of important information which was useful in the identification of the latent factors.

The chi-square value evaluates the null hypothesis that the two matrices are equal against the alternate hypothesis which states that the two matrices differ. The analysis to ascertain whether the correlation matrices differed, resulted in a chi-square of 2513.44 ($p < .001$) with 435 degrees of freedom. This indicates that the correlation matrices differ significantly from one another. However, since sample size has an effect on the value of the chi-square, it is important to consider other measures, in addition to that of the chi-square statistic. Alternative approaches for the evaluation of the closeness of fit have been described (Bentler & Bonet, 1980; Bollen, 1990; Byrne, 1989; Marsh, Balla, & McDonald, 1988). The following five additional approaches were used in this study. First, the Goodness of Fit Index (GFI) is considered to be the relative amount of variance and covariance jointly accounted for by the model. A GFI in the .90's indicates an excellent fit (Byrne, 1989). This test revealed a less than excellent Goodness-of-fit index (GFI = .802). Second, the Adjusted Goodness-of-fit Index (AGFI) is the GFI adjusted for degrees of freedom (Brown, 1986). Third, the relation of the chi square to the number of degrees of freedom (χ^2/df) was computed. A χ^2/df ratio of less than 2 indicates a good fit (Byrne, 1989). The ratio of the chi square to degrees of freedom (χ^2/df = 2.65) indicated that the fit was inadequate. Fourth, the Tucker-Lewis (TL; Tucker and Lewis, 1971) reliability coefficient, is the ratio of the amount of covariation explained by a proposed model to the total amount of available covariation to be accounted for by a null model in which all factors are uncorrelated; See Appendix D). A TL index in the .90's indicates an excellent fit (Byrne, 1989). The TL index (TL = .824) indicated a less than excellent fit between the two matrices. Fifth, the root mean square residual (RMSR) of the fit of the model to the data, which ranges between 0 and 1, is the average discrepancy between the elements in the sample and the hypothesized covariance matrices. A RMSR of less than .05 indicates a good fit (Byrne, 1989). The RMSR indicated that there was a fairly large discrepancy between the elements of the two correlation matrices (RMSR = .117). These additional indices are necessary because as previously described, the chi-square value is directly related to sample size, making it almost impossible not to get a significant value when the number of participants is large. Taken together, these indices indicated that the normal and demented participants' correlation matrices differed between the two groups.

Using this analysis as an indicator of the difference between the two groups, the subsequent Confirmatory Factor Analysis was conducted separately for the Normal and Demented groups.

Model 1

The initial step in a CFA is to estimate factor loadings, which are correlations between measured and latent factors when the variables have been standardized. The values of these factor loadings are indicated as λ's and are either freed to be estimated or fixed at zero. Unlike the exploratory factor analytic approach which requires that all measured variables load on all latent factors, the factor loadings for a CFA are restricted so that each measured variable loads only on the latent factors that they are hypothesized to represent.

In the present model, each measured variable was specified as having a loading on only one latent construct. The latent factors were allowed to be correlated with each other and in the initial model, the unique (i.e., random error) factors were specified as being uncorrelated among any other factors or measured variables. In the initial model, all of the coefficients in the variance-covariance matrix of the latent factors (Phi) were free to be estimated by LISREL. The diagonal elements consisted of the factor variances and the off-diagonals are covariances. The error/uniqueness components of each measured variable (i.e., the diagonal values of the theta delta matrix) were also free to be estimated by LISREL. The off-diagonal elements were initially set to zero since it was assumed that there were no correlations between the errors of the measured variables. In contrast to the common exploratory factor analytic models, LISREL allows for covariances among error/uniqueness (Marsh & Hocevar, 1985, 1987). There were 28 factor loadings, 10 coefficients in the phi variance-covariance matrix, and 28 error/uniqueness components to be estimated. Within each factor one loading was arbitrarily fixed to unity (i.e., 4 factors were set to 1), resulting in 24 estimated factor loadings. This was done in order to identify the model by setting reference indicators, and to set a scale or metric for the factors (Marsh & Hocevar, 1985). The total parameters to be estimated was 62 (see Figure 4).

As previously mentioned, for all of the models in this study, the analyses were performed on the observed correlation matrices. In the first analysis, the basic model shown in Figure 3 was fit separately for the normal and demented participants. In this analysis, the ability of the model to fit the data was tested without respecifying the parameter estimates for either of the two groups. The null model is the case in which no common factors are assumed to be correlated with each other (i.e., measurements are completely independent). This is conducted by forcing the latent factors (i.e., Phi matrix) to be uncorrelated. Bentler and Bonnet (1980), however, regard the null model as a general, most restricted model against which other less restricted models are compared in a nested sequence of models. A nested model consists of a series of similar models which have the same parameters but are ordered according to increasingly more restricted a priori constraints placed on their parameters. In the present study, the null model was one in which the latent factors were uncorrelated. However, in a series of nested models successive models were compared to the initial model using a χ^2 change test (Bentler & Bonett, 1980). A significant change in χ^2 indicates that a less restricted model (i.e., allowing more parameters to be estimated by LISREL) represents a better fitting model. The results for the CFAs, including goodness-of-fit indices, are presented in Table 18. The subsequent models differed from the original in that the modification indices which are provided by LISREL were used in order to increase the fit of the model. The modification indices represent the decrease in the chi-square that would be obtained if certain elements are free to be estimated. Respecifying a model is a form of post hoc analysis which has been criticized because of the problem of capitalization on chance (MacCallum, Roznowski, & Lawrence, 1992). In addition, it can render an analysis out of the realm of confirmatory procedures. However, through careful justification based on the substantive and theoretical consequences of the respecification this problem can be alleviated (Byrne, 1989). Therefore, before any model was respecified, a theoretical basis for any change was required. The concept of model respecification will be discussed later.

Insert Table 18 about here

A shown in Table 18, the indices of the fit of the model indicates that the initial model provides a less than adequate fit for both the Normal ($\chi^2(344)$ = 1054.98; GFI = .705; AGFI = .651; RMSR = .094; χ^2/df = 3.07; TL = .549) and Demented ($\chi^2(344)$ = 1558.03; GFI = .700; AGFI = .646; RMSR = .107; χ^2/df = 4.53; TL = .637) groups. Since an excellent fit has been described as having indices greater than .9, five additional models were tested by freeing or constraining various parameters. The successive models, which are presented in Table 18, are based on allowing the error covariances (i.e., Θ_δ) to be estimated. Such correlated errors can have important meaning in reflecting minor data covariation which can not be explained by the model. This may include covariation such as that resulting from nonrandom error which is introduced by a particular measurement method such as the effect due to item formats associated with the subscales of a particular measuring instrument (Byrne, 1989). In the present case, modifications included estimating the error covariances between: (a) Left and Right Dynamometer, Left and Right Finger Tapping (Model 1a), (b) Buschke Free Recall Trial 1 and Trial 7, and Buschke Total Recall Trial 1 and Trial 7 Model 1b), (c) WAIS-R Digits Forward and Digits Backward (Model 1c), (d) Clock Reading and Clock Setting (Model 1d), and (e)

MAS Mental Status and MAS Orientation (Model 1e). These models were conducted such that each successive model was more restricted than the previous one and included the estimation of error covariances of the previous models.

In the first additional model (Model 1a), the error covariances between Left and Right Dynamometer, and between Left and Right Finger Tapping were set free to be estimated. This model resulted in a significant chi-square for both the Normal ($\chi^2(342) = 754.48$) and Demented ($\chi^2(342) = 1263.84$) groups. The goodness-of-fit indices revealed further that the model was inadequate for both the Normal (GFI = .733; AGFI = .683; RMSR = .085; χ^2/df = 2.21; TL = .736) and Demented (GFI = .730; AGFI = .679; RMSR = .095; χ^2/df = 3.69; TL = .723) groups. The change in chi-square between this model and the original model, demonstrated that the second model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 300.50$, $p < .01$) and Demented groups ($\Delta\chi^2 = 294.19$, $p < .01$).

In the second additional model (model 1b), in addition to those error covariances set free in the previous model, Buschke Free Recall Trial 1 and Trial 7, and Buschke Total Recall Trial 1 and Trial 7 were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(340) = 667.39$) and Demented ($\chi^2(340) = 1084.62$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 87.09$, $p < .01$) and Demented groups ($\Delta\chi^2 = 179.22$, $p < .01$). The goodness-of-fit indices, however, revealed that this model was still inadequate for both the Normal (GFI = .757; AGFI = .710; RMSR = .081; χ^2/df = 1.96; TL = .791) and Demented (GFI = .776; AGFI = .733; RMSR = .081; χ^2/df = 3.19; TL = .775) groups.

In the third additional model (model 1c), in addition to those error covariances set free in the previous models, WAIS-R Digits Forward and Digits Backward were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(339) = 635.41$) and Demented ($\chi^2(339) = 1058.90$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 31.98$, $p < .01$) and Demented groups ($\Delta\chi^2 = 25.72$, $p < .01$). The goodness-of-fit indices, however, revealed that this model was still better than the previous model but was still inadequate for both the Normal (GFI = .769; AGFI = .723; RMSR = .079; χ^2/df = 1.87; TL = .810) and Demented (GFI = .779; AGFI = .736; RMSR = .080; χ^2/df = 3.12; TL = .782) groups.

In the fourth additional model (model 1d), in addition to those error covariances set free in the previous models, Clock Setting and Clock Reading were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(339) = 802.83$) and Demented ($\chi^2(339) = 1026.72$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 32.58$, $p < .01$) and Demented groups ($\Delta\chi^2 = 32.18$, $p < .01$). The goodness-of-fit indices, however, revealed that this model, although better than the previous model, was still inadequate for both the Normal (GFI = .780; AGFI = .736; RMSR = .078; χ^2/df = 1.78; TL = .830) and Demented (GFI = .785; AGFI = .742; RMSR = .081; χ^2/df = 3.04; TL = .790) groups.

In the final model (model 1e), in addition to those error covariances set free in the previous models, MAS Mental Status and MAS Orientation were also set free. This model had a significant chi-square value for both the Normal ($\chi^2(337) = 580.75$) and Demented ($\chi^2(337) = 1010.47$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal

($\Delta\chi^2 = 22.08$, $p < .01$) and Demented groups ($\Delta\chi^2 = 16.25$, $p < .01$). The goodness-of-fit indices, however, revealed that this model, although better than the previous model, was still inadequate for both the Normal (GFI = .788; AGFI = .744; RMSR = .077; χ^2/df = 1.72; TL = .843) and Demented (GFI = .787; AGFI = .744; RMSR = .082; χ^2/df = 3.00; TL = .794) groups. The factor loading matrix (Lambda) and the factor variance-covariance matrix (Phi) of this model (i.e., model 1e) are presented in Table 19.

Insert Table 19 about here

The model was also applied when the samples were collapsed. This was conducted since indices of fit less than the suggested fit of .9 have been used to describe an adequate fit. As previously described, when the significance between correlation matrices was performed the goodness-of-fit indices, although not excellent, were approaching good fitting models. The basic four factor model revealed a significant fit ($\chi^2(344) = 2620.64$, $p < .001$). In addition, the Goodness-of-Fit indices revealed an inadequate model (GFI = .678; AGFI = .620; RMSR = .074; TL = .226). The final model (which was identical to Model 1e) revealed a better fitting model than the original model $\chi^2(337) = 1365.96$, $p < .001$; GFI = .782; AGFI = .737; RMSR = .059; TL = .643; $\Delta\chi^2$(last model - first model) = 1254.69, $p < .01$).

<u>Summary</u>. In addition to the original hypothesized model, five additional models were examined in which various error covariance parameters were free to be estimated. Based on a chi-square change test, each successive model was significantly better than the previous one. However, based on six goodness-of-fit indices which were used for both the Normal and Demented

groups, the final models did not reveal an adequate fit. When the two groups were collapsed, the model was still inadequate.

Model 2

Even when the previous model was modified, an adequate representation of the data did not emerge. In order to further examine the model, a secondary model was tested. This model differed from the previous one in that the Multi-Focus Assessment Scale (MAS; Koch et al., 1984) was not included in the model. The second model was conducted since there is no reported literature in which the MAS items were factor analyzed, thus precluding that the test dimensions are themselves accurate. The nested model consisted of the same respecifications as those imposed on the first model. The model consisted of 18 factor loadings (i.e., 4 factors were set to 1), 10 coefficients in the Phi variance-covariance matrix, and 22 error/uniqueness components to be estimated, a total of 50 parameters. The results of the CFA are reported in Table 20.

Insert Table 20 about here

The results of the second model demonstrated a better fit when the MAS was removed (see Table 20). The initial model provided a less than adequate fit for both the Normal ($\chi^2(203) = 816.41$; GFI = .713; AGFI = .642; RMSR = .103; $\chi^2/df = 4.02$; TL = .239) and Demented ($\chi^2(203) = 993.58$; GFI = .763; AGFI = .704; RMSR = .108; $\chi^2/df = 4.89$; TL = .280) groups. As previously described, an excellent fit has been described as having indices greater than .9.

