Two-Step Screening for Cognitive Impairment: The Clock Test and the 3MS

by

Megan Eleine O'Connell
B.A., University of Saskatchewan, 1998

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

MASTER OF ARTS

in the Department of Psychology

We accept this thesis as conforming
to the required standard

[redacted]

Dr. Holly Tuokko, Supervisor (Department of Psychology)

[redacted]

Dr. Roger Graves, Departmental Member (Department of Psychology)

[redacted]

Dr. Helena Kadlec, Departmental Member (Department of Psychology)

[redacted]

Dr. John Walsh, External Examiner (Faculty of Education)

Supervisor: Dr. Holly Tuokko

ABSTRACT

This project examined the efficacy of using a combination of complementary screening tools to detect cognitive impairment in persons over 65 years old. Using data from the Canadian Study of Health and Aging, we examined whether the addition of the Clock Test components of Clock Drawing, Clock Setting, and/or Clock Reading (Tuokko, Hadjistavropoulos, Miller, Horton, & Beattie, 1995) to the Modified Mini-Mental State Examination (3MS; Teng & Chui, 1987) resulted in increased sensitivity and/or specificity when identifying persons with cognitive impairment. The verification bias in the sample was adjusted. Then each screening tool was compared over all possible cut-off points. The results indicated that the use of the Clock Drawing, Clock Setting, and/or Clock Reading tasks in combination with the 3MS did not improve the identification of persons with and without cognitive impairment over the 3MS alone. Equivalent results were found for different age, education level, and sex groups.

Examiners:

[redacted]

Dr. Holly Tuokko, Supervisor (Department of Psychology)

[redacted]

Dr. Roger Graves, Departmental Member (Department of Psychology)

[redacted]

Dr. Helena Kadlec, Departmental Member (Department of Psychology)

[redacted]

Dr. John Walsh, External Examiner (Faculty of Education)

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor and my committee members for their support and help with this thesis. In particular, I would like to thank Dr. Holly Tuokko, my supervisor, for her help and support during my time as a student at the University of Victoria, and for her helpful suggestions and comments on my thesis. I would also like to thank Dr. Roger Graves for his innovative ideas and helpful suggestions. Finally, I would like to thank Dr. Helena Kadlec for all of her additional mathematical help and helpful suggestions. I extend a special 'thank you' to Helena for her patience and encouragement, and for making me feel more comfortable and confident about mathematics and statistics.

I would also like to thank Jackie Bush, a good friend, for her generous support. In addition, I would like to thank Jackie for her diligent proofreading of both my proposal and my final thesis. I also thank my family. In particular, I thank my mother, Flo Broten, for her judicious proofreading of my thesis and for her support. I thank Cleo for always being there, and I thank my husband, David Dahlem, for all of his support, encouragement, understanding, and love. Finally, I thank David for writing an exceptional computer program, without which I would not have been able to complete this thesis.

Two-Step Screening for Cognitive Impairment: The Clock Test and the 3MS

CHAPTER 1

Introduction

Aging and age-related health care issues are becoming increasingly important as the population gets older. In 1999, Canadians 65 years of age and older comprised 12% of the population (Statistics Canada, 1999a), and the number of older Canadians is expected to increase; 16.5 % of the population by 2016 and 21.4% of the population by 2026 will be over 65 years of age (Statistics Canada, 1999b). This changing age distribution raises a number of serious questions relevant to geriatric research. For example, in Canada in 1991, one quarter of a million people 65 years of age and older suffered from dementia, a degenerative syndrome that involves cognitive impairment in two or more domains and interferes with daily activities of living. The probability of suffering from dementia rises with age; for example, 2.4% of those 65-74 years old, 11.1% of those 75-84 years old, and 34.5% of those 85 years old and over have dementia (Canadian Study of Health and Aging Working Group, 1994). Given this finding and the continued growth of the oldest segments of the population, issues concerning this vulnerable group are particularly relevant.

Accurate detection of cognitive impairment and dementia has implications for clinical management of patients and their families and can result in more efficient and effective use of limited health care resources. Early, accurate detection of dementia is of particular concern because current pharmacological treatments for dementia are recommended as an early intervention. Although current pharmacological treatments for dementia are not useful for many people, some individuals demonstrate improvements or stability in their cognitive abilities, and these benefits are best realized with early intervention (van Reekum, Simard, & Farcnik, 1999). Consequently, the earlier that dementia is diagnosed, the earlier that intervention can be contemplated. Identifying persons with dementia, however, is costly and time-consuming. Screening tools that identify cognitive impairment are a cost-effective means of identifying persons who require the time-consuming intensive neurological and neuropsychological assessment in order to diagnose dementia. This study examines the effectiveness of using a combination of complementary screening tools for detecting dementia in persons over 65 years of age in an attempt to increase the accuracy of screening for dementia. I had hypothesized that the addition of components of the Clock Test (Tuokko, Hadjistavropoulos, Miller, Horton, & Beattie, 1995) to the Modified-Mini Mental Status (3MS) examination (Teng & Chui, 1987) would result in a more accurate screening tool for cognitive impairment.

Chapter 2 begins with an overview of the purpose of screening and discusses characteristics of screening tools, such as their ability to distinguish diseased persons from non-diseased persons. The psychometric issues relevant to screening tools for cognitive impairment due to dementia are discussed. Then a review of the psychometric properties of screening tools for cognitive impairment is provided. Mental status exams and clock tests, with particular emphasis on the 3MS (Teng & Chui, 1987) and the Clock Test (Tuokko et al., 1995) are the screening tools reviewed. The rationale for using the components of the Clock Test (i.e., Clock Drawing, Clock Setting, and Clock Reading) in combination with the 3MS is discussed. Finally, the research hypotheses are articulated.

Chapter 3 describes the methods used to investigate the efficacy of combining the 3MS (Teng & Chui, 1987) and the components of the Clock Test (Tuokko et al., 1995). The characteristics of the participants from the Canadian Study of Health and Aging – 1 (CSHA – 1) are described. The procedure involved first, the investigation of the influence of age, education, and sex on the 3MS. Then, the 3MS was adjusted for the influences of age, education, and sex, and the accuracy of the adjusted-3MS was compared to the original 3MS. Next, the verification bias was corrected and a graph that represented the hit rate of the screening tool (corrected and non-corrected) was constructed. Then, the two areas under the curves of these graphs were compared as this comparison is a determination of which screening tool (i.e., corrected or non-corrected) was better at distinguishing cognitively impaired from cognitively normal persons. Finally, because the two screening tools were given to the same individuals, the correlation between the two tests was accounted for before comparing the areas.

The 3MS was combined separately with each component (i.e., Clock Reading, Clock Setting, and Clock Drawing) to determine whether the combined screening tools were more accurate than any screening tool alone. In addition, the 3MS was combined with the total Clock Test score. Finally, the 3MS was combined with Clock Drawing and Clock Setting based on the results of logistic regression, which suggested that Clock Drawing and Clock Setting contributed unique prediction above the 3MS alone. The methods for comparing the accuracy was performed as described above. Finally, the ability of the combined (i.e., Clock Drawing/3MS, Clock Setting/3MS, and Clock Reading/3MS) versus single screening tools to identify cognitive impairment was investigated for different age, education, and sex groups to see if the combination of complementary screening tools would aid in the detection of cognitive impairment for particular demographic groups.

Chapter 4 presents the results of the analyses. To summarize, I found that the attempt to improve the accuracy of the 3MS by adjusting for the effects of age, education, and sex was unsuccessful: the original 3MS was more accurate than the adjusted-3MS at discriminating between participants with and without cognitive impairment. Further, adding each component of the Clock Test to the 3MS did not increase the accuracy above the 3MS alone. Due to the relatively low sensitivity and specificity of the components of the Clock Test, any addition of the 3MS to the Clock Test resulted in increased accuracy over each component of the Clock Test alone. When assessed separately for different demographic groups, the results were similar: the addition of single components of the Clock Test to the 3MS resulted in equivalent accuracy to the 3MS alone. Finally, for each age, education, and sex group, any addition of the 3MS to the Clock Test resulted in increased accuracy over each component of the Clock Test alone.

Chapter 5 provides a summary and integration of the results, and discusses the consistency of these results with the hypotheses, which were presented in Chapter 2. In addition, the limitations of the current procedure and their impact on some of the findings are discussed. Suggestions for future research are presented. Subsequently, the clinical conclusions that can be drawn from this study are discussed.

CHAPTER 2

Literature Review

The present chapter reviews the purpose of screening and the characteristics of screening tools which are relevant to understanding the literature on screening tools for cognitive impairment. In addition, the present chapter discusses psychometric properties specifically related to screening for cognitive impairment, followed by a literature review of screening for cognitive impairment using mental-status assessment tools, with specific attention to the 3MS (Teng & Chui, 1987). The rationale and success of screening for cognitive impairment with clock tasks will be discussed. In addition, properties of the Clock Test (Tuokko et al., 1995), a screening tool for cognitive impairment will be discussed, followed by the hypotheses of this study.

Purpose of Screening

To save time and money, screening tests are used to identify persons who require further testing. Persons who screen positive on the screening tool are referred for a diagnostic investigation, whereas persons who screen negative are not referred (Ferrer, 1968; Wilson & Jungner, 1968). Screening tools do not constitute a diagnosis, but the screening procedure is helpful in that it minimizes the number of persons who require a full medical and psychological investigation, which is a more costly process than the screening procedure (Essex-Sorlie, 1995). In order to be effective, screening tools should be accurate (Zweig & Campbell, 1993). In addition, screening tools should be efficacious or have practical value. To be efficacious, screening tools should be easy and quick to administer, inexpensive, and acceptable to health professionals and the public (Ferrer, 1968; Zweig & Campbell, 1993). The accuracy of a screening tool is essential, but accuracy alone is not sufficient to ensure a screening tool is effective and efficacious (Zweig & Campbell, 1993). This study investigates the accuracy (i.e., effectiveness) of using a combination of two efficacious screening tools.

Screening for Cognitive Impairment

Screening for dementia is important for secondary prevention and planning of health care services (Ganguli, 1997). Dementia is a degenerative syndrome that involves cognitive impairment in two or more domains. In addition, for a diagnosis of dementia, the degree of cognitive impairment must be severe enough to have interfered with the daily activities of living. A positive result on a screening tool for cognitive impairment does not constitute a diagnosis of dementia. Screening tools only help clinicians decide whether an individual's level of cognitive functioning warrants further investigation (Baker, 1989).

One of the difficulties with using cognitive impairment as an indication of the dementing process is differentiating cognitive change due to age from pathological cognitive impairment (Ritchie, 1988). It is difficult to determine when cognitive change constitutes cognitive impairment. For instance, a clinician may assume that an individual is impaired if they are unable to perform an 'easy' task, which they would normally be expected to perform with little effort (Tuokko & Hadjistavropoulos, 1998); however, many persons with mild cognitive impairment will likely be able to perform these 'easy' tasks.

Many screening tools for cognitive impairment include only 'easy' items that measure extreme impairment (Ritchie, 1988). Most people are able to perform each task with relative ease and very few people are not able to perform most of the tasks (i.e., the distribution is negatively skewed and demonstrates a ceiling effect). Consequently, most screening tools for cognitive impairment are not able to identify persons with mild cognitive impairment (Mitrushina & Fuld, 1996; Tuokko & Hadjistavropoulos, 1998). The inclusion of items that are more difficult in an attempt to identify persons with mild cognitive impairment, however, is problematic if the items are too difficult and persons with intact cognitive abilities are not able to correctly perform the tasks (Ritchie, 1988).

Despite the problems identifying mild cognitive impairment, screening tools for cognitive impairment are popular because they have standardized the way clinicians assess for cognitive difficulties (Baker, 1989) and have quantified clinical observations (Mitrushina & Fuld, 1996). In addition, screening tools for cognitive impairment are brief and not overly demanding for patients. An ideal screening tool for cognitive impairment can be easily administered by trained paraprofessionals and demonstrates few practice effects (Mitrushina & Fuld, 1996).

Characteristics of Screening Tools

Sensitivity and Specificity

The sensitivity of the screening tool is its ability to produce a positive result when the person has the disease (i.e., the conditional probability that a person with the disease has a positive result on the screen). Conversely, the screening tool should indicate a negative result when the individual does not have the disease (i.e., the conditional probability that a person without the disease has a negative result on the screen). The measure of the ability of the screening tool to correctly identify disease free persons is the specificity of the screening tool (Ferrer, 1968; Wilson & Jungner, 1968). The sensitivity and specificity of a screening tool are assessed by applying the screening tool and the diagnostic procedure to the same sample of individuals, referred to as the reference population. The accuracy (i.e., sensitivity and specificity) of the screening tool is determined by its ability to distinguish diseased from non-diseased persons in this sample (Essex-Sorlie, 1995); thus, the estimates of the sensitivity and specificity of the screening tool are dependent on the characteristics of the reference population. These characteristics include the prevalence of the disease in the reference population and the degree of impairment of diseased individuals in the reference population (Ritchie, 1988).

The accuracy of a screening tool is its ability to distinguish between diseased and non-diseased states, which are binary or dichotomous (Wilson & Jungner, 1968). However, the scores obtained with most screening tools are continuous for diseased and non-diseased persons. The cut-off point of the screening tool arbitrarily separates normal (i.e., a negative result) from abnormal (i.e., a positive result) for a continuous screening tool. A cut-off point on a continuous screening tool determines the sensitivity and specificity for that screening tool at that cut-off point (Essex-Sorlie, 1995). It is important to remember, however, that cut-off scores are arbitrary, are based on clinical judgement, and are externally imposed (i.e., not a property of the test; Dwyer, 1996). The costs and benefits of a highly specific or highly sensitive screening tool must be weighed in order to determine the cut-off point (Essex-Sorlie, 1995). Assuming that higher scores on the screening tool are normal, or indicative of non-diseased states, if the cut-off point is lowered, the test will become more sensitive but less specific. This is because sensitivity and specificity are inversely related; as one is raised, the other is lowered (Zweig & Campbell, 1993).

Deciding how to balance the sensitivity and specificity of the screening tool is important because there is usually overlap between diseased and non-diseased persons' performance on the screening tool; therefore, the sensitivity and specificity are not perfectly accurate at distinguishing diseased from non-diseased individuals. In other words, false-positive and false-negative results occur. The false-positive proportion is the probability that a non-diseased patient will have a positive result on the screening tool. The false-negative ratio is the probability that a diseased patient will have a negative result on the screening tool (Essex-Sorlie, 1995). The sensitivity of a screening tool is the complement of a measure of the false-negative proportion (i.e., false-negative = 1 – sensitivity), and the specificity is the complement of the false-positive proportion (i.e., false-positive = 1 – specificity; Essex-Sorlie, 1995; Wilson & Jungner, 1968).