Four of the five additional models which were tested in the previous model were re-examined for the present model. The successive models, which are presented in Table 20, were based on allowing the error covariances (i.e., Θ_δ) to be estimated. These models were conducted such that each successive model was more restricted than the previous one and included the estimation of error covariances of the previous models.

In the first additional model (Model 2a), the error covariances between Left and Right Dynamometer, and between Left and Right Finger Tapping were set free to be estimated. This model resulted in a significant chi-square for both the Normal ($\chi^2(201) = 516.26$) and Demented ($\chi^2(201) = 701.35$) groups. The goodness-of-fit indices revealed further that the model was inadequate for both the Normal (GFI = .752; AGFI = .688; RMSR = .088; χ^2/df = 2.57; TL = .604) and Demented (GFI = .805; AGFI = .755; RMSR = .092; χ^2/df = 3.49; TL = .539) groups. The change in chi-square between this model and the original model, demonstrated that the second model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 300.15$, $p < .01$) and Demented groups ($\Delta\chi^2 = 292.21$, $p < .01$).

In the second additional model (model 2b), in addition to those error covariances set free in the previous model, Buschke Free Recall Trial 1 and Trial 7, and Buschke Total Recall Trial 1 and Trial 7 were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(199) = 440.96$) and Demented ($\chi^2(199) = 620.21$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 75.30$, $p < .01$) and Demented groups ($\Delta\chi^2 = 81.14$, $p < .01$). The goodness-of-fit indices, however, revealed that this model was better, although still inadequate for both the Normal (GFI = .781; AGFI = .722; RMSR = .082; χ^2/df = 2.22; TL = .693) and Demented (GFI = .828; AGFI = .781; RMSR = .078; χ^2/df = 3.12; TL

= .607) groups.

In the third additional model (model 2c), in addition to those error covariances set free in the previous models, WAIS-R Digits Forward and Digits Backward were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(198) = 408.86$) and Demented ($\chi^2(198) = 595.95$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 32.10$, $p < .01$) and Demented groups ($\Delta\chi^2 = 27.26$, $p < .01$). The goodness-of-fit indices, however, revealed that this model was still better than the previous model but was still inadequate for both the Normal (GFI = .796; AGFI = .740; RMSR = .079; χ^2/df = 2.06; TL = .733) and Demented (GFI = .835; AGFI = .789; RMSR = .079; χ^2/df = 2.99; TL = .631) groups.

In the fourth additional model (model 2d), in addition to those error covariances set free in the previous models, Clock Setting and Clock Reading were also set free. This model also had a high chi-square value for both the Normal ($\chi^2(197) = 376.14$) and Demented ($\chi^2(197) = 558.29$) groups. The change in chi-square between this model and the previous one demonstrated that this new model was significantly better than the previous model for both the Normal ($\Delta\chi^2 = 32.72$, $p < .01$) and Demented groups ($\Delta\chi^2 = 34.66$, $p < .01$). The goodness-of-fit indices, however, revealed that this model, although better than the previous model, was still inadequate for both the Normal (GFI = .812; AGFI = .759; RMSR = .078; χ^2/df = 1.91; TL = .771) and Demented (GFI = .845; AGFI = .800; RMSR = .077; χ^2/df = 2.83; TL = .661) groups. The factor loading matrix (Lambda) and the factor variance-covariance matrix (Phi) of this model (i.e., model 1e) are presented in Table 21.

Insert Table 21 about here

The model was also applied when the samples were collapsed. This was conducted since indices of fit less than the suggested fit of .9 have been used to describe an adequate fit. As previously described, when the significance between correlation matrices was performed the goodness-of-fit indices, although not excellent, were approaching good fitting models. The basic four factor model revealed a significant fit ($\chi^2(203) = 1534.68$, $p < .001$). In addition, the Goodness-of-Fit indices revealed an inadequate model (GFI = .754; AGFI = .693; RMSR = .085; TL = .455). The final model (which was identical to Model 2d) revealed a better fitting model than the original model $\chi^2(197) = 603.99$, $p < .001$; GFI = .880; AGFI = .845; RMSR = .042; TL = .829; $\Delta\chi^2$(last model - first model) = 930.69, $p < .01$).

Summary. In addition to the original hypothesized model, four additional models were examined in which various error covariance parameters were free to be estimated. This set of models differed from the previous set in that the MAS was not included in the model. Based on a chi-square change test, each successive model was significantly better than the previous one. Based on six goodness-of-fit indices which were used for both the Normal and Demented groups, the final models approached an adequate fit. Although not an excellent fit, goodness-of-fit indices of .8 have been referred to as reflecting an adequate fit (Byrne, 1989). When the two groups were collapsed, the model was even better. These analyses therefore suggest that the best fitting model is one in which the MAS is not included and in which the error covariances between subscales of a given test (e.g., Tapping Left and Right) are free to be estimated.

Chapter 5
Discussion

The present study examined various psychometric characteristics of the Clock Test as an assessment tool for the diagnosis of dementia (with particular emphasis on Alzheimer's disease). This chapter consists of an interpretation of the results of the study, with attention to their implications for the use of the Clock Test in research and clinical settings. The limitations of the present study will be addressed, and suggestions regarding directions for future research, modifications to the Clock Test components, and improvements to the general scoring protocol will be presented. In the following section, the results will be briefly summarized with respect to the specific hypotheses examined.

Summary of Results

Clock Test Reliability. The first set of research questions addressed the structure of the Clock Test. The first hypothesis was that the three Clock Test components would demonstrate high internal consistency reliability coefficients. This hypothesis was supported by showing that the reliabilities for Clock Setting (Cronbach's α = .95) and Clock Reading (Cronbach's α = .86) were high when examined for all subjects combined. When examined separately by group (i.e., for the Normal and Not Demented groups combined, and for the Unlikely Alzheimer's disease, and the Possible and Probable Alzheimer's disease groups combined), the Clock Setting reliabilities were high for both the Demented (Cronbach's α = .93) and relatively normal (Cronbach's α = .84) groups. The Clock Reading reliability estimates for the relatively normal adults (Cronbach's α = .57), however, was much lower than that of the demented adults (Cronbach's α = .83). These results suggest that the Clock Setting and

Clock Reading were highly reliable for the demented participants, whereas only Clock Setting showed high reliability for the relatively normal participants.

Although these results could indicate that the relatively normal participants performed less consistently on Clock Reading than did the demented participants, a second explanation may be related to the fact that the relatively normal participants were performing at ceiling levels. This would have an effect on reliability since the size of a correlation is reduced when there is a high degree of selection or restriction in the range of a variable (Cohen & Cohen, 1983). In fact, for Clock Reading over 90% of the relatively normal participants, scored between 11 and the maximum possible score of 15 (M score = 13.79, *SD* = 2.18), whereas only 37% of the demented participants had scores within this range. In addition, whereas the distribution of scores for the relatively normal subjects were restricted, the demented participants (M score = 8.22, *SD* = 5.11) demonstrated a fairly consistent distribution of scores. This suggests that for the relatively normal participants, the restriction in scores is a better explanation for the lowered reliability. This indicates, however, that there is a problem with the scale for the relatively normal individuals, since a reliable measure should produce a fairly consistent distribution of scores, and a high estimate of reliability.

Cronbach's alpha was not computed for Clock Drawing. Rather, the following three sets of correlations were performed. First, the correlations between a Clock Drawing variable and the dimension to which it is associated were computed. Second, the correlations between the Clock Drawing variables and the total Clock Drawing score were computed. Third, the correlations between the Clock Drawing dimensions and the total Clock Drawing score were computed. With the exception of Distortion errors (which were negligible for both groups), the correlations between the dimension and total Clock Drawing scores were significant (r = .32 to r = .72) for both the normal and demented groups. However, only 52% (13 of 25) of the correlations between the Clock Drawing variables and total Clock Drawing score were significant. These significant correlations were most predominant for errors of Omission, Perseverations, Rotations, and Misplacements for both the relatively normal and demented groups, as well as for Additions and Substitutions for the latter group. Distortion error items did not correlate with the total score for either group, although for the demented participants each item did correlate significantly to the total category score. Those items which were not significantly correlated to the total Clock Drawing score however, were, in fact correlated to the category total. Together, these sets of correlations suggest that for the demented participants, performance is correlated significantly to either the total Clock Drawing score or to the dimension score. The results, however, were not replicated for the relatively normal participants, since only 38% of the variables correlated to the total Clock Drawing score. In fact, for nine (34.6%) of the variables, correlations could not be computed since these types of errors were not produced (see Table 4).

Tuokko et al. (1990a) stated that the Clock Drawing scoring system involves both a quantitative and qualitative schema implying that in addition to total scores one should examine the number, and types of errors, that are made on each of the categories. This distinction is important since certain errors are scored on a present versus absent basis, so that these items may not correlate significantly to the total score but they do correlate with the category score. For example, a score of two on distortions, while not having a major effect on the total score, indicates that the individual has made two out of three possible distortion errors. In addition, since each category is unequal with respect to the maximum possible score, the item to total correlations may not be as relevant as the item to category correlations. Although the present study examined the relationship of the Clock Drawing variables to the dimension and total score, the present analyses did not examine whether any of the items are redundant or spurious.

Group Differences. The second hypothesis concerned the effect of age, education and diagnostic group on the performance of individuals on the Clock Test components. Specifically, it was hypothesized that: (a) performance would be negatively related to age and positively related to education, (b) performance for the demented groups would be inferior to that of the relatively normal groups, and (c) an Age by Diagnostic Group interaction would emerge, such that the performance of the relatively normal adults would decrease as a function of age, whereas this would not be observed for the demented participants who would be performing at floor levels. The use of orthogonal contrasts within regression analyses allowed for the analysis of a categorical variable (diagnostic group) and continuous variables (age, education) within a single analysis (Pedhazur, 1982).

As hypothesized, diagnostic group differences were observed. Across the three components the relatively normal participants outperformed the demented groups. This is consistent with previous research in which demented participants performed worse than normal controls on each component comprising the Clock Test (Albert & Moss, 1984; Anderson, et al., 1990; Lowenstein et al., 1989; Shulman et al., 1986). In addition, the performance of the two AD groups was inferior to that of the Unlikely AD group (i.e., participants with a diagnosis other than AD). This is consistent with the findings that individuals with Alzheimer's disease experience visuospatial deficits early on (Cummings & Benson, 1983), whereas deficits in visuospatial and constructional abilities for other forms of dementia may not be as severe as those observed in AD. For example, when comparing Alzheimer's disease groups, to groups consisting of individuals with other forms of dementia, the AD individuals usually perform worse. This has been shown in individuals with

Huntington's disease (Brouwers, Cox, Martin, Chase, & Fedio, 1984; Rouleau, et al., 1992), pseudodementia (Dastoor et al., 1991), multi-infarct dementia (Perez, Rivera, Meyer, Gray, Taylor, & Matthew, 1975), and Pick's disease (Cummings, 1985). In addition, for both Clock Setting and Clock Reading, the Normal participants outperformed the Not Demented participants, and Possible AD participants outperformed the Probable AD participants.

Overall, these analyses revealed that for all three tasks, diagnostic group was the best predictor of performance. In addition, for Clock Setting and Clock Reading significant differences were observed between each of the groups (i.e., Normal followed, in turn, by the Not Demented group, the Unlikely AD group, the Possible AD group, and the Probable AD groups.) For Clock Drawing, however, the Normal and Not Demented groups, and the Possible and Probable Alzheimer's disease groups, did not differ from each other. The two former groups performed significantly better than the two latter groups. In addition, the Unlikely Alzheimer's disease group significantly outperformed the two Alzheimer's disease groups.

In previous research, old normal adults over the age of 60 performed more poorly on Clock Drawing and Clock Setting than younger adults (Albert & Kaplan, 1980; Farver, 1975; Goodglass & Kaplan, 1983). Age differences were not revealed in the present study, probably because there were no adults below the age of 54 years. In addition, the regression analyses revealed that most of the variance was accounted for by diagnostic group, leaving little non-extraneous variance to be accounted for by age. Finally, the fact that the diagnostic groups account for so much variance may, in part, be due to the fact that the age range was restricted (range = 54 to 84 years). In order to further examine the age effect, and to demonstrate the hypothesized existence of an Age by Diagnostic group interaction, simple main effects of the analyses were conducted. Simple main effects provide a more precise and quantitative analysis of within group differences than do interactions (May, Masson, & Hunter, 1990; Pedhazur, 1982). The Age by Diagnostic group interaction examined the hypothesis that an age effect would occur for cognitively intact participants but would not be observed in cognitively impaired participants (due to floor performance). Whereas age was a significant predictor of performance for the relatively normal participants, age was not a significant predictor of performance for the demented participants. This analysis therefore supported the hypothesis that across the three tasks, a significant age effect was demonstrated for the relatively normal participants but not for the demented group. This suggests that performance decreases as a function of normal aging, and supports previous research. Further research, in which there is a greater age range, would be required in order to better examine the effect of age on the Clock Test. Unlike Farver (1975) and Albert and Kaplan (1983), no education effect was observed.