False-positives will be identified by further diagnostic investigation, but the individuals who have a false-negative result will be 'missed.' Therefore, it is important for the false-negative rate to be low (Ferrer, 1968; Wilson & Jungner, 1968). However, high numbers of false-positives increase the burden on the health care system through unnecessary investigations of persons who are disease-free (Ferrer, 1968). In addition, the cost of false-positives includes the psychological harm for the patient (Malm, 1999). A good screening tool minimizes error (i.e., false-positive and -negative results) by maximizing both sensitivity and specificity.

Predictive Value

Sensitivity and specificity are based on the knowledge of the disease status, which is identified through further diagnostic investigation or verification. Other useful knowledge about a screening tool is the probability that a person has the disease given a positive or negative result on the screening tool (Essex-Sorlie, 1995). The predictive value of a screening tool is related to the sensitivity and specificity of the screening tool. Predictive value is of greater clinical utility than the sensitivity and specificity alone, because it helps the clinician determine the likelihood that the individual being screened has the disease for which the individual is being screened, given the result on the screening tool. The sensitivity and specificity only determine the accuracy of the screening tool, but the predictive value determines the usefulness of a screening tool (Zweig & Campbell, 1993). The positive predictive value of a screening tool is the probability that a patient with a positive result on the screen actually has the disease (i.e., is a true-positive). The negative predictive value is the probability that a patient with a negative result does not have the disease (i.e., is a true-negative). Predictive values are extremely sensitive to the prevalence of the disease in the reference population, which limits their utility (i.e., they are not readily generalizable). In addition, predictive values are only useful when the prevalence of the disease in the population to be screened is similar to the reference population (Essex-Sorlie, 1995). Predictive values are rarely reported in the literature on screening tools, partially because of the predictive value's limited utility and partially because diseases that are rare will have low predictive values even if the sensitivity and specificity are very high (Zweig & Campbell, 1993). For example, screening tools for dementia, a disease with a low prevalence in the population will have low predictive values.

Reliability and Validity of Screening Tools for Cognitive Impairment

The reliability of a screening tool is an estimate of the extent to which changes in performance reflect true differences in behavior (Retzlaff & Gibertini, 1994). Due to the degenerative nature of dementia, cognitive impairment will increase with time, and clinicians want to ensure that the differences in test scores over time reflect true differences and not the instability of the measure (Retzlaff & Gibertini, 1994). Therefore, the stability of a screening tool for cognitive impairment is important. In addition to using a highly reliable screening tool for cognitive impairment, clinicians should administer and score the tests in a standardized or a uniform way to minimize measurement error (Tuokko & Hadjistavropoulos, 1998), thus, ensuring higher reliability (Allen & Yen, 1979).

The validity of a screening tool for cognitive impairment refers to the test's utility: does the screening tool measure cognitive impairment associated with dementia? Validity is assessed in different ways. Criterion-related validity assesses the performance of the screening tool in relation to other measures (Retzlaff & Gibertini, 1994). Criterion-related validity of screening tools for cognitive impairment can be obtained by comparing the results of the screening tool with the gold standard diagnosis of dementia (Ritchie, 1988). For example, at present, the only way to confirm a diagnosis of Alzheimer's dementia, the most common dementia, is by autopsy. Consequently, the current gold standard for the diagnosis of Alzheimer's Disease is a consensus diagnosis after medical, neurological, and neuropsychological assessments (McKhann et al., 1984). Sensitivity and specificity are measures of the criterion-related validity of a screening tool (Tuokko & Hadjistavropoulos, 1998).

When assessing the criterion-related validity of a screening tool for cognitive impairment, the scores on the screen should not influence the diagnosis. If the scores do influence the diagnosis, criterion contamination has occurred and the results of the assessment will be spuriously high (Tuokko & Hadjistavropoulos, 1998). Many screening tools for cognitive impairment are very similar to the mental status interviews regularly performed by clinicians. The examination of mental status was traditionally a crucial portion of the diagnostic interview to diagnose dementia. The mental status screening tool may contain many of the same items as the diagnostic criteria, which may potentially lead to inflated estimates of the criterion validity (Ritchie, 1988).

In addition, the reference population selected for assessing the criterion-related validity of a screening tool is important. The reference population should resemble the population where the screening tool will be used; for example, the reference population might include individuals of all stages of dementia and include persons with concomitant disorders (Ritchie, 1988), depending on the setting where the screening tool will be used.

Validity is lessened when factors other than cognitive impairment influence the score an individual obtains on the screening tool (Retzlaff & Gibertini, 1994). For instance, the validity of screening tool scores are negatively affected by insufficient visual and auditory acuity (Paolo, 1998). Age-related factors can influence scores on a screening test for cognitive impairment; therefore, it is important to compare the results on the screening tools with equivalent age groups because the scores may decline with age (Ganguli, 1997). Consequently, older persons may appear impaired when compared to younger persons whose performance would be expected to be much higher. In addition, some tasks may depend on the level of education attained by the person being screened (Ritchie, 1988).

Screening Tools for Cognitive Impairment

Modified Mini-Mental Status (3MS) Examination

The Mini-Mental State Examination (MMSE) is a screening tool that was designed to be a quick and easy to administer quantitative assessment of cognitive functioning. Items include questions about orientation, memory, attention, naming ability, ability to follow written and verbal commands, and the ability to draw a figure and write a sentence, and are scored out of a total of 30 points (Folstein, Folstein, & McHugh, 1975).

The Modified Mini-Mental State (3MS) examination (Teng & Chui, 1987) was derived from the MMSE in an attempt to decrease the floor and increase the ceiling to aid in the identification of individuals with cognitive impairment. The 3MS allows for graded scoring on previously dichotomously scored responses on the MMSE. The scores range from 0-100 and assess the same aspects of cognitive functioning as the MMSE (i.e., orientation, memory, attention, naming ability, the ability to follow written, and verbal commands, the ability to draw a figure, and the ability to write a sentence). The 3MS includes more items than the MMSE and assesses temporal and spatial orientation, the ability to see relationships between objects, and verbal fluency. In addition, the 3MS includes more items that assess different aspects of memory, including cued recall, recognition memory, and delayed free and cued recall, and delayed recognition memory. (Teng & Chui, 1987).

Factor analytic study of the 3MS with a sample of nursing home residents between the ages of 71 – 97 years (*M* age = 84.4, *SD* = 6.1) indicated that the 3MS measures the following five domains: psychomotor skills, memory, identification and association, orientation, and, concentration and calculation (Abraham et al., 1993). Items included in the psychomotor skills domain were the ability to read and obey a command, to write a sentence, to copy pentagons, and to follow a three-stage command. Recall of three words after a delay (e.g., first recall after one intervening task, and second recall after 12 intervening tasks) and naming of four-legged animals comprised the memory factor. The identification and association factor was comprised of naming, the ability to see how two things are similar, and the ability to repeat sentences. The orientation factor was composed of orientation to birth, time, and space. Finally, the concentration and calculation factor included the ability to register new words and the ability to count backward and to spell backward (i.e., mental reversal; Abraham et al., 1993). It appears from Abraham et al.'s factor analysis, that the 3MS samples a wide range of cognitive domains. In addition, the 3MS had moderate to high correlations with neuropsychological tests that assess similar domains, which was evidence of its construct validity (Grace et al., 1995).

Tombaugh, McDowell, Kristjansson, and Hubley (1996) compared the sensitivity and the specificity (i.e., criterion validity) of the 3MS and the MMSE using samples of individuals with Alzheimer's disease and normal individuals (i.e., with no cognitive impairment) from CSHA-1, found that the sensitivity of the 3MS and MMSE were similar (with cut-off scores of 77 and 18, respectively). The sensitivity of the 3MS was 92.6% whereas the sensitivity for the MMSE was 90.5%. The difference was not significant at these cut-off points, nor was the difference significant when all possible cut-off points were examined. In both the MMSE and the 3MS the scores varied according to age and education level for individuals without cognitive impairment (scores for individuals with Alzheimer's Disease were not correlated with age and education); therefore, the authors did not supply an overall estimate for the specificity. They found that for younger persons (i.e., ages 65-79 years) with nine or more years of education, the specificity of the MMSE (i.e., 90%) was slightly larger than that for the 3MS (i.e., 81%). Similarly, for older individuals (i.e., ages 80-89 years) with nine or more years of education, the specificity of the MMSE (i.e., 90%) was larger than that for the 3MS (i.e., 72%). The specificity for younger persons with eight or fewer years of education was 74% with the MMSE and 57% with the 3MS. For older individuals with eight or fewer years of education the specificity of the MMSE was 51% and the specificity of the 3MS was 37%. Although the MMSE appeared to have higher specificity than the 3MS, when Tombaugh et al. examined the sensitivity and specificity over all cut-off points for the whole sample overall, they did not find that the 3MS and MMSE differed significantly. However, there appeared to be a trend suggesting that the 3MS was more accurate (i.e., had higher sensitivity and specificity) overall than the MMSE.

The above study by Tombaugh and colleagues (1996) has some methodological errors. For example, there was a verification bias in the sample. Although all of the individuals who were identified by the screening tool as positive for cognitive impairment were asked to participate in a neurological, medical, and neuropsychological investigation to diagnose the aetiology of the cognitive impairment, only a portion of those who were negative on the screen were investigated and had their diagnosis verified.

This procedure leads to a bias in the estimates of the sensitivity and specificity of the screening tool, such that the sensitivity is overestimated and the specificity is underestimated (Zweig & Campbell, 1993). In addition, the methods used by Tombaugh et al. (1996) to compare the two screening tools over all cut-off points (i.e., by comparing the areas under the curves of receiver operating characteristic plots; Zweig & Campbell, 1993) did not account for the fact that the scores from the two screening tools were correlated. The two scores were correlated because both the 3MS and the MMSE scores were obtained from the same persons. The statistical tests that account for this covariance are more sensitive. Consequently, Tombaugh et al. (1996) may have found a significant difference if they had accounted for the covariance between the 3MS and the MMSE. In light of these methodological problems, the results of the Tombaugh et al. study should be interpreted with caution.

A study by McDowell, Kristjansson, Hill, and Hebert (1997), compared the 3MS and the MMSE and found the 3MS to be superior to the MMSE because of the 3MS's extended scoring system. The study by McDowell et al. used a different sample from CSHA-1 than the previous study by Tombaugh et al. (1996). In particular, the McDowell et al (1996) study consisted of individuals who either spoke French or English. McDowell et al. performed analyses for identifying cognitive impairment with the following two groups: those with cognitive impairment (i.e., participants with dementia and cognitive impairment but not demented) versus normal participants. In addition, they looked at the two screening tools' ability to identify persons with and without dementia with the following two diagnostic groups: participants with dementia versus those without dementia (i.e., participants with cognitive impairment but not dementia and normal participants). After correcting for the verification bias, McDowell et al. compared the sensitivities and specificities of the 3MS and MMSE over all possible cut-off points (by comparing the areas under the receiver operating characteristic plots; Zweig & Campbell, 1993). McDowell et al. (1997) used procedures that accounted for the fact that the 3MS and MMSE scores were correlated because both tests were administered to the same sample. They found that the 3MS was superior to the MMSE at almost all cut-off points. Further, the 3MS was better than the MMSE at correctly classifying individuals with and without cognitive impairment when the sensitivity of the test was more highly weighted than the specificity of the test.

Other researchers have compared the criterion validity assessed by sensitivity of the MMSE and the 3MS in different populations. For example, the ability of the MMSE and the 3MS to differentiate between diagnoses in a general psychiatric population (e.g., organic mental disorder, mood disorder, schizophrenia, anxiety disorder, and a personality disorder) was assessed. The results indicated that although the sensitivity was low for both tests, the 3MS was superior to the MMSE (Lamarre & Patten, 1991). In addition, the sensitivity of the 3MS was found to be superior to that of the MMSE when identifying persons with cognitive impairment due to stroke. Both screening tools, however, were found to be insensitive to the localisation of deficits (Grace et al., 1995).

The above data suggest that the criterion-related validity of the 3MS is superior to the MMSE. However, the unique strength of the 3MS may be in its superior reliability. Test-retest stability assessed after an interval of over 19 days in a sample of stroke patients indicated that the correlations for the two test administrations was 0.80 for the 3MS and 0.71 for the MMSE (Grace et al., 1995). Similarly, the stability of the 3MS (*r* = 0.87) was found to be superior to the MMSE (*r* = 0.78) when tested over various time intervals (i.e., 1 – 419 days) in a sample of English-speaking community-dwelling older persons with and without cognitive impairment. French language versions of the two screening tools indicated a similar pattern, with the 3MS (*r* = 0.81) demonstrating superior stability over the MMSE (*r* = 0.75) in French-speaking older persons with and without cognitive impairment (Bravo & Hebert, 1997b). Stability of the 3MS assessed over an interval of 36 hours in a sample of general psychiatric patients found that the Kappa statistic indicated perfect agreement (i.e., Kappa value was 1.0) between the two administrations of the test (Lamarre & Patten, 1991). These data suggest that the reliability of the 3MS, as measured by test-retest reliability, is high.

Other measures of the reliability of the 3MS are also high. For example, the internal consistency of the 3MS as measured by split-half correlations was 0.82 and Cronbach's alpha (which assesses the interrelationships between all of the items on the test) was 0.87 in a sample of English-speaking and French-speaking older persons with and without cognitive impairment (McDowell et al., 1997). Additionally, the inter-rater reliability of the 3MS, as assessed by a Kappa statistic, was moderate at K = 0.67 (Lamarre & Patten, 1991).

Limitations of the 3MS

Although the reliability and validity of the 3MS appear to be adequate, the 3MS does have some limitations. The 3MS is affected by age, level of education (Bravo & Hebert, 1997a; Tombaugh et al., 1996), and sex (Bravo & Hebert, 1997a).

Bravo and Hebert (1997a) found that age and education accounted for 22% (i.e., with simultaneous multiple regression, $R^2 = 0.22$) of the variance in 3MS scores using data from the CSHA-1. In addition, Bravo and Hebert found that age, education, sex, and language of testing accounted for 27% of the variance in 3MS scores. The influence of these demographic characteristics limits the validity of the 3MS when screening for cognitive impairment. For example, both fewer years of education and increased age reduced the specificity of the 3MS (Tombaugh et al., 1996).

Researchers have attempted to find other tools to screen for cognitive impairment that are not as influenced by the demographic characteristics of the individual being screened. One such proposed tool is a clock drawing task. Some clock drawing tasks are apparently not affected by age (Cahn & Kaplan, 1997) nor education level (Sunderland et al., 1989).