Group Differences in Clock Drawing. This set of analyses was specific to the Clock Drawing component. This test component requires the scoring of 25 types of errors which are then classified into 7 categories. These analyses addressed the question of whether the two groups generated different types of errors, or whether they only differed with respect to the number of errors. This was examined through a series of ANOVAs which investigated group differences with respect to errors, and through a series of hierarchically nested Discriminant Function Analyses which investigated the number of dimensions required to classify individuals into groups. As hypothesized, the pattern of errors differed by diagnostic group. The performance of the Normal and Not Demented adults was similar across the seven Clock Drawing dimensions, although the normal participants consistently had the fewest errors for each of the seven categories. Although the Possible AD participants had a better total score than the Probable AD participants, they produced significantly more errors on three of the error categories. The Probable AD participants made most of their errors on Omissions which are considered to be the most severe type of error. One possible explanation for these results may be that the Probable AD participants were manifesting severe gross impairments which hindered their ability to perform the task (e.g., they did not understand the instructions, they had a severely limited attention span), whereas the Possible AD participants were still able to process the task to some degree, but were making many errors due to the dementia. Further research, in which this hypothesis is actually tested, will be required in order to investigate the reasons for the differential performance between the two AD groups.

As a test of global impairment (i.e., when the groups were collapsed), the relatively normal participants performed significantly better than the demented participants across all seven categories. The demented participants made most of their errors on the Omission category, followed by Misplacement errors. Normal participants on the other hand, produced mostly Misplacements errors. This is consistent with the description of the types of errors produced by normal old adults (Albert & Kaplan, 1983; Farver, 1975). With respect to classifying individuals based on the seven Clock Drawing categories, the final four categories of errors (Misplacements, Distortions, Substitutions, and Additions) did not improve the classification power over the first three types of errors (Omissions, Perseverations, and Rotations) in classifying individuals into their respective diagnostic groups. When the groups were collapsed according to cognitive status, only Omission errors were required to classify individuals into cognitively intact versus cognitively impaired groups. The remaining categories did not improve the classification of individuals into groups. Essentially, for diagnostic purposes one needs to score only the first three types of errors without fear of losing important clinical data. The remaining types of errors could be used for anecdotal and descriptive purposes.

Construct Validity. This set of hypotheses examined the convergent and discriminant validities of the Clock Test components. As previously stated, demonstrating construct validity necessitates showing that the target test correlates highly with other tests with which it is theoretically related, and does not correlate with tests from which it theoretically unrelated. As hypothesized, for both the relatively normal and the demented groups, the correlation between Clock Setting and Clock Reading was significantly higher than their respective correlations with Clock Drawing, thus supporting Tuokko et al.'s (1989) claim that Clock Drawing assesses something different from Clock Setting and Clock Reading.

The neuropsychological test battery was divided into eight discriminant validity tests and 20 convergent validity tests. It was hypothesized that the Clock Test components would correlate highly with other tests of visuospatial conceptualization and organization. These tests included the WAIS-R Block Assembly, WAIS-R Object Assembly, and Design Reproduction. For the demented participants, these tests had high correlations with Clock Setting ($\underline{r}$ = .36 to $\underline{r}$ = .64), Clock Reading ($\underline{r}$ = .35 to $\underline{r}$ = .49) and Clock Drawing ($\underline{r}$ = -.30 to $\underline{r}$ = -.42). (The negative associated with Clock Drawing reflects the fact that higher scores (i.e., more errors) reflect greater impairment in this task.) For the relatively normal participants, only Block Design and Object Assembly were significantly correlated to the Clock Test components. Design Reproduction had a low nonsignificant correlation with the Clock Test components. The scores for the relatively normal participants on Design Reproduction were very restricted ($\underline{M}$ = 4.97, *SD* = .2; maximum possible score = 5 points). Since correlations are reduced when there is a restricted range of scores, it is not surprising that this correlation is so low (see Tables 14 to 16).

With respect to the demented participants, WAIS-R Digit Symbol, which was described as a complex problem solving task, had the highest correlation with Clock Drawing (r = -.44), Clock Setting (r = .66), and Clock Reading (r = .58. In addition, the correlations of Design Recall and Design Delayed Recall, which were described as tests of visuospatial memory, were highly correlated with Clock Drawing (r = -.30 to r = -.36), Clock Setting (r = .48 to r = .54), and Clock Reading (r = .47 to r = .48). For Clock Drawing, seven of the ten tests which were described as assessing general knowledge and abilities, and verbal skills were significantly correlated to Clock Drawing (r = -.15 to r = -.36). Furthermore, for Clock Setting (r = .31 to r = .51), and Clock Reading (r = .25 to r = .51), all of these tests correlated significantly with the particular Clock Test component. Lastly, the four tests which were described as measuring working memory and attention were significantly correlated to Clock Setting (r = .23 to r = .48), and Clock Reading (r = .34 to r = .51). Only three of these tests correlated significantly to Clock Drawing (r = -.19 to r = -.34). Although significance tests were not conducted between Clock Test components, it is interesting to note that the correlations for Clock Setting and Clock Reading were similar, and they were higher than those of Clock Drawing (when the negative direction of the correlation is ignored).

The pattern of correlations differed for the relatively normal participants. The WAIS-R Digit Symbol task had a high correlation with Clock Drawing (r = -.24), Clock Setting (r = .46), and Clock Reading (r = .30). However, unlike the demented group these correlation were not the highest. Rather, the correlations of the visuospatial memory tasks, were more highly correlated with Clock Drawing (r = -.25 to r = -.35), Clock Setting (r = .45 to r = .47), and Clock Reading (r = .29 to r = .37). The significance between the difference of the correlations between these tests, however, were not conducted. For Clock Drawing, only four of the ten general knowledge and abilities, and verbal skills tests were significantly correlated to Clock Drawing (r = -.11 to r = -.27). Furthermore, eight of these tests were significantly correlated to Clock Setting (r

= .22 to _r_ = .48), and only three of the ten were significantly correlated to Clock Reading (_r_ = .25 to _r_ = .51). Lastly, only two of the four working memory and attention tests were significantly correlated to Clock Drawing (_r_ = .01 to _r_ = -.37), three of the four were significantly correlated to Clock Reading (_r_ = .14 to _r_ = .26), and all four tests were significantly correlated to Clock Setting (_r_ = .27 to _r_ = .36). Although significance tests were not conducted between the Clock Test components, it is interesting to note that the correlations for Clock Setting were generally the highest, followed in turn by, Clock Reading, and Clock Drawing (when the negative direction is ignored).

These results demonstrate that the hypothesis which stated that the Clock Test components would correlate significantly with other tests with which they are theoretically related was supported for the demented participants. The results demonstrated only limited support for the relatively normal group.

It was also hypothesized that the Clock Test components would not be significantly correlated to tests from which they theoretically differ. In addition, it was hypothesized that these discriminant validity correlation coefficients would be significantly less than the convergent validity correlation coefficients. For the demented participants, the correlations between MAS MUNSH and MAS Social Behaviour, which were described as measures of social behaviour and mood, did not correlate significantly with Clock Drawing (_r_ = -.03 to _r_ = -.12), Clock Setting (_r_ = .02 to _r_ = .07), and Clock Reading (_r_ = .05 to _r_ = .10). The correlations for Finger Tapping and Hand Dynamometer (each of which had a score for the left and right hands), were described as measures of hand strength, intensity and motor speed. Whereas, one of these tests was significantly correlated to Clock Drawing (_r_ = .03 to _r_ = -.17), and two of the tests were significantly correlated to Clock Setting (_r_ = .10 to _r_ = .18), none of the tests were significantly correlated to Clock Reading (_r_ = .07 to _r_ = .12). Finally, with respect to those measures which were described as remote memory tasks, neither task was significantly correlated to Clock Drawing (r = -.08 to r = -.10) or Clock Reading (r = .16 to r = .33). One of the tasks, however, was significantly correlated to Clock Setting (r = .17 to r = .29).

With respect to the demented participants, the correlations demonstrate that in most cases the Clock Test components do not correlate significantly with tests from which they theoretically differ. However, significance tests were conducted in order to determine whether the discriminant validity correlation coefficients were in fact significantly less than the convergent validity correlation coefficients. Each convergent validity coefficient was thereby compared to each discriminant validity coefficient using McNemar's formula for testing the significance between correlated correlations. This resulted in a total of 160 comparisons for each task. As described in the previous chapter, 75% of the Clock Drawing comparisons, 88.1% of the Clock Setting comparisons, and 90.6% of the Clock Reading comparisons were significant in the hypothesized direction. This suggests that for the demented participants, the hypothesis that the discriminant validity correlation coefficients would be significantly less than the convergent validity correlation coefficients was generally supported.

For the relatively normal group, whereas MAS MUNSH (r = -.12) did not significantly correlate to Clock Drawing, the correlation between Clock Drawing and MAS Social Behaviour was significant (r = -.23). These did not correlate significantly with Clock Setting (r = .11 to r = .22), and Clock Reading (r = .05 to r = .06). Whereas, two of the hand strength, intensity and motor speed tests were significantly correlated to Clock Setting (r = .18 to r = .30), none of the tests were significantly correlated to Clock Drawing (r = -.04 to r = -.08), or to Clock Reading (r = .06 to r = .19). Finally, with respect to the remote memory tasks, both tasks were significantly correlated to Clock Drawing (r = -.22 to r = -.33), Clock Setting (r = .28 to r = .44), and Clock Reading (r = .27 to r = .37).

These correlations demonstrate that, except for remote memory, the

Clock Test components do not generally correlate significantly to tests from which they theoretically differ. However, significance tests were once again conducted in order to determine whether the discriminant validity correlation coefficients were in fact significantly less than the convergent validity correlation coefficients. As described in the previous chapter, only 36.9% of the Clock Drawing comparisons, 50.6% of the Clock Setting comparisons, and 35.6% of the Clock Reading comparisons were significant in the hypothesized direction. This suggests that for the normal participants, the hypothesis that the discriminant validity correlation coefficients would be significantly less than the convergent validity correlation coefficients was not generally supported.

Although remote memory was hypothesized to be a measure of discriminant validity, the results do not support this hypothesis. The remote memory task, as described in Chapter 3, is actually the delayed recall trial of the Buschke Cued Recall Test. The participant is administered three trials prior to the delayed remote memory trial (Tuokko & Crockett, 1989). When the first trial (Buschke FR1 and TR1) was correlated to the remote memory trial, the resulting correlations were high for both Free Recall (r = .72, .83; relatively normal and demented, respectively) and Total Recall (r = .82, .72; relatively normal and demented, respectively). This task, therefore, may not be a good measure of remote memory, since a delay of only fifteen minutes does not in fact test remote memory.

Significance tests were conducted in order to determine whether the correlations were significantly different between the relatively normal and demented groups. Five (17.9%) of the 28 correlations between the tests and Clock Drawing, three (10.7%) of the 28 correlations between the tests and Clock Setting, and 11 (39.3%) of the 28 correlations between the tests and Clock Reading differed significantly between the two groups. These differences can be explained in part by the restricted range of scores on the part of the normal subjects, who performed near ceiling on a number of tasks (see Table 17).

A limitation to these analyses was described by Hertzog et al (1989). They state that the convergent and discriminant validity is inadequately assessed by zero-order correlations for the following reasons. First, higher correlations do not necessarily imply that the two variables measure the same construct, rather they may measure different, but correlated constructs. Second, magnitudes of correlations are attenuated by the reliability and validity of the measures. Thus, low correlations may reflect low reliability, low validity, or both, rather than the influence of different constructs on each measures. Furthermore, they state that although inferences that are based on differential magnitudes of pairs of correlations may often be accurate, the possible biasing effect in combination of differential reliabilities and validities can be problematic. The authors suggest that an alternate method for addressing construct validity uses both confirmatory factor analysis and structural equation models. Future research should therefore use these alternate methods since modelling provides a more direct representation of the convergent and discriminant validity.

Factor Structure of the Clock Test. The final set of hypotheses involved the factor structure of the tests administered in the present study. The use of a Confirmatory Factor Analysis (CFA) allows one to test the hypothesized relations of a set of observed variables to the underlying constructs, while allowing these constructs to intercorrelate. Since many psychological tests are in fact multifactorial and multidimensional, the tests are often intercorrelated to some degree. The use of CFA therefore alleviates this problem by not restricting the factors to be uncorrelated. In addition, the analysis determines the degree of correlation between the factors. Although none of the models satisfied conventional criteria for an excellent fit (i.e., goodness-of-fit indices greater than .9), the best fitting model was one in which the MAS was not included, and various error parameters were free to be estimated. These modified parameters included those which were affected by a particular measurement instrument and were modified only on the basis of theoretical justification (see Byrne, 1989). Because the tests used in the present study are often used to screen for dementia, the present results provide important information about the overall factor structure of these instruments. For example, if one were attempting to develop a brief battery of neuropsychological tests, one would attempt to get measures that assess as many different abilities as possible. S/he would, for example, use one or two measures from each of the four dimensions and use them as a brief test. Future research should try to optimize the model in order to get a better fit of the data. In addition, one could examine whether the four factor model would emerge in samples with different diagnostic groups such as head injury, Parkinson's disease, or Huntington's disease. This would aid in the construction of test batteries by obtaining measures which would be most suitable for the study of a given disorder.

The Restricted Range Issue. The restricted range of the various Clock Test components was used in order to explain some of the findings for the relatively normal individuals. Further analyses were conducted in order to examine the issue of a restricted range. Across the three Clock Test components, there was a tendency for the relatively normal individuals to perform near the ceiling for Clock Setting and Clock Reading, and at floor levels for Clock Drawing. As previously described, for Clock Drawing, 91.4% of the relatively normal participants had between 0 and 4 errors (M = 2.07, *SD* = 2.86), whereas, only 40% of the Demented participants (M = 8.83, *SD* = 7.26), had scores within this range. For Clock Setting, 84.1% of the relatively normal participants scored between 11 and the maximum possible score of 15 (M = 12.87, *SD* = 2.69), whereas, only 13.5% of the Demented participants (M = 5.53, *SD* = 4.55) had scores within this range. For Clock Reading, 92.3% of the relatively normal participants scored between 11 and the maximum possible score of 15 (M score = 13.79, *SD* = 2.18), whereas only 37.1% of the demented participants (M = 8.22, *SD* = 5.11) had scores within this range. In addition, whereas the distribution of scores for the relatively normal subjects were restricted, the demented participants demonstrated a fairly consistent distribution of scores.

In order to further examine differences between the Demented and relatively normal participants with respect to the range of scores, the homogeneity of variances were examined. In order to determine whether or not the relatively normal and demented groups were homogeneous with respect to variance, the F_{max} test was conducted. In the F_{max} test, the larger group variance is divided by the smaller group variance. If this value is greater than that of the critical F (for N-1 (larger group variance), and N-1 (smaller group variance) degrees of freedom), then the two groups differ significantly with respect to variance. In other words, the variance between the groups is not homogeneous. The F_{max} test revealed that the variance of the relatively normal individuals was significantly less than that of the demented individuals for each of the three components of the Clock Test. Keppel (1991), however stated that an one should become concerned only if the F_{max} is appreciably greater than three. The F_{max} for the three components of the Clock Test were 6.43, 2.86, and 5.47 (Clock Drawing, Clock Setting, and Clock Reading, respectively).

Further analyses were conducted in order to examine whether the three groups (i.e., Unlikely AD, and Possible and Probable AD) which comprised the Demented group differed with respect to homogeneity of variance. Across the three Clock Test components for the three groups, the F_{max} tests revealed that the variance of the three groups was homogeneous. That is, the F_{max} test did not reveal significant differences between the groups. In addition, F_{max} tests were conducted in order to examine the homogeneity of variance between the Not Demented group and the Normal controls. These analyses revealed that, whereas the variance for Clock Drawing and Clock Reading were homogeneous across the two groups, the variance was not homogeneous for Clock Setting.

Although heterogeneity of variance violates the assumptions underlying the analysis of group differences, Winer (1971) and Keppel (1991) suggest that lowering the alpha level should correct the problem. (The decreased alpha level in these cases were $p < .001$.) These findings indicate that the variances of the Clock Test components differ for the relatively normal and demented groups. On the other hand, there is only modest support for the contention that normal adults performed at absolute ceiling.

Limitations and Modifications

The limitations of the present study must be examined with respect to the data and design. First, the study relied entirely on diagnostic data and included a fairly large number of measured variables. With studies such as this one (i.e., in which diagnostic data are used), there is often a concern with respect to a circularity problem. That is to say, the target task is often tested as a diagnostic tool and used to some degree in the diagnosis of an individual. In the present study, however, the Clock Test data were not used in the diagnostic process, rather the data were simply collected (H. Tuokko, personal communication, July 1992).

An important limitation to the present study concerns the specificity of the diagnoses. Although a diagnosis is based on a large number of tests and is agreed upon by a multidisciplinary team, there is no valid way to evaluate independently the specificity or correctedness of the diagnosis. For example, the diagnostic procedure in this study did not include many of the standard tests (e.g., Boston Naming Test, Immediate and Delayed Logical Memory, Trails A and B of the Halstead-Reitan Battery, the Mini-Mental State Exam (MMSE), Beck Depression Inventory (BDI), or the Geriatric Depression Scale) which are typically used in Alzheimer's disease research. This is a considerable problem, since even with the most rigid and careful testing, the clinical diagnosis of Alzheimer's disease is correct only about 80% of the time (Sunderland, et al. 1989). Therefore, the assignment to diagnostic groups may not have been accurate. (However, H. Tuokko (personal communication, July 1992) stated that for the UBC Alzheimer's Disease and Related Disorders Clinic, the overall hit rate between correct diagnosis and autopsy is about 90%.) In addition, it has been shown in previous studies that the Functional Rating Scale (FRS; Crockett et al., 1989), which was used in the diagnostic process, had a correct classification rating of over 85%. Nevertheless, the present study must be evaluated with regard to this limitation.

Since the participants were categorized into groups based on their diagnoses, there was no experimental control with such concepts as random assignment into groups, or random sampling from the general population. This lack of control, which is a concern in all studies involving diagnostic groups, limits any form of randomness which is a goal when conducting psychological research and when performing statistical analyses. An additional concern is that the normal controls did not undergo the same testing procedures as did the demented participants. The FRS was not administered to them, and they were not assessed by the multidisciplinary team. It is therefore possible that some of these normal participants were in fact cognitively impaired with a dementing syndrome. However, when the performance of the relatively normal participants on the neuropsychological tests were compared to published norms, as a group, their performance was within the normal range (see Table 17). Finally, there was no control with respect to such disorders as arthritis or hearing impairments, which could have an effect on performance of the Clock Test. Participants with visual impairments (self reported or observed) were not included in the present study. However, the presence of visual impairments were not tested for, thus a participants could produce a flawed clock because s/he misperceived the stimulus. It is therefore possible that those individuals who performed poorly on the Clock Test did so due to impaired visual capacity, rather than due to a cognitive impairment associated with dementia. This is an important point that must be addressed in future research.

Further limitations include the fact that neither the Clock Test components nor the items making up any of the components were randomized. The test was given in a prespecified order to all of the participants. This confound makes it impossible to determine whether there is an item effect (i.e., are some items more difficult than others?) or a practice effect (i.e., does performance increase or decrease with practice?). This is even more relevant given the fact that the same times are used across the three tasks. Because the same times are used across the three tasks, there is even greater potential for learning to take place. Future research must address this problem by determining whether (a) there actually is a practice effect, and (b) there is an item, or difficulty, effect. In addition, the full neuropsychological battery was given to the participants in a prespecified order. Performance on the Clock Test, which was the last test given, may have been affected by an attentional deficit that is associated with dementia (Lezak, 1983). Because the length of the administration of the neuropsychological battery required approximately 90 minutes to complete, it is possible that the demented participants performance was reduced due to a problem of sustained attention over such a long period of time. In a future study, the administration of the Clock Test, within the test protocol should controlled.

Furthermore, as digital clocks become even more common, it is quite possible that participants make use this type of clock, rather than the conventional clock face used in the present study. Although this may present a thorny problem for the future of the Clock Test, it may be of less concern in the present study. First, for the participants in the present study (i.e., they were over 54 years of age), telling time on an analog clock was assumed to be an overlearned response which would not be affected until a marked decline in cognitive performance has occurred. Although no data were collected pertaining to the clocks normally used the participants, it was assumed that the performance of the participants were valid indicators of their ability to tell time. Therefore, errors could be assumed to be indicators of some cognitive dysfunction. Second, analog clocks are still so prevalent in our society that it is probable that individuals who use digital clocks would still be familiar with analog clocks. One important issue is that older adults may use digital clocks because they are having difficulty reading an analog clock. Furthermore, telling time from a digital clock may be a compensatory mechanism for visual impairment or even mild cognitive dysfunctions (for a review of compensation theory in older adults see Dixon & Bächman, in press). Although the issue of analog versus digital clocks will become more prevalent if digital clocks continue to replace analog clocks, the Clock Test can still be used with the current older cohort. This problem, however, must be addressed in further research by asking participants (or an informant) the types of clocks they use in their homes, and then determining whether the use of analog or digital clock is related to performance.

The participants were never asked to identify the minute and hour hands. It is quite possible that a low score, rather than being indicative of cognitive dysfunction, may in fact reflect the fact that the participant didn't know what the minute and hour hands were (i.e., they could not tell time). Therefore, prior to testing, the participant should be asked to identify the type (or types) of clocks that they have in their house, and at some point during the testing process they should be asked to identify the minute and hour hands. This information would allow for a more direct assessment of the Clock Test and would provide beneficial information which can be used for validation purposes.

An additional concern involves the lack of a true measure of depression in the present study. Since individuals with a dementing disorder are often depressed, this lack of a depression measure may be problematic since there is no way to ascertain whether one's performance is being affected by dementia, depression, or both. Although the MUNSH verifies the presence or absence of general demoralization it does not assess depression. This information would be important for the purpose of further validation.

In an earlier section some potential advantages of the Clock Test were noted. These included the assertions that the Clock Test was relatively free of cultural biases, and that it was non-threatening to the patient (Sunderland et al., 1989). It should be emphasized that neither of these advantages have in fact been empirically tested, rather they were considered as advantages based on anecdotal information. Whereas many neuropsychological tests have demonstrated profound cultural differences (e.g., WAIS-R Vocabulary, WAIS-R Information), it has been assumed that the Clock Test is unaffected by these biases. Whereas cultural biases have not been examined for Clock Drawing or Clock setting, Lowenstein et al. (1989), stated that Clock Reading does not appear to be subject to cultural biases when compared to common measures of language and cognition. In fact, Clock Reading has been translated and administered to individuals from various cultural backgrounds. With respect to the issue of threat to the patient, it has been assumed that since the Clock Test is a simple test, it was non-threatening to the participants. The fact that the test is simple may in fact have an effect opposite to that which is expected. Based on self-efficacy theory (for a review of the theory see Bandura, 1977), one's expectation of performance may in fact have an effect on both how one perceives the task, and how one performs on the task. If a difficult task is administered to an individual with memory problems, the individual may perceive the task, as being difficult and may therefore have an expectation of failure. The individual may be less threatened by a task that is perceived to be beyond his/her abilities. However, if that individual is given a simple task, s/he may expect to complete it successfully. Furthermore, if the individual fails to perform adequately on the simple task, s/he may therefore feel threatened by the test because of the previous expectation of successful performance, and this could have an effect on future performance. The question of whether or not the Clock Test is threatening to cognitively impaired adults should be explored in future research.

A limitation to the present study includes the fact that males and females may have performed differently on the Clock Test components and on the remaining tests of the neuropsychological test battery. The possible effects of gender on the previously described results were not examined. Specifically, the diagnostic groups differ with respect to the ratio of females with males. Differences between the groups in the number of males to females may in fact play a role in explaining some of the observed group differences. A simple examination of the effect of gender was examined through a series of ANOVAs which were conducted on the three components of the Clock Test. These analyses revealed that across the five groups of participants, there were no significant differences between the performance of males and females for any of the Clock Test components. In addition, the Gender by Diagnostic group interactions were not significant. Furthermore, ANOVAs conducted on the seven dimensions of Clock Drawing did not reveal any gender effects, nor did they reveal Gender by Diagnostic group interactions. Although these results suggest that there are no gender differences with respect to the Clock Test components, the gender effect could be examined further.

Lastly, although the present study included a fairly large number of measured variables, it was based on those variables that were collected for diagnostic purposes. With respect to the convergent and discriminant validities, additional tests would have been valuable. For example, a copying test such as the Rey-Osterrieth Complex Figure Test or the Bender-Gestalt Test would have been useful. These tests assess both visuospatial constructional ability and visuospatial memory (Spreen & Strauss, 1991). In addition, those test mentioned above (Boston Naming Test, Immediate and Delayed Logical Memory, Trails A and B of the Halstead-Reitan Battery, and the Mini-Mental State Exam) would have been worthwhile, as they are often used in the assessment of dementia. The use of these tests would strengthen the present study, as they are commonly used in the diagnostic process. Furthermore, additional tests of discriminant validity, such as a sensory-perceptual abilities test (e.g., tactile or auditory stimulation), an additional motor test (e.g., Purdue Pegboard test), or a personality test (e.g., MMPI) would be beneficial.

Suggestions for Future Research and Modifications to the Clock Test

Although the present study answered a number of important questions with respect to the psychometric properties of the Clock Test, a number of important and interesting question remain unanswered. The first issue for future research involves the establishment of normative data which would be truly representative of the entire elderly population. The Clock Test score of a given individual could then be compared to normative expectations based upon individuals of similar age, gender, education, and possibly other demographic variables. Bayles and Kaszniak (1987) stress the importance of normative data stating that its use could lessen diagnostic errors. The establishment of normative Clock Test data using the present scoring scheme is currently being undertaken and will be analyzed shortly (H. Tuokko, personal communication, June, 1991).

Future research should also attempt to match individuals with respect to various demographic variables and level of severity and test them in a controlled experiment. As stated earlier, a lack of experimental control is somewhat inherent in diagnostic research. However, even manipulations of the order of the test components may be important. For example, Clock Drawing is given first, followed by Clock Setting and Clock Reading. If the order were manipulated, it is possible that a different pattern of results may be obtained. Furthermore, if the individual items which comprise Clock Setting and Clock Reading were manipulated (e.g., from easiest to hardest, hardest to easiest, or randomly), the results may vary from those observed in the present study. In addition, through the use of a normal population, control over the order of the test components and items would not interfere in the diagnostic process. This may at least determine the effect of practice on the test. Lastly, the diagnosis of the subjects should be established independently of, and prior to, any experimental manipulations.