In addition to being affected by the age and education of patients, the MMSE and 3MS have other limitations that may be improved with the use of clock drawing tasks. For example, The MMSE and the 3MS are screening tools that assess cognitive impairment in domains which could be characterized as more 'cortical' (Royall & Polk, 1998), such as concentration and calculation, psychomotor skills, memory, identification and association, and orientation (Abraham et al., 1993). However, the MMSE and 3MS offer little in terms of screening for impairments in executive function. Consequently, cognitive impairment that can be characterized as 'sub- cortical' will likely be under identified with sole reliance on a screening tool like the MMSE or 3MS (Royall & Polk, 1998). The ability to correctly draw a clock after a verbal request requires intact auditory comprehension, a memory of a spatial representation of a clock, intact visual perception, intact attentional processes, and an intact executive system to oversee motor output (Freedman et al., 1994). In addition, clock drawing assesses constructional apraxia, a parietal lobe function (Critchley, 1966). The cognitive complexity in clock drawing makes the clock drawing task ideal for screening for a wide variety of cognitive impairments (Libon, Swenson, Barnoski, & Sands, 1993; Royall & Polk, 1998).

Finally, the clock drawing task may be particularly sensitive to Alzheimer's disease, the most common form of dementia, because clock drawing may assess parietal lobe functions (Critchley, 1966). In addition, research suggests that patients with Alzheimer's disease may have an impaired mental representation of a clock (Rouleau, Salmon, & Butters, 1996) or, possibly, a deficit in the semantic representations of a clock (Rouleau, Salmon, Butters, Kennedy, & McGuire, 1992). Consequently, the addition of clock drawing to the 3MS might offer additional discriminability of persons with and without cognitive impairment.

Clock Drawing

There are many different versions of clock drawing tasks, which are commonly used in clinical practice. Typically, a patient is asked to draw the face of a clock, usually within a pre-drawn circle. In some versions of the clock test, the patients are asked to draw the circle, however, this has disadvantages if the circle is drawn too small to accommodate all of the numbers from the clock face (Freedman et al., 1994).

In addition, some versions of the clock drawing task ask patients to draw the clock hands to indicate a specific time. Successful completion of this task may depend on an intact memory if the command to set the time is given before the clock face is drawn. Asking the patient to set the clock to different times may involve different processes. Asking the patient to set the hands of the clock to indicate 10 minutes after 11 o'clock requires the patient to put hands on either side (i.e., left and right visual fields) of the clock face; consequently, clock drawing may be sensitive to hemi-neglect. In addition, asking the patient to set the clock places a great demand on executive functions. For example, in order to set the time to 10 minutes past 11 o'clock the '10 after' must be recoded as '2' on the clock face. Further, persons with dementia may be more likely to make stimulus bound errors when setting the time to 10 minutes after 11 o'clock and be 'captured' by the '10' on the clock face next to the '11'; consequently, they may set the time to 10 minutes to 11 (Freedman et al., 1994).

Different researchers and clinicians use different versions of clock drawing. Some researchers present the pre-drawn circle and some do not. Some researchers require that the hands be drawn to represent specific times, although no one time is universally used. In addition to different versions of the task, there are different scoring techniques for the clock drawing task. Scoring techniques include, the clinician's ratings of clock drawings (Cahn-Weiner et al, 1999; Sunderland et al., 1989), comparison of the clock drawn by the patient and a template (Wolf-Klein, Silverstone, Levy, Brod, & Breuer, 1989), and deciding where on the clock face an error occurs (Watson, Arfken, & Birge, 1993). Other scoring methods assess the qualitative errors made while drawing the clock face and the hands of the clock (Kozora & Cullum, 1994; Shulman, Shedletsky, & Silver, 1986). In addition, some methods of scoring quantitatively assess the quality of the drawing and qualitatively assess the errors (e.g., Rouleau et al., 1992; Tuokko, Hadjistavropoulos, Miller, & Beattie, 1992; Tuokko et al., 1995).

Various versions of the clock drawing task are useful for screening for persons with cognitive impairment due to dementia (e.g., Cahn et al., 1996; Lam et al., 1998; Shulman et al., 1986; Tuokko et al., 1992; Wolf-Klein et al., 1989). The estimated sensitivity of the clock drawing task when screening for dementia ranges from 83% to 92% and the specificity ranges from 72% to 92.7% (Cahn et al., 1996; Lam et al., 1998; Shulman et al., 1986; Tuokko et al., 1992; Wolf-Klein et al., 1989). The wide range in the estimated sensitivity and specificity of the clock drawing task is due, in part, to the different scoring techniques used in each of these studies. One study, which compared the Shulman et al. (1986), Tuokko et al. (1992), Watson et al. (1993) and Wolf-Klein et al. (1989) scoring procedures, found that the Shulman et al. (1986) and Tuokko et al. (1992) techniques had superior sensitivity, the Wolf-Klein et al. (1989) technique had superior specificity, and the Watson et al. (1993) technique had inferior sensitivity and specificity (Tuokko, Hadjistavropoulos, Rae, & O'Rourke, 2000).

It has been suggested that the clock drawing task may be useful for differentiating between Alzheimer's disease and multi-infarct/vascular dementia (Libon, Malamut, Swenson, Sands, & Cloud, 1996; Wolf-Klein et al., 1989), and Alzheimer's disease and the dementia due to Huntington's disease (Rouleau et al., 1992). However, other studies have not found the clock drawing task useful for differentiating between types of dementia (Lam et al., 1998). Consequently, it is unclear whether clock drawing is sensitive to different types of cognitive impairment due to different dementias.

Clock drawing has been correlated with other screening tools for cognitive impairment. Specifically, the clock drawing task and the MMSE are highly correlated (Juby, 1999), clock drawing performance can predict MMSE scores (Death, Douglas, & Kenny, 1993), and clock drawing is useful in identifying cases of Alzheimer's disease categorized as mild based on MMSE scores above 23 or 24 (Esteban-Santillan, Praditsuwan, Ueda, Geldmacher, 1998; Manos, 1999), partially because clock drawing performance is reportedly not affected by normal aging (Cahn & Kaplan, 1997). However, many mildly demented persons draw normal clocks; consequently, the sensitivity is poor in samples of mildly demented persons. The estimates of the sensitivity and specificity in samples of mildly demented persons range from 67% to 73% for the sensitivity and 77% to 97% for the specificity, depending on the scoring system (Lee, Seanwick, Coen, & Lawlor, 1996).

In addition to the diminished validity in screening for mild dementia, the clock drawing task might be affected by education level. Ainslie and Murden (1993) showed that a clock drawing task was affected by education level. However, Rouleau et al. (1996) and Ferrucci et al. (1996) found clock drawing to be independent of education. The complexity of the relations between clock drawing and education is evident in Lam et al.'s (1998) finding that a clock drawing task was correlated with education level for demented individuals but not correlated for normal elderly controls (there were no differences in education level between the control and demented groups). Other research has indicated that the clock drawing task is not as affected by education level or language difficulties as the MMSE is (Borson et al., 1999). As a result, the clock drawing task may be more useful than the MMSE and, potentially, the 3MS in identifying cognitive impairment due to dementia in individuals who have low levels of education (Borson et al., 1999).

Other Clock Tasks

Some clock tasks include not only clock drawing ability but also measure clock reading and clock setting (i.e., Lam et al., 1998; Tuokko, Hadjistavropoulos, Miller, & Beattie, 1992). Clock reading involves the reading of a pre-set clock, and clock setting involves drawing hands on a pre-drawn clock to indicate a specific time.

Inclusion of a pre-drawn clock in the setting task has an advantage over setting the clock time after drawing the clock face as is done in some clock drawing tasks. When a pre-drawing clock is not used in the setting portion of the clock drawing task, the ability to set the clock may be affected by a poorly drawn circle, misplacement of the clock numbers, or problems with fine motor coordination. Clock setting tasks where the clock face is pre-drawn allows for an untainted assessment of the ability to set a clock at a specific time (Freedman et al., 1994). Clock reading tasks may be similarly affected by a poorly drawn clock face. Thus, presentation of a pre-drawn clock assesses the ability to read time as presented on a clock.

Assessment of the ability to set time and read time on a clock face is important because these tasks are conceptual and involve abstract reasoning (Tuokko et al., 1995). Furthermore, both the ability to read time and the ability to set time on a clock face require spatial thinking, a parietal lobe function (Critchley, 1966). These tasks may be particularly sensitive to cognitive impairment due to Alzheimer's disease. Evidence suggests that Alzheimer's disease affects the parietal lobes (Critchley, 1966) and may be associated with a disturbance in the conceptualization of time (Tuokko et al., 1992).

Scoring practices for clock tasks which combine clock reading, clock setting, and clock drawing have used techniques which look at both the degree of impairment and the type of errors made by patients (Lam et al., 1998; Tuokko et al., 1992). Lam et al. (1998) found that clock reading and clock setting added to the predictive value of the clock drawing task when screening for dementia (i.e., combined predictive value of 98%). The authors noted that the clock reading and clock setting tests were relatively insensitive but highly specific.

The Clock Test

The Clock Test is a published test that measures Clock Setting, Clock Reading, and Clock Drawing and includes a standardized scoring procedure that includes a qualitative and quantitative assessment of errors made for Clock Drawing and the number of correct responses for Clock Reading and Clock Setting (Tuokko et al., 1995). Kurzman (1992) found that the components of the Clock Test are correlated. Specifically, the Clock Drawing task shows a moderate positive relationship with the Clock Setting ($r = 0.64$) and Clock Reading tasks ($r = 0.60$), and the Clock Setting task shows a strong positive relation with the Clock Reading task ($r = 0.79$).

Tuokko et al. (1992) assessed the sensitivity and specificity of the Clock Test in 62 normal elderly people (*M* age = 71.3 years, *SD* = 8.07) and patients with Alzheimer's disease who were diagnosed by consensus after medical, neurological, and neuropsychological investigations (*M* age = 70.6 years, *SD* = 7.45). They found that Clock Drawing had a sensitivity of 86% and a specificity of 92%, Clock Setting had a sensitivity of 97% and a specificity of 87%, and Clock Reading had a sensitivity of 85% and specificity of 92%. Contrary to the findings reported by Lam et al., (1998), Tuokko et al., (1992) found that the Clock Reading and Clock Setting components of the Clock Test were as valuable as the Clock Drawing task in identifying those with dementia.

Although the Clock Test does not provide a composite score, overall performance on the Clock Test may be assessed by using a criteria of positive results (i.e., scores below the cut-off point) on two or more of the three Clock Test components as an indication of impairment. With this method, the sensitivity of the Clock Test was 93% and the specificity was 94%; overall, the Clock Test has a correct classification rate of 93% (Tuokko et al., 1992). Using the same method (i.e., positive results on two or more of the Clock Test components), Tuokko et al. (1995) found that the overall sensitivity was 80% and the overall specificity was 82% after adjusting cut-off scores on the component tests by age groups (i.e., 65-79 and 80+ years old). The Tuokko et al. (1995) study used a large population-based sample of normal elderly people (*N* = 1753, *Mdn* age = 75 years) and a clinical sample of Alzheimer's disease patients (*N* = 269, *Mdn* age = 73). The use of a population-based sample may explain the lower estimates of the overall discriminability of the Clock Test, because community-based samples are typically more heterogeneous than clinic-based samples.

In addition to the adequate concurrent criterion validity (i.e., sensitivity and specificity) of the Clock Test, superior predictive criterion validity has been demonstrated. O'Rourke, Tuokko, Hayden & Beattie, (1997) found that Clock Test scores from time 'one' differentiated individuals who were diagnosed with dementia at time 'two' from individuals who were not demented at time 'two.' The sample was clinic-based and all individuals were without dementia at time 'one.' At time one and time two, participants were assessed medically, neurologically, and neuropsychologically, and the diagnosis was by consensus. The time interval between time 'one' and time 'two' was, on average, 22 months. Using the criteria of positive results (i.e., scores below the cut-off point) on two or more of the three Clock Test components as an indication of impairment, O'Rourke et al. found that the predictive sensitivity was 91% and the predictive specificity was 95%.

Other evidence of validity of the Clock Test has been established. In particular, the construct validity of the Clock Test was demonstrated by concurrent correlations with neuropsychological tests that assessed similar domains. Divergent validity was demonstrated by low correlations of the Clock Test components to unrelated domains (e.g., mood and social behaviour; Hadjistavropoulos, Tuokko, & Beattie, 1991; Kurzman, 1992).

In addition to possessing adequate validity, the Clock Test demonstrated reliability. Inter-rater reliability assessed in a clinic sample of patients with Alzheimer's disease and normal elderly persons, was 0.90 to 0.95 for the Clock Test

(Tuokko et al., 1992). Analysis of inter-rater reliability across specific items in each component indicated reliabilities of 0.95 to 0.98 for Clock Setting items and 0.99 for Clock Reading items. Clock Drawing error inter-rater reliabilities ranged from 0.73 to 1.00 (Kurzman, 1992). Using a clinic sample and a normal elderly sample, the stability of scores on the Clock Test was found to be 0.70 over a four day interval (Tuokko et al., 1992). The internal consistency for Clock Setting and Clock Reading, as measured by Cronbach's alpha, were 0.95 and 0.86, respectively (Kurzman, 1992). Clock Drawing is not suitable for this assessment of reliability as scoring involves assessment of different types of errors, and, thus, would not be internally consistent.

Combining the Clock Test and the 3MS

Two screening tools may be combined in an attempt to increase the sensitivity and specificity of screening for a particular disease. In order for this combination to be effective, both screening tools must be valid and independently add to the discriminability of diseased and non-diseased persons (Lin, 1999). Preliminary evidence with simultaneous multiple regression suggests that the Clock Setting component of the Clock Test predicts cognitive impairment due to dementia over and above the 3MS (Tuokko, 1993 as cited in Tuokko et al., 1995). Although these results suggest that combining Clock Setting with the 3MS may add to the differentiation between cognitively normal and cognitively impaired persons, there are no data on whether the added sensitivity and specificity is significantly better (i.e., are the hit rate and correct rejection rate better?) than with one screening tool alone. Simultaneous multiple regression informs about unique prediction but does not inform about whether the added prediction is statistically and/or clinically useful (i.e., is the hit rate better and if so, by how much?).