Many cognitive measures are insensitive to subtle changes in skills which are affected in the initial phase of Alzheimer's disease and other dementias. An interesting question for research is whether, over time, the Clock Test is sensitive to subtle cognitive changes. If it can be shown that subtle cognitive changes are reflected through Clock Test scores, then this would further clarify the value of this test as a diagnostic tool.

Finally, although some evidence for the reliability and validity of the Clock Test has been demonstrated, several suggestions to improve the test are proposed. First, the strategies used by an individual when proceeding with the task should be described. The examination of strategies adopted when completing the test has been acknowledged to be important when giving neuropsychological tests since global achievement does not address the nature and effectiveness of the strategies that an individual employs enroute to either a correct or incorrect solution (e.g., Anderson et al., 1990). In fact Werner (1956, cited in Kaplan, 1989) stated that careful monitoring of behaviour towards a solution (i.e., the process) is more likely to provide useful information than that obtained from final scores (i.e., achievement). This distinction between process and product approaches has also been the basis of the revision of the WAIS-R for neuropsychological testing (WAIS-NI; Kaplan, Fein, Morris, & Delis, 1991). Impairments in function may be evident in the strategy that an individual employs even if the final product is correct. An incorrect final solution may reflect either varying degrees of dysfunction or differential focal pathology (Albert & Kaplan, 1980).

A second methodological revision concerns the use of a center point in the materials presented to the participants. A center point is a point in the center of the clock face from which the participant would draw the hands. The center point is used in various testing procedures (e.g., Anderson et al., 1990; Sunderland et al., 1989). Although the effect of the center point on the performance of individuals has not yet been tested, its benefit in orienting the participant to the task has been acknowledged (Anderson et al., 1990). This orientation can be important, since in the present scoring system a hand is defined as any line "arising from the center or near-center of the clock" (Tuokko et al., 1990a, p. 8). The presence of a center point will simplify scoring decisions. A number of participants, rather than drawing the hands from the center of the clock to the correct number, notched or drew a line over the correct number. Since the line did not emanate from the center or near center of the clock, they did not receive credit for getting the correct time (although one can infer than they still possess the concept of time). By orienting the individual with a center point this problem may occur less frequently. A third proposed change concerns the fact that the test is not timed and there is no maximum time limit to finish the tasks. If the test were timed this may result in additional information which could be used for validation purposes.

With respect to the Clock Setting component, an individual can only receive a third point for the relative lengths of the hands if they correctly set the minute and hour hand. Therefore an individual who gets one of the hands correct and the relative lengths correct will only get a score of 1 thereby making it seem as if s/he got 2 errors rather than 1. The scoring should therefore be modified, such that even if the times are incorrect, one still receives a point for the relative lengths of the hands.

Finally, a fourth task should be appended to the Clock Test. Because there may be a dissociation based on the demands of a task, and since previous research has demonstrated that performance may improve for copying a clock from a model versus spontaneous drawing (Goodglass & Kaplan, 1983; Rouleau et al., 1992), a participant should be required to copy a clock from a model. This would allow one to observe whether performance increases during the copying trial. Moore and Wyke (1984) found that copies, although containing errors, generally include more details than drawings. A number of researchers already use clock copying in addition to Clock Reading and Clock Setting when testing individuals. With the addition of Clock Copying, a complete picture will be obtained with respect to visuospatial abilities and the abstract conceptualization of time.

General Conclusions

Although the findings of the present study point to some deficiencies in the Clock Test, it may still make a contribution to neuropsychological assessment. This test has shown some evidence of reliability and validity. With further refinements, it may be able to assist in the discrimination between cognitively intact and cognitively impaired individuals. The advantages of the Clock Test over other neuropsychological tests include the measure's ease, the minimal time required to administer it, and the simplicity in scoring. The suggested revisions, which should be based on further research findings, would likely improve the test's utility as a diagnostic instrument.

References

Albert, M., & Kaplan, E. (1980). Organic implications of neuropsychological deficits in the elderly. In L. W. Poon, J. L. Fozard, L. S. Cermack, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging: Proceedings of the George A. Talland memorial conference (pp. 403-432). Hillsdale, NJ: Erlbaum.

Albert, M., & Moss, M. (1984). The assessment of memory disorders in patients with Alzheimer's disease. In L. Squire & N. Butters (Eds.), Neuropsychology of memory (pp. 224-235). New York: Guilford.

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

Anastasi, A. (1988). Psychological testing (6th ed.). New York, NY: Macmillan.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103, 411-423.

Anderson, M., Dobbs, A., & Rule, B. (1990). Clock drawing in Alzheimer patients and normals. Unpublished Manuscript. University of Alberta.

Atkinson, L., & Cyr, J. J. (1984). Factor analysis of the WAIS-R: Psychiatric and standardization samples. Journal of Consulting and Clinical Psychology, 52, 714-716.

Bäckman, L., & Dixon, R. A. (in press). Psychological compensation: A theoretical framework. Psychological Bulletin.

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

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Bayles, K. A., & Kaszniak, A. W. (1987). Communication and cognition in normal aging and dementia. Boston, MA: Little, Brown.

Beck, N. C., Horwitz, E., Seidenberg, M., Parker, G., & Frank, R. (1985). WAIS-R factor structure in psychiatric and general medical patients. Journal of Consulting and Clinical Psychology, 53, 402-405.

Belsky, J.K. (1984). The psychology of aging: Theory, research and practice. Monterey: Brooks/Cole Publishing.

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Benton, A. L., & Fogel, M. L. (1962). The assumption that a common disability underlies failure in copying drawings and in copying two- and three-dimensional block patterns is not justified. Archives of Neurology, 7, 347.

Benton, A. L., & Hamsher, K. (1978). Multilingual Aphasia Examination (Revised). Iowa City: University of Iowa.

Black, F. W., & Strub, R. L. (1976). Digit repetition performance in patients with focal brain damage. Cortex, 14, 12-21.

Blazer, D. (1989). Current concepts: Depression in the elderly. New England Journal of Medicine, 320, 164-166.

Blieszner, R., & Shifflett, P. A. (1990). The effects of Alzheimer's disease on close relationships between patients and caregivers. Family Relations, 39, 57-62.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. Psychological Bulletin, 107, 256-259.

Bornstein, R. A. (1985). Normative data on selected neuropsychological measures from a nonclinical sample. Journal of Clinical Psychology, 41, 651-659.

Bornstein, R. A. (1986). Normative data on intermanual differences on three tests of motor performance. Journal of Clinical and Experimental Neuropsychology, 8, 12-20.

Borod, J. C., Goodglass, H., & Kaplan, E. (1980). Normative data on the Boston Diagnostic Aphasia Examination, Parietal Lobe Battery, and the Boston Naming Test. Journal of Clinical Neuropsychology, 3, 209-215.

Botwinick, J., & Storandt, M. (1974). Memory, related functions, and age. Springfield, Il: C. C. Thomas.

Brink, T. L. (1984). Limitations of the GDS in cases of pseudodementia. Clinical Gerontology, 2, 60-61.

Brouwers, P., Cox, C., Martin, A., Chase, T., & Fedio, P. (1984). Differential perceptual-spatial impairment in Huntington's and Alzheimer's dementias. Archives of Neurology, 41, 1073-1076.

Brown, R. L. (1986). A comparison of the LISREL and EQS programs for obtaining parameter estimates in confirmatory factor analysis studies. Behavior Research Methods, Instruments, and Computers, 18, 382-388.

Browne, M. W. (1978). The likelihood ratio test for the equality of correlation matrices. British Journal of Mathematical and Statistical Psychology, 31, 209-217.

Buschke, H. (1973). Selective reminding for analysis of memory and learning. Journal of Verbal Learning and Behavior, 12, 543-550.

Buschke, H. (1984). Cued recall in amnesia. Journal of Clinical Neuropsychology, 6, 433-440.

Byrne, B. M. (1988). Measuring adolescent self-concept: Factorial validity and equivalency of the SDQ III across gender. Multivariate Behavioral Research, 23, 361-375.

Byrne, B. M. (1989). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer-Verlag.

Campbell, D. T. (1960). Recomendations for APA test standards regarding construct, trait, or discriminant validity. American Psychologist, 15, 546-553.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Canadian Collaborative Study of the Epidemiology of Dementia's, Including Alzheimer's Disease. (1990). Core Grant Application. Ottawa, Ontario: Author.

Cliff, N. (1987). Analyzing multivariate data. San Diego, CA: Harcourt Brace Jovanovich.

Cohen, J. (1957a). A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. Journal of Consulting Psychology, 6, 451-457.

Cohen, J. (1957b). The factorial structure of the WAIS between early adulthood and old age. Journal of Consulting Psychology, 21, 283-290.

Cohen, J., & Cohen, P. (1983). Applied multiple regression / correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Costa, L. D. (1975). The relation of visuospatial dysfunction to digit span performance in patients with cerebral lesions. Cortex, 11, 31-36.

Coval, M., Crockett, D., Holliday, S., & Koch, W. (1985). A Multifocus Assessment Scale for use with elderly populations. Canadian Journal on Aging, 4, 101-109.

Crawford, M. S., & Boer, D. P. (1985). Content bias in the WAIS-R Information subtest and some Canadian alternatives. Canadian Journal of Behavioural Science, 17, 79-86.

Critchley, M. (1953/1966). The parietal lobes. New York: Hafner.

Crockett, D., Tuokko, H., Koch, W., & Parks, R. (1989). The assessment of everyday functioning using the Present Functioning Questionnaire and the Functional Rating Scale in elderly samples. Clinical Gerontologist, 8, 3-25.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach, L. J. (1990). Essentials of psychological testing (5th ed.). New York, NY: Harper & Row.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.

Cummings, J. L. (1985). Clinical neuropsychiatry. Orlando, Fl: Grune & Straton.

Cummings, J. L., & Benson, D. F. (1983). Dementia: A clinical approach. Boston, MA: Buttersworth.

Dastoor, D. P., Schwartz, G., & Kurzman, D. (1991). Clock drawing: An assessment technique in dementia. Journal of Clinical and Experimental Gerontology, 13, 69-85.

Delis, D. C., Freeland, J., Kramer, J. H., & Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. Journal of Consulting and Clinical Psychology, 56, 123-130.

desRosiers, G., & Kavanagh, D. (1987). Cognitive assessment in closed head injury: Stability, validity and parallel forms for two neuropsychological measures of recovery. International Journal of Clinical Neuropsychology, 9, 162-173.

Dewis, E., & Baumann, A. (1982). Alzheimer's disease: The silent epidemic. The Canadian Nurse, 7, 33-35.

Dixon, W. J. (1988). BMDP Statistical Software Manual. Berkeley: University of California Press.

Dodrill, C. B. (1978). The Hand Dynamometer as a neuropsychological measure. Journal of Consulting and Clinical Psychology, 46, 1432-1435.

Dodrill, C. B., & Troupin, A. S. (1975). Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. Journal of Nervous and Mental Disease, 161, 185-190.

Doyon, J., Bouchard, R., Morin, G., Bourgeois, C., & Côté, D. (1990). Nouveau système de cotation quantatif du test dl'horloge: Sensibilité et spécificité dans la démence de type Alzheimer. Presented at the IVe Congrès International Francophone de Gérontologie: Abrégés des Communications Scientifiques. Montreal, Quebec.

Edlund, M. (1987). Psychological time and mental illness. New York: Gardner.

Erber, J. T., Botwinick, J., & Storandt, M. (1981). The impact of memory on age differences in Digit Symbol performance. Journal of Gerontology, 36, 586-590.

Farver, P. (1975). Performance of older adults on a test battery designed to measure parietal lobe functions. Unpublished master's thesis, Boston University College of Allied Health Professions, Boston, MA.

Ferguson, G. A. (1981). Statistical analysis in psychology and education (5th ed.). New York: McGraw Hill.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical guide for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-198.

Friedman, P. J. (1991). Clock drawing in acute stroke. Age and Ageing, 20, 140-145.

Gainotti, G., Caltagirone, C., Masullo, C., & Miceli, G. (1980). Patterns of neuropsychological impairment in various diagnostic groups of dementia. In L. Amaducci, A. N. Davison, & P. Antuono (Eds.), Aging of the brain and dementia. New York: Raven.

Gallagher, D. (1986). Assessment of depression by interview methods and psychiatric rating scales. In L. W. Poon (Ed.), Clinical memory assessment of older adults (pp. 202-213). Washington, DC: American Psychological Association.

Gaylord, S. A., & Marsh, G. R. (1975). Age differences in the speed of a spatial cognitive process. Journal of Gerontology, 39, 674-678.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. Journal of Consumer Research, 11, 572-580.

Golden, C. J., Purisch, A. D., & Hammeke, T. A. (1976). Luria-Nebraska Neuropsychological Battery: Forms I and II (Manual). Los Angeles, CA: Western Psychological Services.

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd Edition). Philadelphia: Lea and Febiger.