Alternatively, components of the Clock Test that do not uniquely discriminate above and beyond the 3MS in simultaneous multiple regression may offer unique discriminability for particular age, education, and/or sex groups. Although the 3MS may be affected by age, education level (Bravo & Hebert, 1997a; Tombaugh et al., 1996), and sex (Bravo & Hebert, 1997a), and the Clock Test may be affected by age (Tuokko et al., 1995) and education level (Corney, 2000), it is unclear if these two tests are equally affected at all age and education levels. Some research suggests that performance on the Clock Test appears to be uncorrelated with level of education (Tuokko et al., 1995), but the relation between the Clock Test and education is unclear, as poor performers on the Clock Test tend to have a low levels of education (Corney, 2000). In addition, research has indicated that clock drawing is not as affected by education level as is the MMSE (Borson et al., 1999). If follows, therefore, that the Clock Drawing task may not be as affected by education as the 3MS, which was derived from the MMSE. These results suggest that adding the other components of the Clock Test (i.e., Clock Drawing and Clock Reading) to the 3MS is an appropriate investigation for the different demographic groups (i.e., age, education, and sex groups).

Hypotheses

The hypotheses of the proposed study can be divided into three parts. The first set of hypotheses addresses the influence of age, education, and sex on the 3MS. The second set of hypotheses addresses whether the addition of the components of the Clock Test to the 3MS adds to the identification of persons with and without cognitive impairment as indicated by greater sensitivity and/or specificity. The third set of hypotheses investigates the influence of age, education, and sex on the components of the Clock Test and then investigates whether adding the components of the Clock Test to the 3MS enhances to the identification of cognitive impairment for different groups based on age, education, and sex.

Effects of Age, Education, and Sex on the 3MS

Previous research has found that 3MS scores were affected by age, education level (Bravo & Hebert, 1997a; Tombaugh et al., 1996), and sex (Bravo & Hebert, 1997a). Older persons obtained lower scores on the 3MS (Bravo & Hebert, 1997a), and increasing age was found to have reduced the specificity of the 3MS (Tombaugh et al., 1996). Further, 3MS scores were lower for persons with fewer years of education (Bravo & Hebert, 1997a); however, lower levels of education affected the 3MS by reducing the specificity and increasing the sensitivity (Tombaugh et al., 1996). Finally, females obtained higher scores on the 3MS than did males (Bravo & Hebert, 1997a). Since the influence of these demographic variables likely reduced the validity (i.e., sensitivity and/or specificity) of the 3MS, I adjusted for the influence of these demographic variables in an attempt to increase the validity in terms of increased sensitivity and/or specificity of the 3MS.

1. I had no reason to expect that age, education, and sex in this sample would affect performance on the 3MS any differently than in previous research. Specifically, I expected that the performance of younger individuals (i.e., those between 65-79 years old) on the 3MS would be superior to the performance of older individuals (i.e., those between 80-89 years of age). In addition, I expected the 3MS

to be more accurate, in terms of greater sensitivity and/or specificity, for younger than for older persons. I also expected that persons with eight or fewer years of education would perform more poorly on the 3MS than would persons with nine or more years of education; consequently, I expected the 3MS to be more accurate for persons with nine or more years of education than for persons with eight or fewer years of education. Finally, I expected that females would obtain higher scores on the 3MS than males, and I expected that the 3MS would be more accurate for females than for males.

2. I expected that 3MS scores adjusted for the effects of age, education, and sex would be more accurate at identifying persons with and without cognitive impairment than 3MS scores not corrected for the effects of age, education, and sex.

Accuracy of Combined Screening Tools

I investigated whether the use of two complementary screening tools increased the accuracy of identification of persons with and without cognitive impairment as indicated by greater sensitivity and/or specificity over any one screening tool alone.

1. Based on the previous research, which suggested that the Clock Setting component of the Clock Test predicted cognitive impairment over and above the 3MS (Tuokko, 1993 as cited in Tuokko et al., 1995), I hypothesized that the combined Clock Setting/3MS measure would have significantly greater discriminability than the 3MS and Clock Setting measures alone.

2. I predicted that the accuracy of the combined Clock Drawing/3MS and Clock Reading/3MS would be greater than the accuracy of the Clock Test components alone (i.e., Clock Drawing and Clock Reading). However, I predicted

that the combined Clock Drawing/3MS and Clock Reading/3MS would be less accurate than the 3MS alone.

Demographic Influences on Components of the Clock Test

Previous research found that age was correlated with performance on the Clock Test (Tuokko et al., 1995) and that both age and education affected performance on the 3MS (Tombaugh et al. 1996). Based on their finding that age affected performance on the Clock Task, Tuokko et al. (1995) presented different cut-offs for the following five age groups: less than 69 years old, 70-74 years old, 75-79 years old, 80-84 years old, and over 85 years old. Tombaugh et al. (1996) separated their normative data into the following four groups: 65-79 years old with eight or fewer years of education, 65-79 years old with nine or more years of education, 80-89 years old with eight or fewer years of education, and 80-89 years old with nine or more years of education.

I proposed to investigate two groups that differ with respect to their total years of education: persons with eight or fewer years of education and those with nine or more years of education. The age groups specified by Tombaugh et al. (1996) were less specific but compatible with the age groups specified by Tuokko et al. (1995). Therefore, I proposed to investigate the influences of age on the screening measures in persons between 65-79 years of age and those above 80 years of age.

Next, the ability of the combined versus single screening tools to discriminate between persons with and without cognitive impairment for specific demographic groups (i.e., age, education, and sex groups) was examined.

1. I expected that the components of the Clock Test would be affected by age such that persons over 80 years of age would perform more poorly than persons between 65 and 79 years of age; consequently, I expected that the components of the Clock Test would be more accurate at identifying persons with and without cognitive impairment for younger than for older persons. Consistent with previous research by Tuokko et al. (1995), I did not expect performance on components of the Clock Test to be affected by education, and I expected that the accuracy of each component of the Clock Test would be equivalent for the two age groups. I did not expect the average score of the components of the Clock Test to be significantly different for males or females., Consequently I did not have any *a priori* hypotheses about the accuracy of the components of the Clock Test for males and females.

2. Previous research with a clock drawing task and the MMSE indicated that a clock drawing was not as affected by education as was the MMSE (Borson et al., 1999). Based on this research, I hypothesized that the combination of Clock Drawing and the 3MS would result in increased sensitivity and/or specificity over any one measure alone for individuals who were less educated (i.e., had eight or fewer years of education). I did not have any *a priori* hypothesis about the superiority of the combined Clock Drawing/3MS measure over either measure alone for the two age groups (i.e., 65-79 years old and 80 years and over) or sex groups.

3. Based on the hypothesis that education affected the 3MS but not the Clock Setting task, I hypothesized that the superiority of the combined Clock Setting/3MS measure over either measure alone would be especially pronounced in persons with eight or fewer years of education. I did not have any *a priori*

hypotheses about the superiority of the combined Clock Setting/3MS measure over either measure alone for the two age groups or sex groups.

4. Because Clock Reading appeared to be unaffected by education (Tuokko et al., 1995), I hypothesized that the Clock Reading component of the Clock Test combined with the 3MS would add to the discriminability of persons with fewer years of education (i.e., eight or fewer years of education). I did not have any *a priori* hypotheses about the superiority of the combined Clock Reading/3MS measure over either measure alone for the two age or sex groups.

CHAPTER 3

Method

Participants

This study used data collected as part of the Canadian Study of Health and Aging (CSHA-1), a national, epidemiological study of the health of older Canadians, 1990-1991. This sample consisted of randomly selected community-dwelling persons and residents of institutions who were 65 years of age and older. A sub-sample of the 10,263 persons who took part in CSHA-1 were administered both the 3MS and the Clock Test (n = 2071). Of these, 88 persons either did not complete all trials of the Clock Test or had missing data for either age, education level, or sex. Consequently, the final sample included 1983 persons. This sample (N = 1983) was composed of 883 males and 1100 females who had between zero and 30 years of formal education (M = 11.28, SD = 3.67) and whose ages ranged from 65 to 98 years (M = 75.38, SD = 6.95). All persons who tested positive on the 3MS screen (i.e., scored below 77 on the 3MS) went on for verification of their diagnosis whereas only a random sample of those who tested negative went on for verification: in total 309 persons had their diagnosis verified. The different methods of sampling based on the results of the screen meant that the sample of participants was influenced by a verification bias (corrected as described later).

Participants who had their diagnosis verified (n = 309) were assessed by nursing, medical, and neuropsychological staff. Following this, a case conference was held, and after consensus, each participant received one of the following diagnoses: no cognitive impairment (n = 181), cognitive loss without dementia (n = 74), probable Alzheimer's disease ($n = 17$), possible Alzheimer's disease ($n = 19$), vascular dementia ($n = 13$), other dementia ($n = 2$), or unclassifiable dementia ($n = 3$). Diagnosis of dementia was based on the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders third edition, revised (American Psychiatric Association, 1987). Cognitive loss without dementia was diagnosed by clinical judgement when participants had cognitive impairment that was not sufficient for a diagnosis of dementia (i.e., impairment was only in one cognitive domain).

For the purpose of this study, participants were either classified as cognitively normal (i.e., no cognitive impairment) or as cognitively impaired, which could be either due to dementia or due to cognitive loss without dementia. These two groups (i.e., cognitively impaired and not cognitively impaired) were chosen to reflect the diagnostic concerns of clinicians who are interested in identifying persons who require a thorough examination to elucidate whether or not their cognitive impairment it is due to a dementing process. The cognitively impaired group ($n = 128$) was older ($M = 81.64$, $SD = 6.32$) and had fewer years of education ($M = 9.27$, $SD = 4.36$) than the cognitively normal group (M age = 79.08, $SD = 6.93$; M years education = 10.53, $SD = 3.71$; $t = -3.32$; $p = 0.001$, $t = 2.74$, $p = 0.006$; respectively). Both groups were comprised of equivalent numbers of males ($n = 165$) and females ($n = 144$): 53% of the cognitively impaired and cognitively normal groups were male ($\chi^2_1 = 0.500$, $p > 0.05$; $\chi^2_1 = 0.934$, $p > 0.05$; respectively).

Measures

The Clock Test and the 3MS were administered to all participants (N = 1983). The 3MS is not a copyrighted test and is widely available. The following items were included in the 3MS: the ability to read and obey a command, write a sentence, copy pentagons, follow a three-stage command, register three new words to remember and then recall the words after a delay, orientation to birth, time, and space, naming words, and naming of four-legged animals, abstract similarities in two things, repeat sentences, count backward, and spell backward. All the questions were presented aurally, and the answers were given orally for most questions, except for the items that measured the abilities to read and follow a written command, write, and copy a design. The scores on the 3MS ranged from 0-100 and include a mix of dichotomous right/wrong scores and graded scoring (Teng & Chui, 1987). In addition, in an attempt to reduce variability and ambiguity during testing, a manual was used for scoring and administration (Kristjansson, 1990). See Appendix A for a copy of the 3MS.

In the Clock Test, the Clock Drawing component was given first. Participants were presented with a pre-drawn circle and asked to write all the numbers as they would appear on a clock. Then participants were asked to draw in the hands of the clock to make it read 10 past 11. In the Clock Setting component, participants were presented with circles with dashes corresponding to where the numbers should be. They were then asked to draw the hands of the clock to read the following five times: 1:00, 11:10, 3:00, 9:15, and 7:30. In the Clock Reading component, participants were presented with five clocks set to the above five times, but in a different order, and were asked to tell the examiner the time indicated (Tuokko et al, 1995).

The Clock Test provided three separate scores: (1) a Clock Reading total score, (2) a Clock Setting total score, and (3) a Clock Drawing total score. The Clock Reading score was based on the correct response for each time. One point was awarded if either the hour or minute hand was correctly identified, and three points were awarded if both the hour and minute hand were correctly identified. The Clock Reading component had a maximum of three points for each time and a total maximum of 15 points. The Clock Setting total score was based on the correct placement of the hands according to a template. One point was awarded for correct placement of the hour hand, one point for the minute hand, and one point was awarded for the relative lengths of the hour and minute hands where the minute hand was sufficiently longer (i.e., by one-half a centimeter) than the hour hand. The Clock Setting component had a maximum of three points for each time, and a total maximum of 15 points. Finally, the Clock Drawing score was a frequency tally of types of errors (i.e., omissions, preservations, rotations, misplacements, distortions, substitutions, and additions). Consequently, higher scores on Clock Drawing corresponds with poorer performance (Tuokko et al., 1995). The Clock Test is published by and available from Multi-Health Systems.

In this study, the maximum Clock Drawing score was 50 (i.e., 50 errors). To ensure that higher scores indicated better performance, the Clock Drawing score was reversed: for each individual, 50 minus the total Clock Drawing score was computed. This ensured that lower scores indicated impairment, which was the same as the other components of the Clock Test (i.e., Clock Setting and Clock Reading). Consequently, a score of 50 on Clock Drawing – Reversed indicated perfect performance on the Clock Drawing task.

Procedure

First Step: Effects of Age, Education, and Sex on the 3MS

The influence of age, education, and sex on the 3MS was investigated to see if the 3MS could be improved as a screening tool for cognitive impairment by removing the influences of demographic variables. First, the 3MS score was predicted by age, education, and sex with simultaneous multiple regression. Next, the simultaneous multiple regression weights were used to adjust the 3MS for significant predictors. Consequently, the adjusted 3MS score for each individual was their original 3MS score minus the effects of age, education, and sex, each weighted according to their regression weights. Because it was based on regression weights, the adjusted 3MS score tended to be a fraction. In order to reduce the number of possible cut-off points, all fractions were rounded to whole numbers (e.g., 98.5 became 99, 99.4 became 99), and the result was the adjusted 3MS score.

Second Step: Correcting the Verification Bias

The ability of the adjusted-3MS and original 3MS scores to differentiate between individuals with a final diagnosis of cognitively normal versus cognitively impaired was determined at all possible cut-off points while correcting for the verification bias. Because all persons who screened positive were verified by diagnosis, whereas only a portion of those who screened negative were verified, there is a bias in the estimate of the accuracy of the screening tools for this sample. Unless corrected, the verification bias would inflate the sensitivity of a screening tool and underestimate the specificity (Begg & Greenes, 1983).

Begg and Greenes (1983) proposed a mathematical procedure based on Bayes's probability theorem to correct the bias in the estimate of the accuracy (i.e., sensitivity and specificity) of the screening tool. The probability of a result on the screening tool (i.e., positive or negative) given a specific disease status (i.e., cognitively impaired or cognitively normal) was mathematically derived from the probability of a verified disease status given the result on the screening tool. In order to apply Bayes's theorem, the likelihood of having the disease verified must not be dependent on the disease status. The separation between probability of having the disease verified and the disease status only occurs when the likelihood of having the disease verified depends on the result of the screening tool (i.e., all positives were verified and some negatives were verified; Begg & Greenes, 1983). These conditions existed in the CSHA-1 sample.