Gooddy, W. (1988). Time and the nervous system. New York: Praeger.

Gorsuch, R. L. (1983). Factor analysis (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

Green, J. A. (1992). Testing whether correlation matrices are different from each other. Developmental Psychology, 28, 215-224.

Grossi, D., & Orsini, A. (1978). The visual crosses test in dementia: An experimental study of 110 subjects. Acta Neurologica, 33, 170-174.

Grossi, D., Orsini, A., & Michele, G. (1978). The copying of geometric drawings in dementia. Acta Neurologica, 33, 355-360.

Gurland, B. J., & Cross, P.S. (1986). Public health perspectives on clinical memory testing of Alzheimer's disease and related disorders. In L. W. Poon (Ed.), Clinical memory assessment of older adults (pp. 11-20). Washington, DC: American Psychological Association.

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

Hertzog, C., Hultsch, D. F., & Dixon, R. A. (1989). Evidence for the confergent validity of two self-report metamemory questionnaires. Developmental Psychology, 25, 687-700.

Jorm, A. F., Korten, A. E., & Henderson, A. S. (1987). The prevalence of dementia: A quantitative integration of the literature, Acta Psychiatra Scandinavia, 76, 465-479.

Johansson, B., Zarit, S. H., & Berg, S. (1992). Changes in cognitive functioning of the oldest old. Journal of Gerontology: Psychological Sciences, 47, 75-80.

Jöreskog, K. G., & Sörbom, D. (1989). LISREL VII: Analysis of linear structural relationships by the method of maximum likelihood. Mooresville, IN.: Scientific Software.

Kahn, R. L., Goldfarb, A. I., Pollack, M., & Peck, A. (1960). Brief objective measures for the determination of mental status in the aged. American Journal of Psychiatry, 117, 326-328.

Kaplan, E. (1989). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), Clinical neuropsychology and brain function: Research, measurement, and practice. (pp. 125-168). Washington, DC: American Psychological Association.

Kaplan, E., Fein, D., Morris, R., & Delis, D. C. (1991). The Wechsler Adult Intelligence Scale - Revised as a neuropsychological instrument. San Antonio, TX: The Psychological Corporation.

Katzman, R. (1986). Differential diagnosis of dementing illnesses. Neurologic Clinics, 4, 329-340.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Englewood Cliff, NJ: Prentice Hall.

Khattab, A. M., Michael, W. B., & Hocevar, D. (1982). The construct validity of higher order structure-of-intellect abilities in a battery of tests emphasizing the product of transformations: A confirmatory maximum likelihood factor analysis. Educational and Psychological Measurement, 42, 1089-1105.

King, D. A., Caine, E. D., Conwell, Y., & Cox, C. (1991). The neuropsychology of depression in the elderly: A comparative study of normal aging and Alzheimer's disease. Journal of Neuropsychiatry, 3, 163-168.

Koch, W. J., Coval, M., Crockett, D. J., & Anderson, A. (1984). Multi-Focus Assessment Scale Manual: For evaluation of cognitive, mood and language functioning in the frail elderly. Vancouver, BC: University of British Columbia Press.

Kozma, A., & Stones, M. J. (1980). The measurement of happiness: Development of the Memorial University of Newfoundland Happiness Scale (MUNSH). Journal of Gerontology, 35, 906-912.

Lachman, M. E., Baltes, P., Nesselroade, J. R., & Willis, S. L. (1982). Examination of personality - Ability relationships in the elderly: The role of the Contextual (Interface) Assessment Mode. Journal of Research in Personality, 16, 485-501.

Lawton, M. P. (1971). The functional assessment of elderly people. Journal of the American Geriatrics Society, 19, 465-481.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. Journal of Clinical Psychology, 42, 332-342.

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

Lowenstein, D. A., Amigo, E., Duara, R., Guterman, A, Hurwitz, D., Berkowitz, N., Wilkie, F., Weinberg, G., Black, G., Gittelman, B., & Eisdorfer, C. (1989). A new scale for the assessment of functional status in Alzheimer's Disease and related disorders. Journal of Gerontology: Psychological Sciences, 44, 114-121.

Luria, A. R. (1963). Restoration of function after brain injury. New York: Pergamon.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111, 490-504.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. Psychological Bulletin, 97, 562-582.

Marsh, H. W., & Hocevar, D. (1987). The factorial invariance of students' evaluations of college teaching. American Educational Research Journal, 21, 341-366.

May, R. B., Masson, M. E. J., & Hunter, M. A. (1990). Application of statistics in behavioral research. New York, NY: Harper & Row.

McCann, J. T. (1991). Convergent and discriminant validity of the MCMI-II and MMPI personality disorder scales. Psychological Assessment, 3, 9-18.

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

McNemar, Q. (1969). Psychological statistics (4th ed.). New York: Wiley.

Miller, W. (1975). Psychological deficits in depression. Psychological Bulletin, 82, 238-260.

Mittenberg, W., Seidenberg, M., O'Leary, D. S., & DiGiulio, D. V. (1989). Changes in cerebral functioning associated with normal aging. Journal of Clinical and Experimental Neuropsychology, 11, 918-932.

Moore, V., & Wyke, M. A. (1984). Drawing disability in patients with senile dementia. Psychological Medicine, 14, 97-105.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C. D. (1989). Evaluation of goodness-of-fit indexes for structural equation models. Psychological Bulletin, 105, 430-445.

Mueller, J. H., & Overcast, T. D. (1976). Free recall as a function of test anxiety, concreteness, and instructions. Bulletins of the Psychonomic Society, 8, 194-196.

Murphy, K. R., & Davidshofer, C. O. (1988). Psycholohical Testing: Principles and applications. Englewood Cliffs, NJ: Prentice-Hall.

Nussbaum, P. D., Fields, R. B., & Starrat, C. (1992, February). Comparison of three scoring procedures for clock drawing. Paper presented at the annual meeting of the International Neuropsychological Society, San Diego, CA.

Parker, K. (1983). Factor analysis of the WAIS-R at nine age levels between 16 and 74 years. Journal of Consulting and Clinical Psychology, 51, 302-308.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). NY: Holt, Rinehart, and Winston.

Perez, F. I., Rivera, V. M., Meyer, J. S., Gay, J. R. A., Taylor, R. L., & Mather, N. T. (1975). Analysis of intellectual and cognitive performance in patients with multi-infarct dementia, verebro-basilar insufficiency with dementia and Alzheimer's disease. Journal of Neurology, Neurosurgery, and Psychiatry, 38, 533-540.

Poon, L. W. (1985). Differences in human memory with aging: Nature, causes, and clinical implications. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (2nd ed., pp. 427-462). New York: Van Nostrand Reinhold.

Reitan, R. M., & Wolfson, D. (1985). The Halstead-Reitan Neuropsychological Test Battery: Theory and interpretation. Tucson, AZ: Neuropsychology Press.

Robertson, D., Rockwood, K., & Stolee, P. (1982). A short mental status questionnaire. Canadian Journal on Aging, 1, 16-20.

Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C., & McGuire, K. (1992), Quantitative and qualitative analyses of clock drawing in Alzheimer's and Huntington's disease. Brain and Cognition, 18, 70-87.

Sattler, J. (1982). Age effects on Wechsler Adult Intelligence Scale - Revised tests. Journal of Consulting and Clinical Psychology, 50, 785-786.

Saunders, D. R. A. (1960). A factor analysis of the Information and the Arithmetic items of the WAIS. Psychological Reports, 6, 367-383.

Savage, R. D., Britton, P. G., Bolton, N., & Hall, E.H. (1973). Intellectual functioning in the aged. New York: Harper & Row.

Shulman, K. I., Shedletsky, R., & Silver, I. L. (1986). The challenge of time: Clock drawing and cognitive function in the elderly. International Journal of Geriatric Psychiatry, 1, 135-140.

Sinoff, G. (1991, October). Depression, dementia, and clock drawing in the elderly. Paper presented at the annual meeting of the Canadian Association on Gerontology, Toronto, Ontario.

Smith, D. M. (1989). The occurrence of Alzheimer's disease and related diseases in Canada. Ottawa, LCDC Typescript Report.

Spitz, H. H. (1972). Note on immediate memory for digits: Invariance over the years. Psychological Bulletin, 78, 183-185.

Spreen, O., & Strauss, E. (1991). A compendium of neuropsychological tests: Administration, norms, and commentary. New York: Oxford.

SPSS-X. (1988). Statistical Packages for the Social Sciences: User's Guide (3rd ed.). New York: McGraw-Hill.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87, 245-251.

Strauss, E., & Wada, J. (1988). Hand preference and proficiency and cerebral speech dominance determined by the carotid amytal test. Journal of Clinical and Experimental Neuropsychology, 10, 169-174.

Sunderland, T., Hill, J. L., Mellow, A. M., Lawlor, B. A., Gundersheimer, J., Newhouse, P. A., & Grafman, J. H. (1989). Clock drawing in Alzheimer's disease: A novel measure of dementia severity. Journal of the American Geriatric Society, 3, 725-729.

Tabachnick, B. G., & Fidell, L. S. (1983). Using multivariate statistics. New York: Harper & Row.

Terry, R., & Katzman, R. (1983). Senile dementia of the Alzheimer's type: Defining a disease. In R. Katzman & R. Terry (Eds.), The neurology of aging (pp. 51-84). Philadelphia: F. A. Davis.

Tuokko, H., & Crockett, D. (1989). Cued recall and memory disorders in dementia. Journal of Clinical and Experimental Neuropsychology, 11, 278-294.

Tuokko, H., Beattie, B., Crockett, D., & Horton, A. (1989, June). Validation of a new approach to the clock test. Paper presented at the annual meeting of the Canadian Psychological Association, Halifax, NS.

Tuokko, H., Crockett, D., Beattie, B., Horton, A., & Wong, M. (1986, February). The use of rating scales to assess psycho-social functioning in demented patients. Paper presented at the annual meeting of the International Neuropsychological Society, Denver, CO.

Tuokko, H., Hadjistavropoulos, T., Miller, J. A., & Beattie, B. L. (1992). The Clock Test: A sensitive measure to differentiate normal elderly from those with Alzheimer disease. Journal of Clinical and Experimental Neuropsychology.

Tuokko, H., Horton, A., Hadjistavropoulos, T., & Beattie, B. (1990a). The Clock Test: Manual for administration and scoring. University of British Columbia, BC.

Tuokko, H., Hadjistavropoulos, T., & Beattie, B. (1990b, October). Clock Drawing: A quantitative and qualitative scoring system. Presented at the annual scientific and educational meeting of the Canadian Association on Gerontology, Victoria, BC.

Tuokko, H., Olson, S., Peters, L. Horton, A., Beattie, B., & Crockett, D. (1988, October). A validation study of a new approach to the clock test. Presented at the 41st annual scientific meeting of the Gerontological Society of America, San Francisco, CA.

Tuokko, H., Vernon-Wilkinson, R., Weir, J., & Beattie, B. L. (1991). Cued Recall and early identification of dementia. Journal of Clinical and Experimental Neuropsychology.

Van der Horst, L. (1932). The psychology of constructive apraxia: Psychological views on the conception of space. Psychiatria en Neurologia, 36, 661-677.

Violato, C. (1986). Canadian version of the Information subtests of the Wechsler Tests of Intelligence. Canadian Psychology, 27, 69-74.

Warrington, E. K. (1969). Constructional apraxia. In P. J. Vinken & G. W. Bruyn (Eds.), Handbook of clinical neurology. (vol. 4, pp. 180-201). Amsterdam: New-Holland.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale - Revised Manual. New York: The Psychological Corporation.

Weinberg, J., Diller, L., Gerstman, L., & Schulman, P. (1972). Digit span in right and left hemiplegics. Journal of Clinical Psychology, 28, 361.

Weingartner, H. (1986). Automatic and effort-demanding cognitive processes in depression. In L. W. Poon (Ed.), Clinical memory assessment of older adults (pp. 218-226). Washington, DC: American Psychological Association.

Weingartner, H., Kaye, W., Smallberg, S. A., Ebert, M. H., Gillin, J. C., & Sitaram, N. (1981). Memory failure in progressive idiopathic dementia. Journal of Abnormal Psychology, 90, 187-196.

Werner, H. (1956). Microgenesis and aphasia. Journal of Abnormal and Social Psychology, 52, 347-353.

Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1976). Comparison of correlations, variances, covariances, and regression weights with or without measurement error. Psychological Bulletin, 83, 1007-1013.

Weschler, D. (1955). Weschler Adult Intelligence Scale. Manual. New York: Psychological Corporation.

Weschler, D. (1981). Manual for the WAIS-R. New York: Psychological Corporation.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wolf-Klein, G. P., Silverstone F. A., Levy A. P., Brod, M. S., Breuer, J. (1989). Screening for Alzheimer's Disease by clock drawing. Journal of the American Geriatrics Society, 37, 730-734.

Yaguchi, K. (1981). A study of tree-drawings in aged groups: An examination of formal indices of the drawings. Journal of Child Development, 17, 32-34.

Yesavage, J. A. (1986). The use of self-rating depression scales in the elderly. In L. W. Poon (Ed.), Clinical memory assessment of older adults (pp. 213-218). Washington, DC: American Psychological Association.