If R represents the resulting value of the screening tool and D represents the disease status, then the conditional probability of a test result given a disease status would be $Pr(R \mid D)$. Further, V- represents cases not selected for verification of their disease status and V+ represents cases selected for verification. In this sample, the only information available was the conditional probability of a test result given a disease status for the verified sample, $Pr(R \mid D, V+)$. However, $Pr(R \mid D)$ does not equal $Pr(R \mid D, V+)$, and there is no information regarding the disease status (D) for the unverified sample (V-). Bayes's theorem allows estimation of Pr(R) from all persons (i.e., V- and V+) using $Pr(R \mid D, V+)$, which is estimated from the verified sample. Bayes's theorem is shown in Equation 1.

$$\Pr(\mathrm{R|D}) = \frac{\Pr(R)\Pr(D \mid R, V+)}{\sum_{R} \Pr(R)\Pr(D \mid R, V+)} \quad (1)$$

Using these proportions, Donald (1996) calculated the unbiased sensitivity and specificity, where R+ was a positive result on the screening tool and R- was a negative result on the screening tool. These values (i.e., R+ and R-) could only be obtained when a cut-off point had been established. Consequently, the verification bias had to be corrected at each possible cutoff point. The correction for the sensitivity is shown in Equation 2.

$$\text{Sensitivity} = \Pr(\mathrm{R+|D+}) = \frac{\Pr(D+ \mid R+)\Pr(R+)}{\Pr(D+ \mid R+)\Pr(R+) + \Pr(D+ \mid R-)\Pr(R-)} \quad (2)$$

In order to calculate the unbiased sensitivity at each cut-off point, the following frequencies were required: the number of diseased persons (D+, V+) who scored positive on the screen (R+), and the number of diseased persons (D+, V+) who scored negative on the screen (R-), the number who scored positive on the screen with a verified diagnosis (R+, V+), the number who scored negative on the screen with a verified diagnosis (R-, V+), the total number who scored positive on the screen (R+) regardless of whether their diagnosis was verified, the total number who scored negative on the screen (R-) regardless of whether their diagnosis was verified, and the total number in the sample (N = 1983). See Table 1 for a summary of all mathematical symbols and what they represent.

The calculation of the numerator of the unbiased sensitivity involved three steps. First, the number of diseased persons (D+, V+) who scored positive on the screen (R+) was divided by the number who scored positive on the screen with a verified diagnosis (R+, V+). This is shown in Equation 3.

$$\Pr(\text{D+} \mid \text{R+}) = \frac{D+R+,V+}{R+,V+} \tag{3}$$

Second, the total number who scored positive on the screen (R+) regardless of whether their diagnosis was verified was divided by the total number in the sample (*N*). This is shown in Equation 4.

$$\Pr(\text{R+}) = \frac{R+}{N} \tag{4}$$

Finally, Equation 3 and Equation 4 were multiplied to provide the numerator of the corrected sensitivity. The calculation of the denominator of the corrected sensitivity also involved multiple steps. First, the number of diseased persons (D+, V+) who scored negative on the screen (R-) was divided by the number who scored negative on the screen with a verified diagnosis (R-, V+). This is shown in Equation 5.

$$\Pr(\text{D+} \mid \text{R-, V+}) = \frac{D+R-,V+}{R-,V+} \tag{5}$$

Second, the total number who scored negative on the screen (R-) regardless of whether their diagnosis was verified was divided by the total number in the sample (*N*). This is shown in Equation 6.

$$\Pr(\text{R-}) = \frac{R-}{N} \tag{6}$$

Third, Equation 5 and Equation 6 were multiplied and their product was added to the product of Equation 4 multiplied by Equation 5 to provide the denominator of the equation for the corrected sensitivity. The corrected sensitivity was equal to the numerator divided by the denominator (see Equation 2). This procedure was repeated at all possible cut-off points separately for the 3MS and the adjusted-3MS.

Table 1. Summary of Mathematical Abbreviations used to Calculate Sensitivity and Specificity and the Sub-samples that they Represent.

Abbreviation	Sub-Sample of Participants
D+, V+	Number of diseased persons
D-, V+	Number of disease free persons
R+ D+, V+	Number of diseased persons who scored positive on the screen
R- D+, V+	Number of diseased persons who scored negative on the screen
R+ D-, V+	Number of disease free persons who scored positive on the screen
R- D-, V+	Number of disease free persons who scored negative on the screen
R+, V+	Total number of persons who scored positive on the screen who had their disease status verified, regardless of whether they were found to be diseased or disease free
R-, V+	Total number of persons who scored negative on the screen who had their disease status verified, regardless of whether they were found to be diseased or disease free
R+	Total number of persons who scored positive on the screen, regardless of whether their disease status was verified or not
R-	Total number of persons who scored negative on the screen, regardless of whether their disease status was verified or not

The corrected specificity of the 3MS and adjusted-3MS is shown in Equation 7.

$$\text{Specificity} = \Pr(\text{R-}|\text{D-}) = \frac{\Pr(D-|R-)\Pr(R-)}{\Pr(D-|R-)\Pr(R-)+\Pr(D-|R+)\Pr(R+)} \quad (7)$$

The calculation of the unbiased specificity required similar frequencies to those required in calculating the unbiased sensitivity. In order to calculate the unbiased specificity at each cut-off point, the following frequencies were required: the number of disease free persons (D-, V+) who scored positive on the screen (R+), and the number of disease-free persons (D-, V+) who scored negative on the screen (R-), the number who scored positive on the screen with a verified diagnosis (R+, V+), the number who scored negative on the screen with a verified diagnosis (R-, V+), the total number who scored positive on the screen (R+) regardless of whether their diagnosis was verified, the total number who scored negative on the screen (R-) regardless of whether their diagnosis was verified, and the total number in the sample ($N = 1983$).

The calculation of the numerator of the unbiased specificity also involved three steps. First, the number of disease-free persons (D-, V+) who scored negative on the screen (R-) was divided by the number who scored negative on the screen with a verified diagnosis (R-, V+). This is shown in Equation 8.

$$\Pr(\text{D-}\,|\,\text{R-, V+}) = \frac{D-R-,V+}{R-,V+} \quad (8)$$

Second, the total number who scored negative on the screen (R-) regardless of whether their diagnosis was verified was divided by the total number in the sample

(N). This is shown in Equation 9.

$$\Pr(\text{R-}) = \frac{R-}{N} \tag{9}$$

Finally, Equation 8 and Equation 9 were multiplied to provide the numerator of the corrected specificity. To calculate the denominator of the corrected specificity first, the number of disease free persons (D-, V+) who scored positive on the screen (R+) was divided by the number who scored positive on the screen with a verified diagnosis (R+, V+). This is shown in Equation 10.

$$\Pr(\text{D-} \mid \text{R+}) = \frac{D - R+, V+}{R+, V+} \tag{10}$$

Second, the total number who scored positive on the screen (R+) regardless of whether their diagnosis was verified was divided by the total number in the sample (N). This is shown in Equation 11.

$$\Pr(\text{R+}) = \frac{R+}{N} \tag{11}$$

Third, Equation 10 and Equation 11 were multiplied and their product was added to the product of Equation 8 multiplied by Equation 9 to provide the denominator of the equation for the corrected specificity. The corrected specificity was then equal to the numerator divided by the denominator (see Equation 7). This was repeated for all possible cut-off points for the 3MS and then for the adjusted-3MS.

In conclusion, correction for the verification bias changed the sensitivity and specificity of each screening tool from their previously distorted proportions. Correcting for the verification bias increased the specificity and decreased the sensitivity of each screening tool. For example, the corrected 3MS had a higher specificity but lower sensitivity than the uncorrected 3MS (see Figure 1 for an illustration of the differences between the 3MS corrected for the verification bias and the biased or uncorrected 3MS).

Third Step: Calculating the AUCs for the 3MS and Adjusted-3MS

Once the unbiased sensitivity and specificity of the 3MS and adjusted-3MS were obtained for each cut-off point, the results were plotted on a receiver operating characteristic (ROC) plot. ROC plots illustrate the screening tool's ability to identify diseased and disease-free persons over all of the possible combinations of sensitivity and specificity (i.e., all possible cut-off points; Zweig & Campbell, 1993). The curve of the ROC plot measures the overall accuracy of the screening tool (Swets, 1988). Although in clinical practice one must choose a single cut-off point in order to determine whether or not someone is cognitively impaired, reporting only one number for the sensitivity and one number for the specificity of a screening tool at one cut-off point does not convey the overall accuracy of the screening tool. When assessing the effectiveness of a screening tool, it is best to look at all possible cut-off points (Zweig & Campbell, 1993).

A ROC plot is the proportion of true-positives, or sensitivity, versus the proportion of false positives (one minus the specificity; Swets, 1988). In nonparametric ROC plots, the raw estimates of the sensitivity and specificity are plotted, which is the most appropriate type of ROC plot to use when the scores on the screening tool are continuous (Zweig & Campbell, 1993), as they are in this study. Once the ROC plot is constructed, the area under the curve (AUC) of the ROC plot can be calculated. The AUC is a representation of the accuracy of the screening tool at all possible cut-off points. The greater the AUC (and the higher the plot is in the left-hand corner), the more accurate the screening tool is. The AUC is one number and, therefore, is an efficient means of

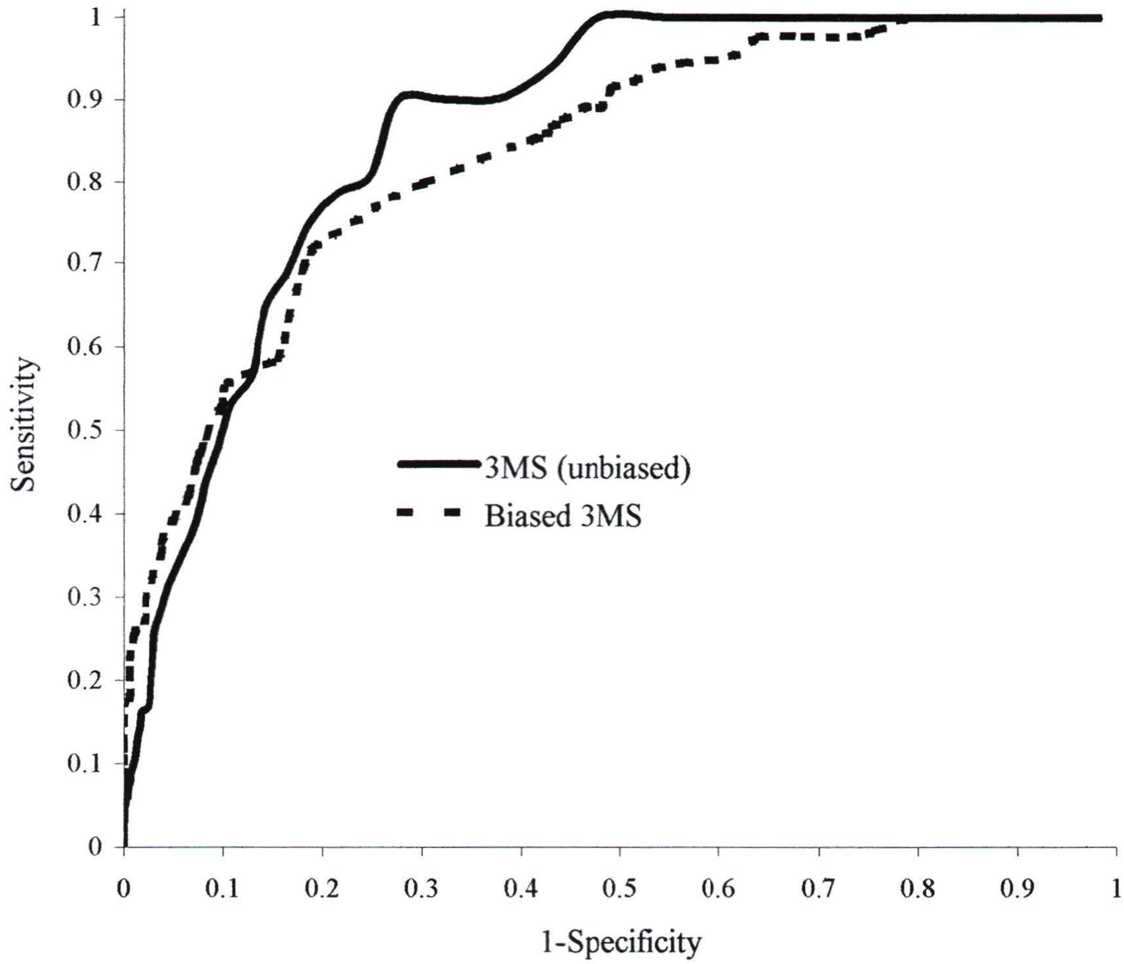


Figure 1. ROCs of the biased 3MS and (unbiased) 3MS.

communicating the effectiveness of the screening tool at all possible cut-off points (Zweig & Campbell, 1993).

The AUC of a nonparametric ROC plot is irregular and cannot be described by a mathematical function. The area under the curve is estimated by adding up the area of the unusual rectangles (i.e., trapezoids) that make up the irregularly shaped polygon (Bamber, 1975; Hanley & McNeil, 1982). The equation for the AUC is shown in Equation 12.

$$\text{AUC} = \sum_i [1/2(\textit{sensitivity}_i + \textit{sensitivity}_{i+1})] * [(1 - \textit{specificity}_{i+1}) - (1 - \textit{specificity}_i)] \quad (12)$$

This was calculated for the 3MS and the adjusted-3MS.

Due to sampling error, a few of the trapezoids were found to be negative. The negative values occurred when the specificity of an earlier cut-off point was higher than the specificity for the next cut-off point (i.e., *i* plus 1). This anomaly occurred at very low cut-off points where there were few participants whose scores were similar to these cut-off points. All negative values were close to zero and, consequently, were rounded to zero in the computation of the AUCs.

Finally, because the theoretical sampling distribution of an AUC was unknown, the standard error of the estimate for each AUC was calculated from a sample of 1000 AUCs (see Figure 2 for an example of the distribution of AUCs for the original 3MS). Similar to the procedure used by Zhou and Higgs (2000), the 1000 AUCs were computed from screening test scores and disease results (i.e., cognitive impairment, no cognitive impairment, or not validated), which were randomly sampled ($N = 1983$) with replacement (i.e., using bootstrapping). The sensitivities and specificities used to

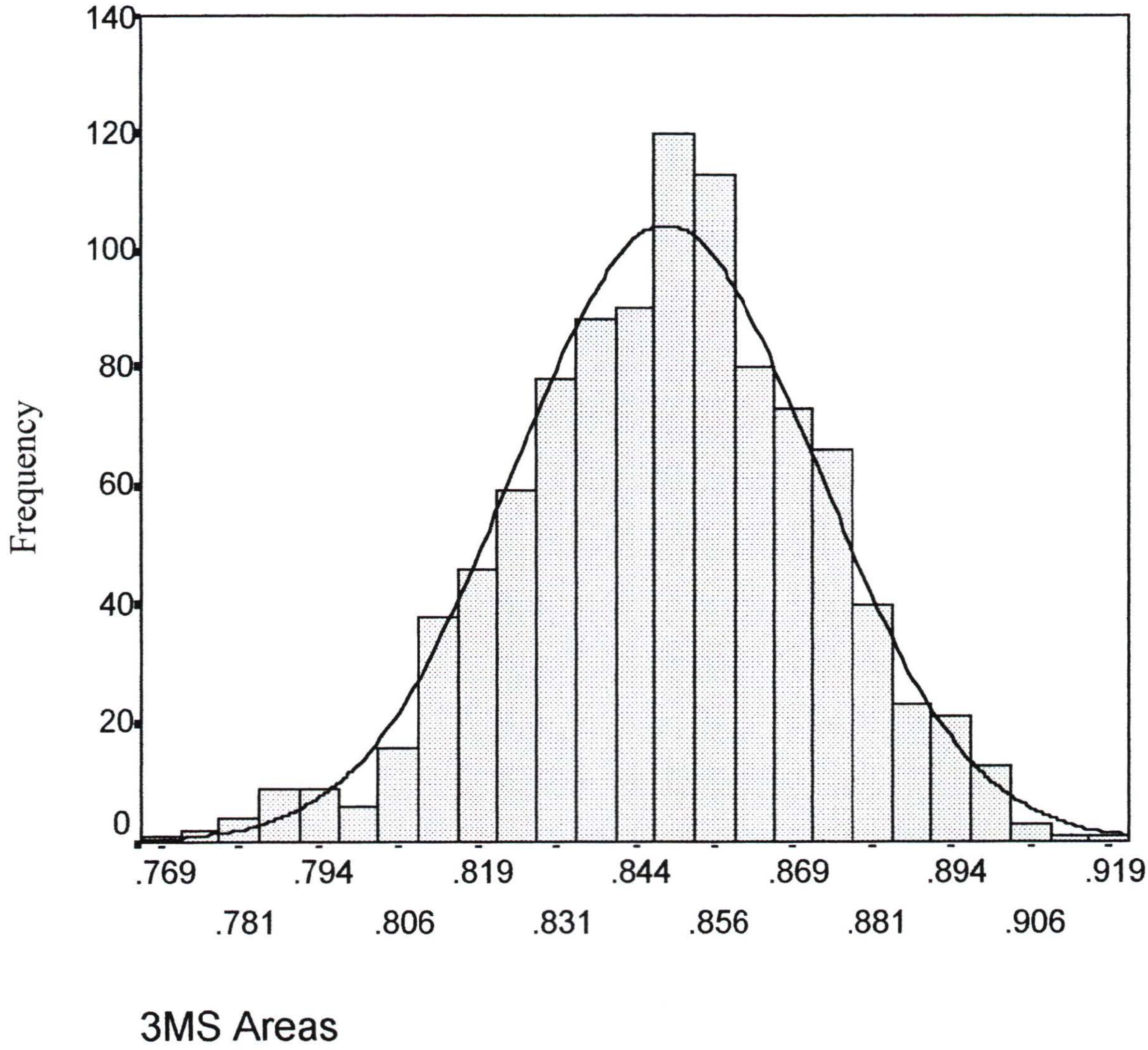


Figure 2. Sampling distribution of 1000 AUCs for the 3MS.

calculate the AUC using Equation 12 were corrected for the verification bias using Equation 2 and Equation 7 (as outlined in the Second Step). A custom-made computer program written in Visual Basic carried out this computationally intensive procedure. For each AUC, the standard error was equal to the standard deviation of the sampling distribution of the 1000 AUCs (i.e., for the 3MS and adjusted-3MS).

Fourth Step: Comparing the AUCs of the 3MS and Adjusted-3MS

Since the two screening tools were administered to the same person for the same disease, the results were likely to be positively correlated (Zweig & Campbell, 1993). The results of correlated screening tools have smaller sampling variability than do the results of non-correlated screening tools. The comparison of the difference between two correlated AUCs can be made more sensitive by taking into account the amount of correlation between the two ROC curves (Bloch, 1997; DeLong, DeLong, & Clarke-Pearson, 1988; Hanley & McNeil, 1983).

Hanley and McNeil (1983) provided a table that approximates the correlation between two tests (r) based the average AUC of the two tests and the average correlation between the two tests for the diseased and non-diseased participants. Hanley and McNeil's table only provided correlation coefficients for average areas of 0.700 or higher; consequently, any two screening tools with average AUCs less than 0.700 were assigned a correlation coefficient of 0 when calculating the *z*-value (see below).

Hanley and McNeil (1983) also provided a modified z-test, which accounted for the correlation between the two AUCs (r). Equation 13 for the z-test is shown for two AUCs with standard errors, SE_1 and SE_2, the equation follows:

$$z = \frac{AUC_1 - AUC_2}{\sqrt{SE_1^{\ 2} + SE_2^{\ 2} - 2rSE_1SE_2}} \tag{13}$$

The AUC of the 3MS was compared to the AUC of the adjusted-3MS with Equation 13 to determine whether they were significantly different from each other. In addition, the magnitude of this difference was calculated (i.e., AUC_1-AUC_2) to investigate how much more accurate one test was over the other. This was done to determine whether the difference in the accuracy of the two tests was clinically, as well as statistically, significant.

Fifth Step: Combining Screening Tools

The next step was to combine components of the Clock Test with the 3MS in an attempt to make a more accurate screening tool. If the adjusted-3MS was found to be more accurate than the 3MS, the adjusted-3MS would have been used in combination with the components of the Clock Test. However, if the adjusted-3MS was found to be equal to or less accurate than the 3MS (i.e., smaller or equal AUCs), then the 3MS would have been used in combination with the components of the Clock Test.

Next, forward stepwise logistic regression with 3MS in the first block and Clock Drawing - Reversed, Clock Setting, and Clock Reading as predictors of disease status in the second block was used to determine which components of the Clock Test, if any, added discriminative power above the 3MS. Neither a constant nor any higher order interactions were included in the model because the goal of the regression was to determine which components of the Clock Test added discriminative power above the 3MS.

Combined scores were created by adding unweighted scores from components of the Clock Test with the 3MS. The following combined scores were computed in this manner: combined Clock Drawing-Reversed/3MS, combined Clock Setting/3MS, and combined Clock Reading/3MS. In addition, a Total Clock Test score was computed by adding all of the unweighted scores from the components of the Clock Test (i.e., Clock Drawing – Reversed, Clock Setting, and Clock Reading). The Total Clock Test score was combined with the 3MS in order to determine whether the combined Total Clock Test/3MS was more accurate than any screening test alone. Finally, a combined score was computed based on the results of the forward stepwise logistic regression where the unweighted Clock Test components were added to the unweighted 3MS to determine whether this combined score was better at predicting cognitive impairment than any screening score alone.

Sixth Step: Correcting the Verification Bias for the Combined Screening Tools and Components of the Clock Test

The ability of the single and combined measures to differentiate between individuals with a final diagnosis of cognitively normal versus cognitively impaired was determined over all possible cut-off points. The corrected sensitivities and specificities were calculated as outlined in the Second Step of the procedure for the following measures: Clock Drawing - Reversed, Clock Setting, Clock Reading, combined Clock Drawing – Reversed/3MS, combined Clock Setting/3MS, combined Clock Reading/3MS, and combined Total Clock Test/3MS.

Seventh Step: Comparing Combined versus Single Screening Tools

First, ROC plots for the combined and single screening tools were constructed as described in the Third Step of the procedure. The AUC of each curve and the standard error for each AUC was computed, as described in the Third Step of the procedure. Next, all of the combined screening tools (i.e., Clock Drawing - Reversed, Clock Setting, Clock Reading, Clock Drawing – Reversed/3MS, Clock Setting/3MS, Clock Reading/3MS, and Total Clock Test/3MS) were compared individually to each single screening tool that made up the combination (i.e., 3MS, Clock Drawing - Reversed, Clock Setting, and Clock Reading). The Type 1 Error rate was adjusted for family-wise comparisons. For each group of comparisons (e.g., the combined Clock Drawing – Reversed/3MS task versus the 3MS and the combined Clock Drawing – Reversed/3MS task versus the Clock Drawing – Reversed task), the alpha level of 0.05 was divided by two to equal a critical alpha of 0.025 two-tailed for each *z*-test.

Eighth Step: Comparing Screening Tools Across Age, Education, and Sex Groups

One of the goals of this research was to compare the effectiveness of the combined versus single screening tools for discrete age, education, and sex groups. Consequently, age, education, and sex groups were constructed. First, age, education, and sex were used as predictors of each Clock Test component with simultaneous multiple regression to determine which demographic variables were related to specific components of the Clock Test. Plots of age and education with each component of the Clock Test were examined to determine the best groups. If the relations between these continuous variables (i.e., age and education) appeared linear, then pre-determined groups were constructed. For education level, the two pre-determined groups were those with eight or fewer years of education versus those with nine or more years of education. For age, the two pre-determined groups were those between 65-79 years old versus those aged 80 years and older.

The groups compared for each combination depended on whether age, education, and/or sex was determined to be related to either of the measures in the combination. The specific groups depended on the analysis (as described in the First Step of the procedure). For example, if the 3MS were related to age, education, and sex, then the components of the Clock Test would be compared with the 3MS for these groups to determine if the components of the Clock Test added discriminative power above the 3MS across these groups.

Ideally, the 3MS combined with components of the Clock Test would have been compared across the following eight groups: females 65-79 years old with 8 or fewer years of education, females 80 years old and over with 8 or fewer years of education, females 65-79 years old with 9 or more years of education, females 80 years old and over with 9 or more years of education, males 65-79 years old with 8 or fewer years of education, males 80 years old and over with 8 or fewer years of education, males 65-79 years old with 9 or more years of education, and males 80 years old and over with 9 or more years of education. However, the number of participants in each group was too small (e.g., for females 65-79 years old with 8 or fewer years of education, there were only 7 with cognitive impairment and 14 who were cognitively normal). Consequently, comparisons of the effectiveness of the combined versus single screening tools were made separately across sex, age groups, and education groups. For example, the accuracy of Clock Drawing – Reversed/3MS was compared to the 3MS and then to Clock Drawing – Reversed separately for males and females, then for persons aged 65-79 years old and those over 80 years of age, and then for persons with 8 or fewer years of education and those with 9 or more years of education. In total, there were six demographic groups created (i.e., 65 – 79 years old, n = 1466; 80 plus years old, n = 517; 8 or fewer years of education, n = 451; 9 or more years of education, n = 1532; males, n = 883; and females, n = 1100).

Next, the corrected sensitivities and specificities were calculated for the 3MS, Clock Drawing – Reversed, combined Clock Drawing – Reversed/3MS, Clock Setting, combined Clock Setting/3MS, Clock Reading, and combined Clock Reading/3MS across each age, education, and sex group. The calculation of the corrected sensitivities and specificities was performed as described in the Second Step of the procedure. Then, as described in the Third Step of the procedure, ROC curves were constructed, the AUCs under each curve were computed, and the standard error of each AUC was determined.

Subsequently, the accuracy of each screening tool was compared across each demographic group. Specifically, the ability of the 3MS to identify younger (i.e., 65 – 79 years old) persons with cognitive impairment was compared to the 3MS's discriminability with the older age group (i.e., 80 plus years old). Similarly, the accuracy of the Clock Drawing – Reversed, Clock-Drawing – Reversed/3MS, Clock Setting, Clock Setting/3MS, Clock Reading, Clock Reading/3MS was compared across these two age groups. The accuracy of the 3MS, Clock Drawing – Reversed, Clock-Drawing – Reversed/3MS, Clock Setting, Clock Setting/3MS, Clock Reading, Clock Reading/3MS was compared for persons with 8 or fewer years of education and for persons with 9 or more years of education. Finally, the accuracy of the 3MS, Clock Drawing – Reversed, Clock-Drawing – Reversed/3MS, Clock Setting, Clock Setting/3MS, Clock Reading, Clock Reading/3MS was compared for males versus females. Since these analyses involved separate groups, the AUCs were not correlated. Consequently, a z-test for differences between independent means was used. The equation is shown in Equation 14.

$$z = \frac{AUC_1 - AUC_2}{\sqrt{SE_1^{\,2} + SE_2^{\,2}}} \tag{14}$$

Finally, the accuracy of the combined versus single screening tools was compared as described in the Fourth Step of the procedure. The following analyses were conducted for persons 65-79 years of age: 3MS versus Clock Drawing – Reversed/3MS, Clock Drawing – Reversed/3MS versus Clock Drawing, 3MS versus Clock Setting/3MS, Clock Setting/3MS versus Clock Setting, 3MS versus Clock Reading/3MS, and Clock Reading/3MS versus Clock Reading. These analyses were repeated for persons 80 plus years old, males, females, persons with 8 or fewer years of education, and persons with 9 or more years of education. The Type 1 Error rate was adjusted to account for the fact that two *z*-tests were required to determine whether or not the combination of complementary screening tools were more accurate than any of the screening tools alone for each demographic group. Therefore, the error rate was 0.025 two-tailed for each *z*-test.

Final Step: Description of a More Effective Screening Tool

Provided that one or more combinations of complementary screening tools were significantly more accurate than any one measure alone, descriptions of the best cut-off scores would have been reported, as would the predictive power of the combined screening tool.

CHAPTER 4

Results

Briefly, the results indicated that the original 3MS was marginally more accurate than the adjusted-3MS at identifying participants with and without cognitive impairment; consequently, the regression adjustment for the influences of age, education, and sex did not improve the accuracy of the (original) 3MS. Attempts to increase the accuracy of the identification of participants with and without cognitive impairment by adding complementary screening tools did not result in increased accuracy over the 3MS alone. In addition, each component of the Clock Test combined separately with the (original) 3MS and the Total Clock Test score combined with the 3MS resulted in statistically equivalent accuracy in terms of correctly identifying participants with and without cognitive impairment when compared to the 3MS alone. Finally, the combination of Clock Drawing – Reversed and Clock Setting with the 3MS was found to be less accurate than the 3MS alone at identifying participants with and without cognitive impairment. However, because of the relatively low sensitivity and specificity of the components of the Clock Test, any addition of the 3MS to the Clock Test resulted in increased accuracy over each component of the Clock Test alone.