Yesavage, J. A., Bink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression rating scale: A preliminary report. Journal of Psychiatric Research, 17, 37-49.

Yeudall, L.T., Fromm, D., Reddon, J.R., & Stefanyk, W. O. (1986). Normative data stratified by age and sex for 12 neuropsychological tests. Journal of Clinical Psychology, 42, 918-946.

Appendix A

Classification of Clock Errors

0) No errors in clock, spacing is accurate, time is clearly 'Ten minutes past eleven o'clock'.

1) Visuospatial:
   a) Mildly impaired spacing of times
   b) Draws times outside circle
   c) Draws in lines (spokes) to orient spacing

2) Errors in denoting time as 10 minutes past 11 o'clock:
   a) Omits minute hand
   b) Omits both hands
   c) Draws a single line from 11 to 12, 11 to 2, 12 to 2
   d) Indicates time by writing in words '10 past 11'
   e) Turns page while writing, so numbers appear upside down

3) Visuospatial:
   a) Moderately impaired spacing in time (so that the time can not be accurately denoted)
   b) Omits numbers
   c) Continues beyond 12 to 13, 14, 15, etc.
   d) Writes numbers counter clockwise, right-left reversal, right or left side neglect, dysgraphia (unable to write numbers accurately)

4) Severely disorganized spacing: Confuses 'Time' - writes in minutes, times of day, month or season

5) Unable to make any reasonable attempt at the clock (Exclude severe depression or other psychotic states)

Note. The original classification of errors had no category for clocks with no errors.

Adapted by Dastoor, Schwartz, & Kurzman, (1990) from Shulman, Shedletsky, & Silver, (1986).

Appendix B

Scoring Criteria for Clock Drawing

Error type	Defining Criteria
Omissions	
Neglect	Part of the clock face has been neglected. That is, no entries have been made within the space defined by two or more consecutive 5 minute segments.
Number	Number(s) from 1 through 12 is left out. (1 error per number left out).
Hands	The stem of the clock hand, arising from the center or near center of the clock is left out (1 error per hand left out). If scored as a Hand error, it can't be scored as a misplaced hand or substitution error.
Perseverations	
Repetition of Numbers	Numbers have been repeated (1 error per repetition).
Hand perseveration	More than 2 hands are drawn on the clock. Score 1 error for each extra hand.

Error type	Defining Criteria
Sequence of Numbers	Numbers, in sequence, following 12 are written in. Each additional number is scored as an error.
Rotations	
Clock Face Rotation	All numbers are displaced either to the left or right or are written counter clockwise. An error is given when numbers are written counterclockwise even if some numbers fall in the correct number locations.
Number Rotation	Numbers of numbers not drawn right side up (i.e, greater than 45 degrees rotation from the horizontal).
Reversal Rotation	Numbers are drawn backwards. One error is scored for each number drawn in reverse. Note that for double digit numbers it is only necessary for one digit to be written in reverse for this number to be scored as an error.

Error type	Defining Criteria
Hand Rotation	Time is shown in its mirror image. The hands are pointing to the correct locations but the relative lengths of the hands are reversed. That is the hands are drawn to indicate 5 minutes to 2 (rather than 10 minutes past 11). If scored as this error type, do not score as a misplacement error.
Hand Rotation	Time is shown in its mirror image. The hands are pointing to the correct locations but the relative lengths of the hands are reversed. That is the hands are drawn to indicate 5 minutes to 2 (rather than 10 minutes past 11). If scored as this error type, do not score as a misplacement error.
<u>Misplacements</u>	
Misplaced Number	The number of number locations not having numbers placed within half a centimeter either to the left or right of them are summed. Each empty number location is scored as one error. Do not score if scored as Number error.

Error type	Defining Criteria
Sequence	If the numbers are not in sequence from 1 to 12, an error is recorded.
Misplaced Hand(s)	Each hand, arising from center or near center, that does not fall within half a centimeter either to the left or the right of the correct number (regardless of number location) is scored as an error.
Misplaced Outside	The number of numbers placed entirely outside of the clock face is summed.
<u>Distortions</u>	
Used	The clock face is used to draw a recognizable feature other than a clock.
Horizontal/Vertical	The numbers on the clock face are arranged in a horizontal or vertical manner, regardless of how many numbers are used.
Circularity	The placement of the numbers is not guided by the circular outline of the clock face. That is, at least 2 or more numbers are placed more than 1 centimeter inside the circle.

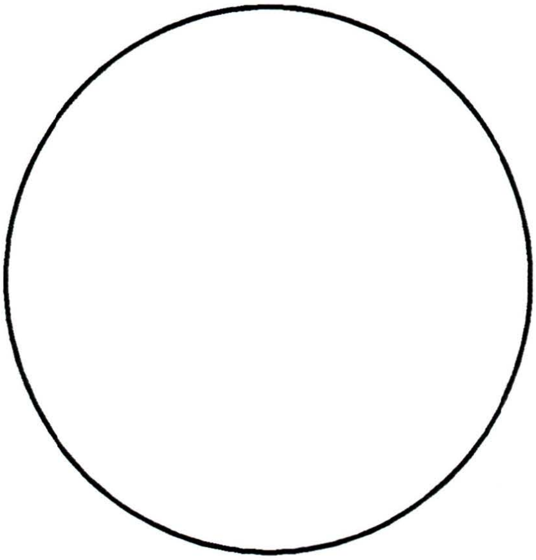
Error type	Defining Criteria
Substitutions	
Letters/Words	Letters or words are used in the place of numbers. This may take the form of numbers expressed as words or any other words or letters in the place of numbers. One error is scored for each distinct occurrence.
Scribble	One or more unintelligible scribbles are present in the place of numbers. One error is scored for each distinct occurrence.
Words	Words indicating the time are substituted for hands. For example the words 'Ten past Eleven' are written on the clock face. One error is scored for each hand replaced by a word.
Additional Number	Numbers other than 1 through 12 are present and are not in direct continuation after 12 (eg., 20, 25). One error is scored for each distinct occurrence.

Error type	Defining Criteria
Relative Lengths	No two hands, arising from the center or near center of the clock, differ in length by half a centimeter or more.
Additions	
Irrelevant Words	Irrelevant words or phrases, or apparent words even if illegible, are present. Only score when words are not in the place of numbers or hands.
Irrelevant Scribble	Irrelevant scribble or lines is present
Irrelevant Figure	A figure, in addition to the clock face is present.

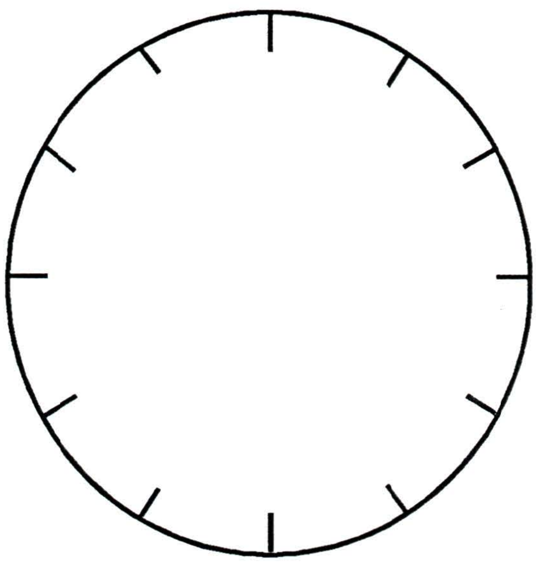
Note. From The Clock Test: Manual for administration and Scoring (pp. 8-12) by H. Tuokko, A. Horton, T. Hadjistavropoulos, & B. L. Beattie, (1990), University of British Columbia, BC. Reprinted by permission.

Appendix C

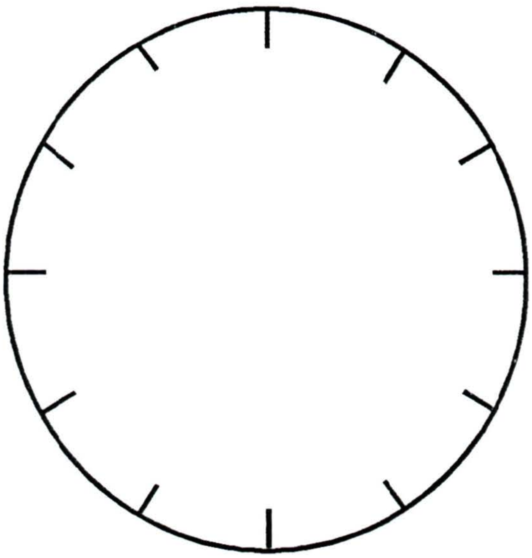
The Clock Test



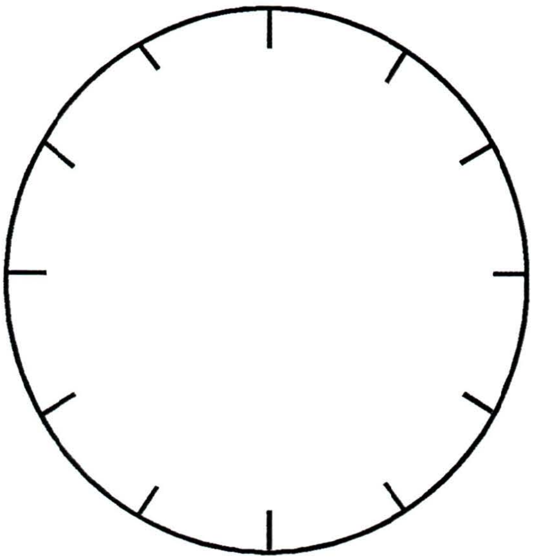
"Draw a Clock"