When assessed separately for different groups based on age, education level, and sex, the results were similar: the addition of single components of the Clock Test to the 3MS resulted in equivalent accuracy in terms of correctly identifying participants with and without cognitive impairment compared to the 3MS alone. Finally, for each age, education, and sex group, any addition of the 3MS to the components of the Clock Test resulted in increased accuracy over each component of the Clock Test alone.

Effects of Age, Education, and Sex on the 3MS

The influence of age, education, and sex on the 3MS were investigated with simultaneous multiple regression. The correlation between the 3MS and these demographic variables is shown in Table 2. Age, education, and sex together predicted an individual's 3MS score ($F_{(3, 1979)} = 262.21, p < 0.001, R^2 = 0.284$). Each variable uniquely predicted the 3MS score: age ($t_{1979} = -19.97, p < 0.001$), education ($t_{1979} = 15.71$, $p < 0.001$), and sex ($t_{1979} = 5.36, p < 0.001$) were significant. On average, for every additional year in age, the 3MS score decreased by 0.563 of a point, and, on average, every additional year of education increased the 3MS score by 0.836 of a point. On average, if the participant was female her score on the 3MS increased by 1 (i.e., 1.08) point. These regression coefficients were used to make the adjusted-3MS as outlined in the First Step of the procedure.

The average 3MS and adjusted-3MS scores were significantly different for cognitively normal participants and participants with cognitive impairment (3MS $t_{307} =$ 12.02, $p < 0.001$; adjusted-3MS $t_{307} = 11.05, p < 0.001$). Cognitively normal participants scored higher on the 3MS ($M = 82.35$, $SD = 10.62$) and the adjusted-3MS ($M = 115.01$, $SD = 9.57$) than did participants with cognitive impairment (3MS $M = 65.54$, $SD =$ 13.94; adjusted-3MS $M = 100.70$, $SD = 13.21$). Consequently, all further analyses used the (unadjusted) 3MS.

Table 2. Correlation Matrix of 3MS and Clock Test Components with Demographic Variables.

	3MS	CD[a]	CS[b]	CR[c]	Sex	Age	Education
3MS	1.000	0.339	0.533	0.449	0.078	-0.427	0.368
CD[a]	--	1.000	0.440	0.342	-0.010	-0.231	0.127
CS[b]	--	--	1.000	0.490	-0.098	-0.380	0.281
CR[c]	--	--	--	1.000	-0.096	-0.277	0.194
Sex	--	--	--	--	1.000	0.078	0.019
Age	--	--	--	--	--	1.000	-0.163
Education	--	--	--	--	--	--	1.000

[a] CD = Clock Drawing – Reversed, [b] CS = Clock Setting, [c] CR = Clock Reading.

The ROC plots (i.e., sensitivity versus 1 minus the specificity at each cut-off point) for the 3MS and the adjusted-3MS are shown in Figure 3. The original 3MS (AUC = 0.8465, *SE* = 0.0239) was marginally better than the adjusted-3MS (AUC = 0.8140, *SE* = 0.0280; $z = 1.93$, $p = 0.054$) at correctly identifying participants with and without cognitive impairment.

<u>The 3MS Combined with the Clock Test</u>

Forward stepwise logistic regression analysis was performed with disease status (i.e., cognitively normal or cognitively impaired) as the outcome measure and with the first block with only the 3MS as a predictor. In the second block, the 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks were entered as predictors. There was no constant included in the model and only first-order associations were in the model. The final model, the 3MS, the Clock Drawing – Reversed, and the Clock Setting tasks, significantly predicted whether individuals were cognitively impaired or cognitively normal ($\chi^2_3 = 92.90$, $p < 0.001$, Cox & Snell $R^2 = 260$). 3MS uniquely predicted disease status (Wald statistic = 30.14, $p < 0.001$), as did Clock Drawing – Reversed (Wald statistic = 40.94, $p < 0.001$), and Clock Setting (Wald statistic = 18.07, $p < 0.001$). However, in the full model, Clock Reading did not uniquely predict disease status (Wald statistic = 0.070, $p = 0.792$) above the 3MS, Clock Drawing – Reversed, and Clock Setting predictors. These data suggest that the Clock Drawing – Reversed and Clock Setting tasks add unique predictability to the 3MS when discriminating persons with and without cognitive impairment; therefore, Clock Drawing – Reversed and Clock Setting appear to be appropriate to add to the 3MS. Finally, the logistic regression weights, which were based on raw scores, were close to 1 for each component of the

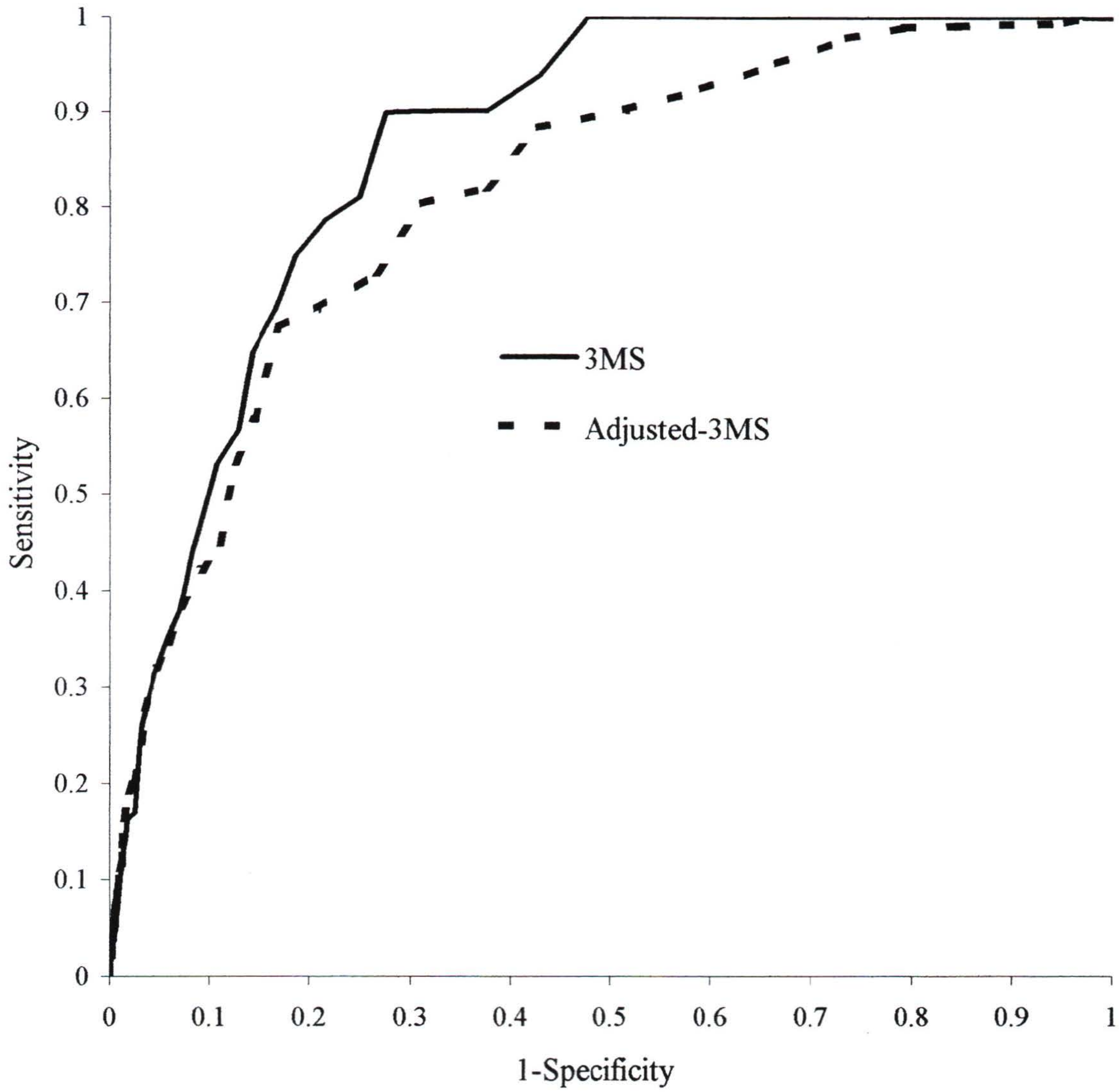


Figure 3. ROCs of the Adjusted-3MS and the (original) 3MS.

Clock Test. In addition, unweighted scores were less complicated and more likely to be used clinically. Consequently, unweighted scores for the components of the Clock Test and 3MS were used for all subsequent analyses.

Unweighted scores for Clock Setting, Clock Drawing – Reversed, and the 3MS were added together to make a combined Clock Drawing – Reversed/Clock Setting/3MS score, which was compared with the accuracy of each screening test score alone. Figure 4 shows the ROCs for the combined Clock Drawing – Reversed/Clock Setting/3MS, 3MS, Clock Drawing – Reversed, and Clock Setting tasks. The Combined Clock Drawing – Reversed/Clock Setting/3MS (AUC = 0.7894, *SE* = 0.0335) was less accurate than the 3MS alone (AUC = 0.8465, *SE* = 0.0239, $z = 3.29$, $p < 0.001$). Overall, the 3MS alone was 5.7% more accurate than the combined Clock Drawing – Reversed/Clock Setting/3MS was at identifying participants with and without cognitive impairment. However, the combined Clock Drawing – Reversed/Clock Setting/3MS (AUC = 0.7894, *SE* = 0.0335) was more accurate than both the Clock Drawing – Reversed task alone (AUC = 0.4416, *SE* = 0.0244, $z = 8.39$, $p < 0.001$, with a 34.8% increase in accuracy overall) and the Clock Setting task alone (AUC = 0.5583, *SE* = 0.0288, $z = 5.23$, $p < 0.001$, with a 23.1% increase in accuracy overall).

Finally, the combined Total Clock Test/3MS, which was computed from the unweighted addition of the Clock Drawing-Reversed, Clock Setting, Clock Reading, and 3MS scores was compared to the 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks alone to determine whether the combined Total Clock Test/3MS score was more accurate than any screening test alone. AUC analysis revealed that the combined Total Clock Test/3MS (AUC = 0.8449, SE =

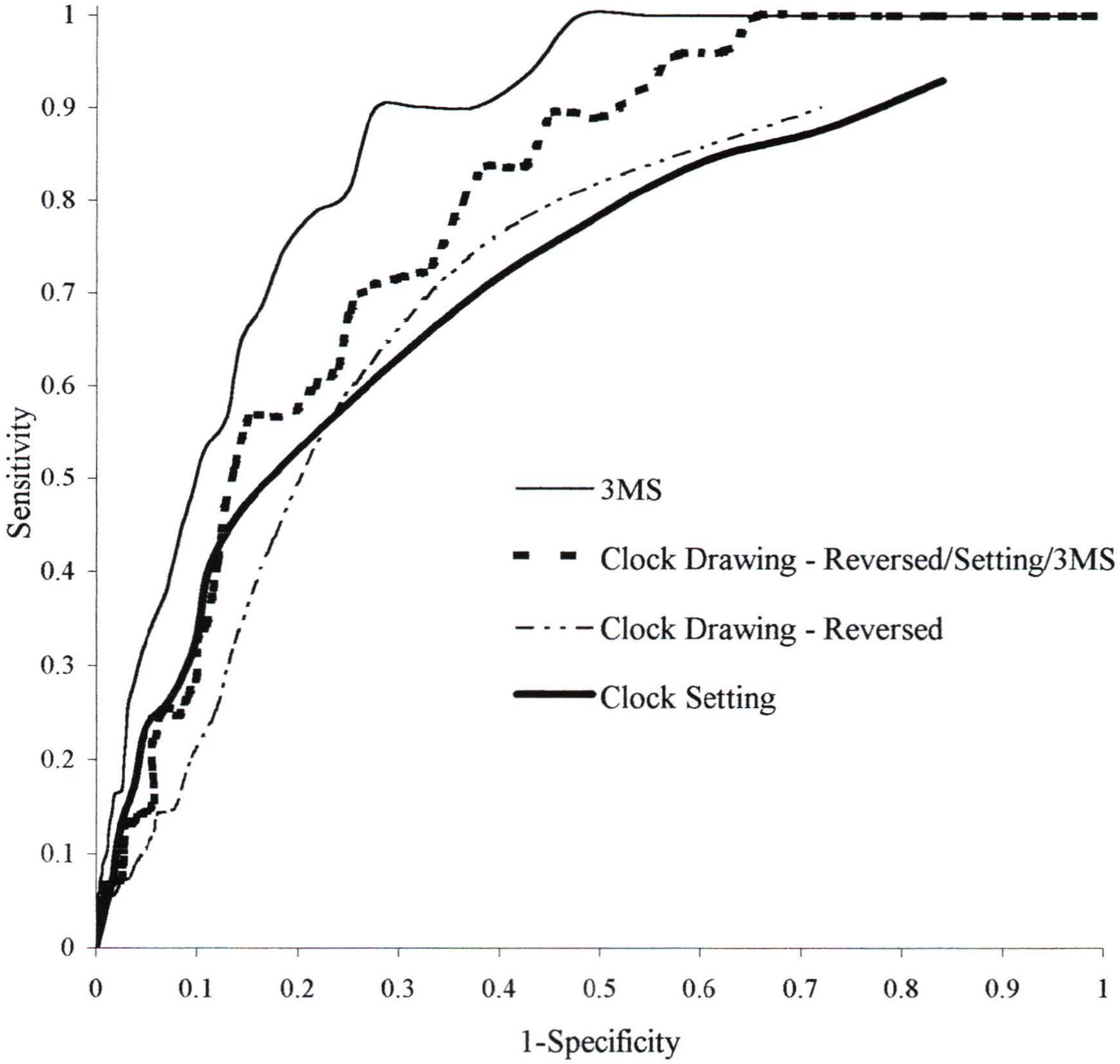


Figure 4. ROCs of the 3MS, Combined Clock Drawing - Reversed/Clock Setting/3MS, Clock Drawing – Reversed, and Clock Setting tasks.

0.0262) was equally accurate as the 3MS alone (AUC = 0.8465, SE = 0.0239, z = -0.12, p = 0.904). However, the combined Total Clock Test/3MS (AUC = 0.8449, SE = 0.0262) was more accurate than each of the components of the Clock Test alone: the AUCs of Clock Drawing – Reversed (AUC = 0.4416, SE = 0.0244, z = 11.26, p <0.001, with a 40.3% increase in accuracy overall), Clock Setting (AUC = 0.5583, SE = 0.0288, z = 7.64, p <0.001, with a 28.7% increase in accuracy overall), and Clock Reading (AUC = 0.0974, SE = 0.0088, z = 27.00, p <0.001, with a 74.8% increase in accuracy overall) were all significantly smaller than the combined Total Clock Test/3MS AUC. Figure 5 illustrates the AUCs of the total Clock Test/3MS, 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks.