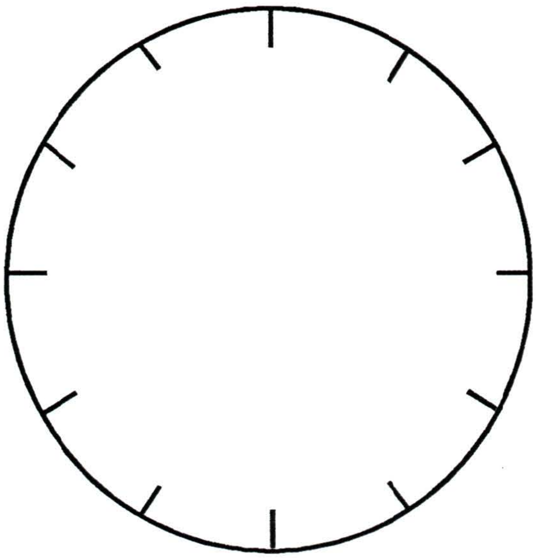
one o’clock



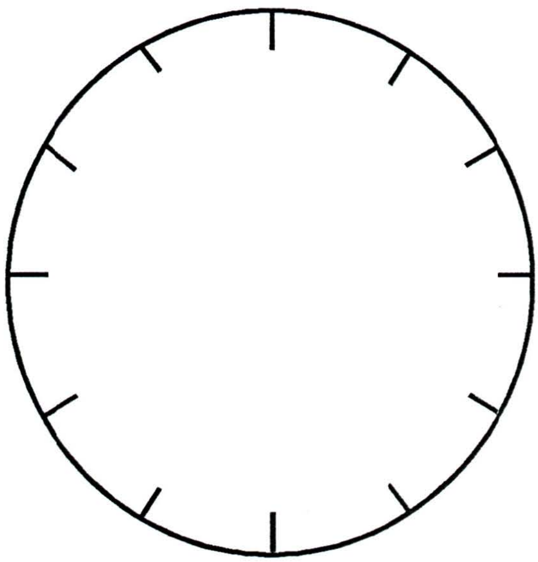
ten past eleven



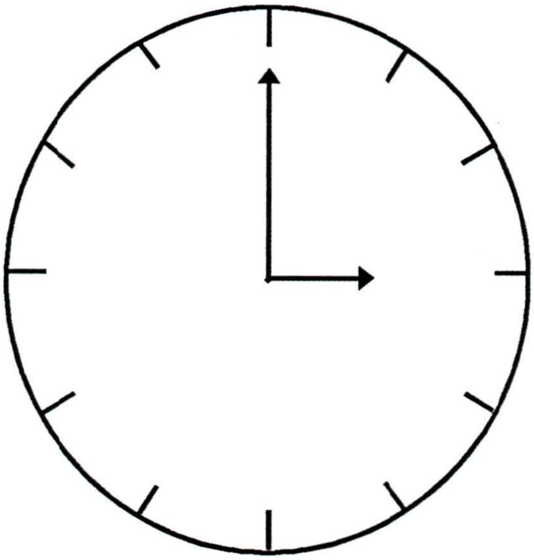
three o'clock



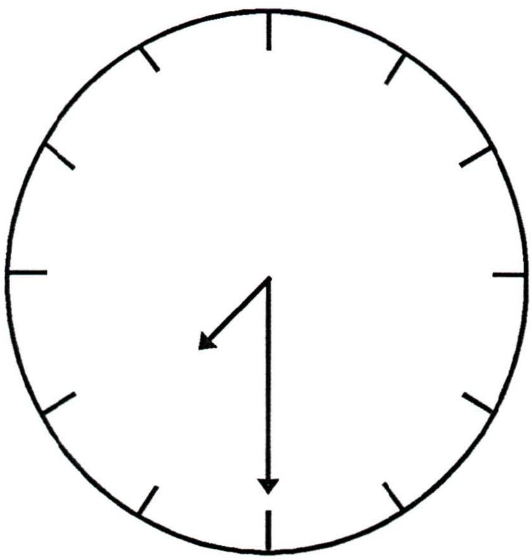
nine fifteen



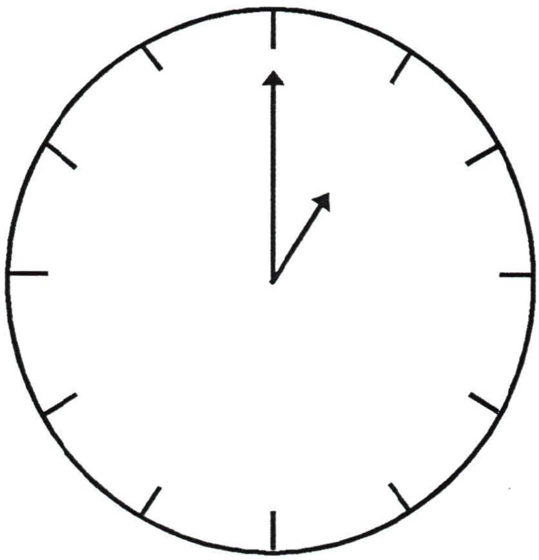
seven thirty



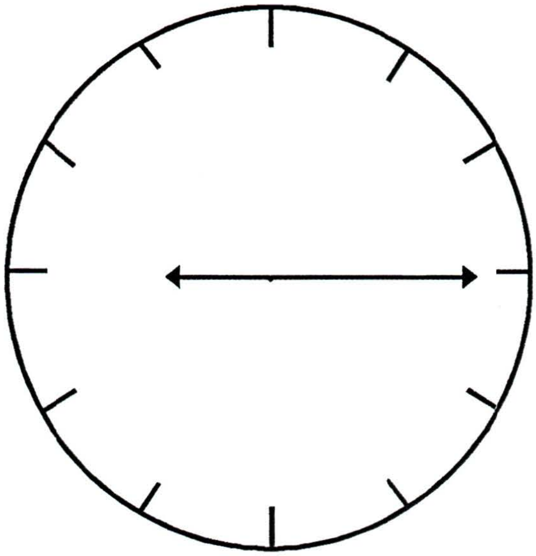
What time is it?



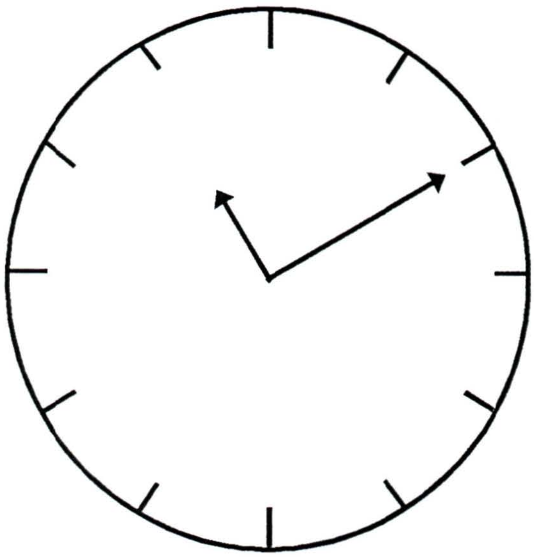
What time is it?



What time is it?



What time is it?



What time is it?

Note. From The Clock Test: Manual for administration and Scoring by H. Tuokko, A. Horton, T. Hadjistavropoulos, & B. L. Beattie, (1990), University of British Columbia, BC. Reprinted by permission.

CLOCK DRAWING RECORD FORM

Name: ______________________________ Age: _____

Sex: ________ Severity Rating: ___________

Diagnosis: ______________________________

Rater: ______________________________

Error Type	Number of Errors	Total Errors
Omissions		
Perseverations		
Rotations		
Misplacements		
Distortions		
Substitutions		
Additions		

Total Number of Errors: ____________

2. Clock Setting

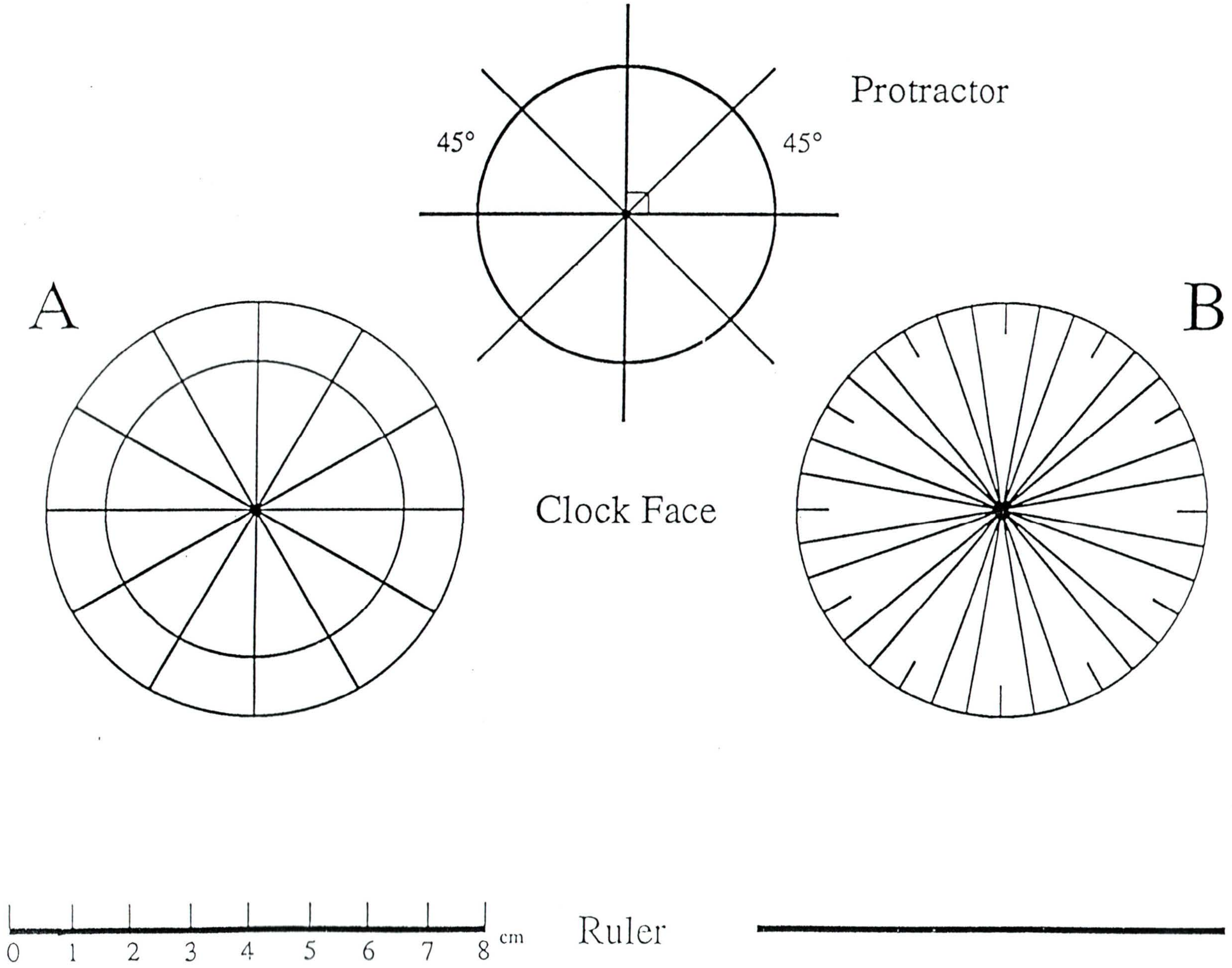
Clock Number	Score (max. = 3 points)
1	________________
2	________________
3	________________
4	________________
5	________________
Total (Max. = 15)	________________

2. Clock Reading

Clock Number	Score (max. = 3 points)
6	________________
7	________________
8	________________
9	________________
10	________________
Total (Max. = 15)	________________

Scoring Template for Clock Drawing

Scoring Template



Appendix D

Formulas used in the Present Study

Lambda Change Test.

$$\Lambda_D = \frac{\Lambda_2}{\Lambda_1}$$

Where Λ_2 is the lambda from the step with the larger number of redictors, and Λ_1 is the lambda from the step with the fewer number of predictors.

Significance of the Change in Lambda.

$$F(df_1, df_2) = \left(\frac{1 - y}{y}\right)\left(\frac{df_2}{df_1}\right)$$

Where, $y = \Lambda^{1/s}$

$$s = \sqrt{\frac{p^2(df_{effect})^2 - 4}{p^2 + (df_{effect})^2 - 5}}$$

Furthermore, $df_1 = p(df_{effect})$

$$df_2 = s\left[(df_{error}) - \frac{p - (df_{effect}) + 1}{2}\right] - \left[\frac{p(df_{effect}) - 2}{2}\right]$$

and, p = the number of dependent variables
df_{effect} = the number of groups minus 1, or k - 1
df_{error} = the number of groups times n - 1, where n is the number of cases per group.

Note. These formulas were taken from Using Multivariate Statistics (2nd ed.; pgs. 388, 514, & 546-547) by B. G. Tabachnick and L. S. Fidell, 1989, New York: Harper & Row.

Significance of the Difference Between Correlated Correlations.

$$t = \frac{(r_{xy} - r_{vy})\sqrt{(n-1)(1 + rxv)}}{\sqrt{2\left(\frac{n-1}{n-3}\right)|R| + \left(\frac{r_{xy} + r_{vy}}{2}\right)^2 (1 - r_{xv})^3}}$$

$$|R| = 1 - r^2_{xy} - r^2_{vy} - r^2_{xv} + 2r_{xy}r_{vy}r_{xy}$$

$$df = n - 3$$

Note. This formula was taken from Psychological statistics (4th ed.) by Q. McNemar, 1969, New York: Wiley.

Significance of the Difference Between Two Correlation Coefficients for Independent Samples.

Fishers z transformation of r.

$$z'_r = \tfrac{1}{2}[\ln(1 + r) - \ln(1 - r)]$$

$$z = \frac{z'_{r1} - z'_{r2}}{\sqrt{1 / (N_1 - 3) + 1 / N_2 - 3)}}$$

Furthermore, a value of 1.96 and 2.58 are required for significance at $p < .05$ and $p < .01$, respectively.

Note. This formula was taken from "Tests for comparing elements of a correlation matrix" by J. H. Steiger, 1980, Psychological Bulletin, 87, p. 245-251.

Tucker-Lewis Formula.

$$TL = \frac{(\chi^2_{null} / df_{null}) - (\chi^2_{equal} / df_{equal})}{(\chi^2_{null} / df_{null}) - 1}$$

Where, χ^2_{null} and df_{null} are for the null model (i.e., where all of the proposed factors, or correlations are equal to zero), and χ^2_{equal} and df_{equal} are for the actual model being tested. A TL coefficient in the .90s indicates an excellent representation of the data.

Note. This formula was taken from "Testing whether correlation matrices are different from each other" by J. A. Green, 1992, Developmental Psychology, 28, 215-225.

Table 1. Descriptive Information on the Sample

	Diagnostic Group					
Variable	Unlikely AD	Possible AD	Probable AD	Not Demented	Normal Subjects	Total
n =	21	105	154	111	65	456
Age[a]	70.43 (10.75)[b]	72.37 (6.60)	71.88 (7.50)	65.62 (11.55)	70.94 (8.35)	70.25 (8.95)
SES[a, c]	3.60 (2.91)	4.99 (3.20)	4.82 (2.99)	4.15 (2.82)	3.02 (2.58)	4.11 (2.90)
Education[a]	11.95 (3.65)	11.03 (2.83)	10.69 (2.95)	11.61 (3.09)	13.33 (3.72)	11.72 (3.25)
FRS Score[a]	25.90 (6.28)	25.75 (4.26)	26.99 (5.10)	16.56 (3.25)	8.00[d] (---)	20.64 (8.02)
Handedness[e]						
Right	19	99	145	104	62	429
Left	2	5	5	4	2	18

(table continues)

	Diagnostic Group					
Variable	Unlikely AD	Possible AD	Probable AD	Not Demented	Normal Subject	Total
Gender[e]						
Male	13	50	50	50	26	189
Female	8	55	104	61	39	267
Language[e]						
English	17	87	126	93	61	384
Other	3	17	28	14	4	68
Marital Status[e]						
Single	--	1	6	4	3	14
Married	15	81	103	86	36	321
Widowed	5	20	36	14	18	93
Separated	1	3	8	7	8	27

Notes.
[a] Group Mean.
[b] Brackets refer to the Standard Deviation.
[c] Lower SES numbers represent higher status.
[d] Normal subjects were not tested with the FRS. By default they received a score of 1 (healthy) for each of the eight FRS dimensions. Scores are summed across the eight dimensions.
[e] Number of subjects.