<u>Components of the Clock Test added to the 3MS</u>

The average Clock Drawing – Reversed and combined Clock Drawing – Reversed/3MS scores were significantly different for cognitively normal participants and participants with cognitive impairment (Clock Drawing - Reversed $t_{307} = 4.82$, $p < 0.001$; Clock Drawing – Reversed/3MS $t_{307} = 11.39$, $p < 0.001$). Cognitively normal participants scored higher on the Clock Drawing – Reversed task (M = 46.13, SD = 4.84) than did participants with cognitive impairment scored (M = 42.91, SD = 6.92). In addition, the average score for cognitively normal participants on the combined Clock Drawing – Reversed/3MS was higher (M = 128.48, SD = 13.42) than the average score for cognitively impaired participants (M = 108.45, SD = 17.47).

In addition, the average Clock Setting and combined Clock Setting/3MS scores were significantly different for cognitively normal and impaired participants (Clock Setting $t_{307} = 7.60$, p < 0.001; Clock Setting/3MS $t_{307} = 12.14$, $p < 0.001$). Participants

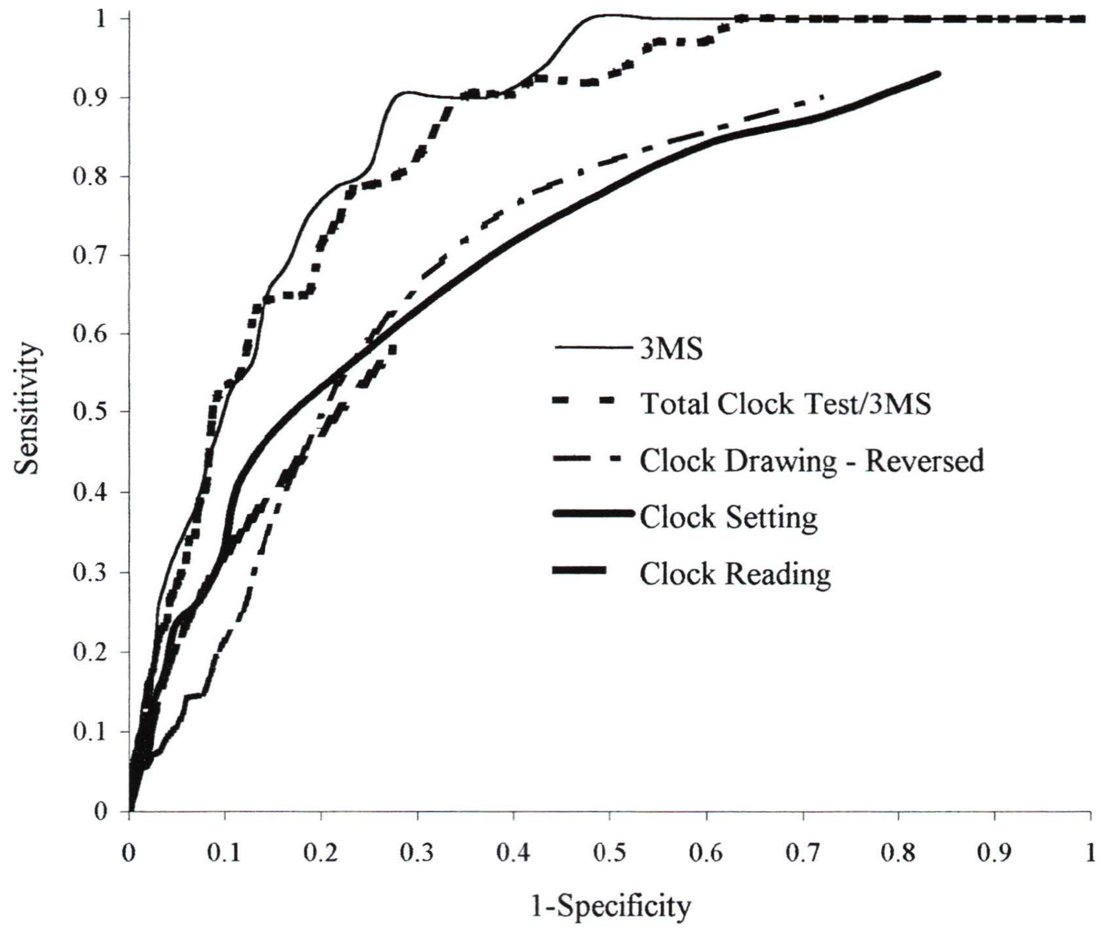


Figure 5. ROCs of the 3MS, Total Clock Test/3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks.

with cognitive impairment scored lower on both tests (Clock Setting *M* = 6.13, *SD* = 4.53; Clock Setting/3MS *M* = 92.08, *SD* = 12.91) than did cognitively normal participants (Clock Setting *M* = 9.73, *SD* = 3.78; Clock Setting/3MS *M* = 76.67, *SD* = 16.63).

Finally, the average Clock Reading and combined Clock Reading/3MS scores were significantly different for cognitively normal participants and participants with cognitive impairment (Clock Reading $t_{307} = 6.65$, $p < 0.001$; Clock Reading/3MS $t_{307} = 11.99$, $p < 0.001$). Cognitively normal participants scored higher on the Clock Reading test (*M* = 13.25, *SD* = 2.76) than participants with cognitive impairment (*M* = 10.66, *SD* = 4.11). In addition, the average score for cognitively normal participants on the combined Clock Reading/3MS was higher (*M* = 95.60, *SD* = 12.06) than the average score for cognitively impaired participants (*M* = 76.20, *SD* = 16.38).

The components of the Clock Test when combined with the 3MS were equally accurate compared to the 3MS alone at identifying persons with and without cognitive impairment; however, the combined scores were more accurate than each component of the Clock Test alone. Table 3 provides the AUCs, standard errors, *z*-values, and percentage difference (if relevant) for each combined screening tool compared to their constituent screening tools alone. The ROC plots of the 3MS, the Clock Drawing – Reversed task, and the combined Clock Drawing – Reversed/3MS task are shown in Figure 6. The ROC plots, which show the ability of the Clock Setting, combined Clock Setting/3MS, and 3MS tasks to identify cognitive impairment over all cut-off points, are shown in Figure 7. Figure 8 shows the ROC plots of the combined Clock Reading/3MS, Clock Reading, and 3MS.

Table 3. The AUCs, Standard Errors (*SEs*), *z*-values, and Differences in Percent (Diff in %) for the Combined and Single Screening Tools.

Screening Tool	AUC	*SE*	*z*-Value	Diff in %
3MS	0.8465	0.0239		
			-0.27	N/A
Clock Drawing – Reversed/3MS	0.8499	0.0257		
			11.54*	40.8%
Clock Drawing - Reversed	0.4416	0.0244		
3MS	0.8465	0.0239		
			-0.97	N/A
Clock Setting/3MS	0.8585	0.0243		
			13.67*	30.0%
Clock Setting	0.5583	0.0288		
3MS	0.8465	0.0239		
			-0.47	N/A
Clock Reading/3MS	0.8522	0.0240		
			29.54*	75.5%
Clock Reading	0.0974	0.0089		

* $p < 0.001$ two-tailed.

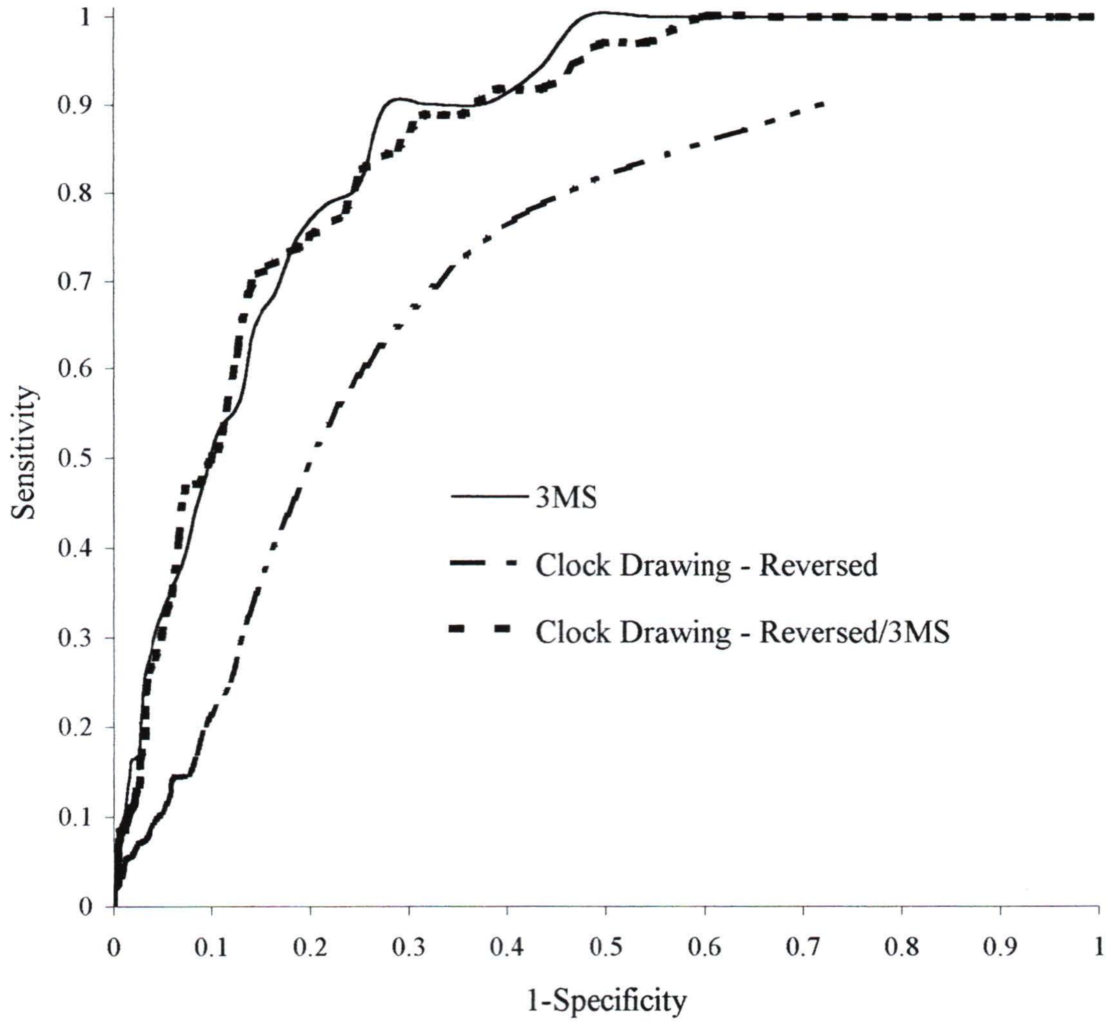


Figure 6. ROCs of the 3MS, Clock Drawing – Reversed, and Combined Clock Drawing – Reversed/3MS tasks.

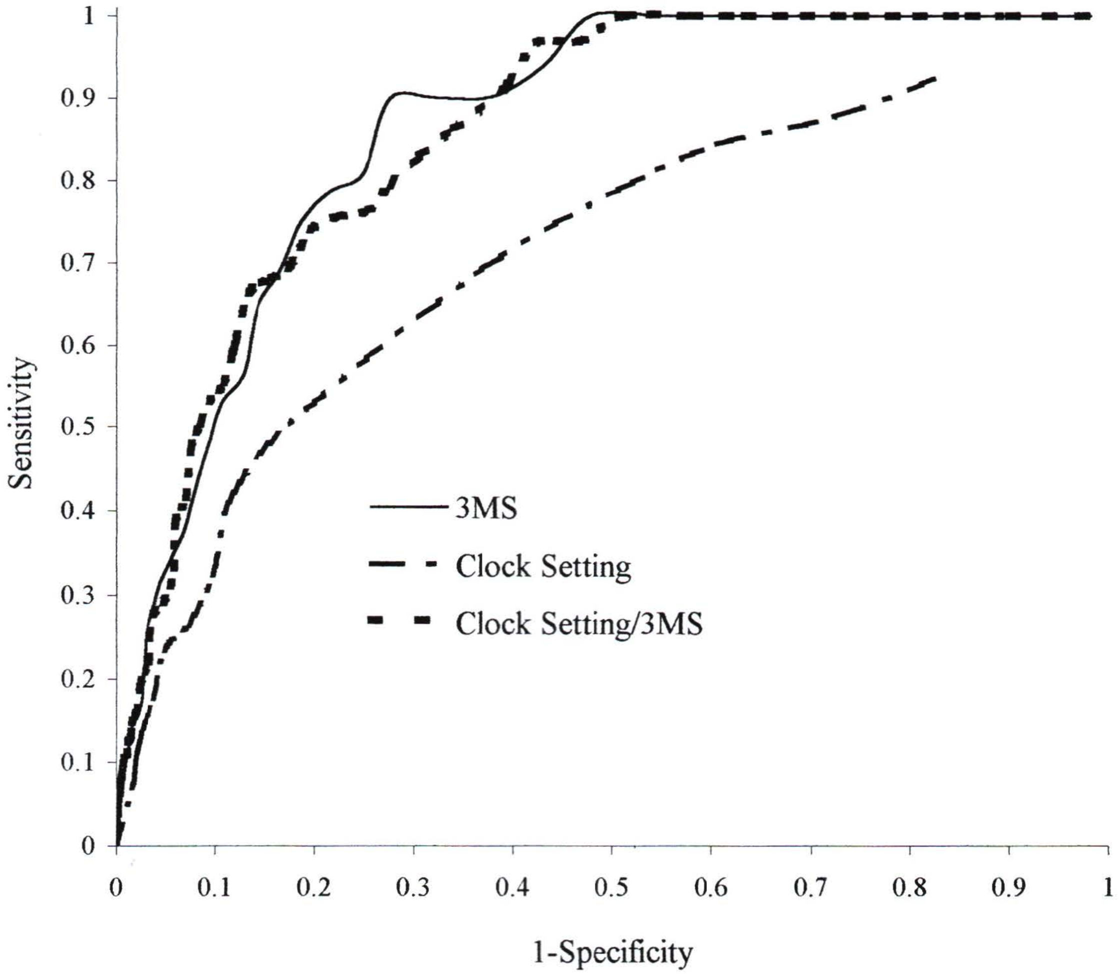


Figure 7. ROCs of the 3MS, Clock Setting, and Combined Clock Setting/3MS tasks.

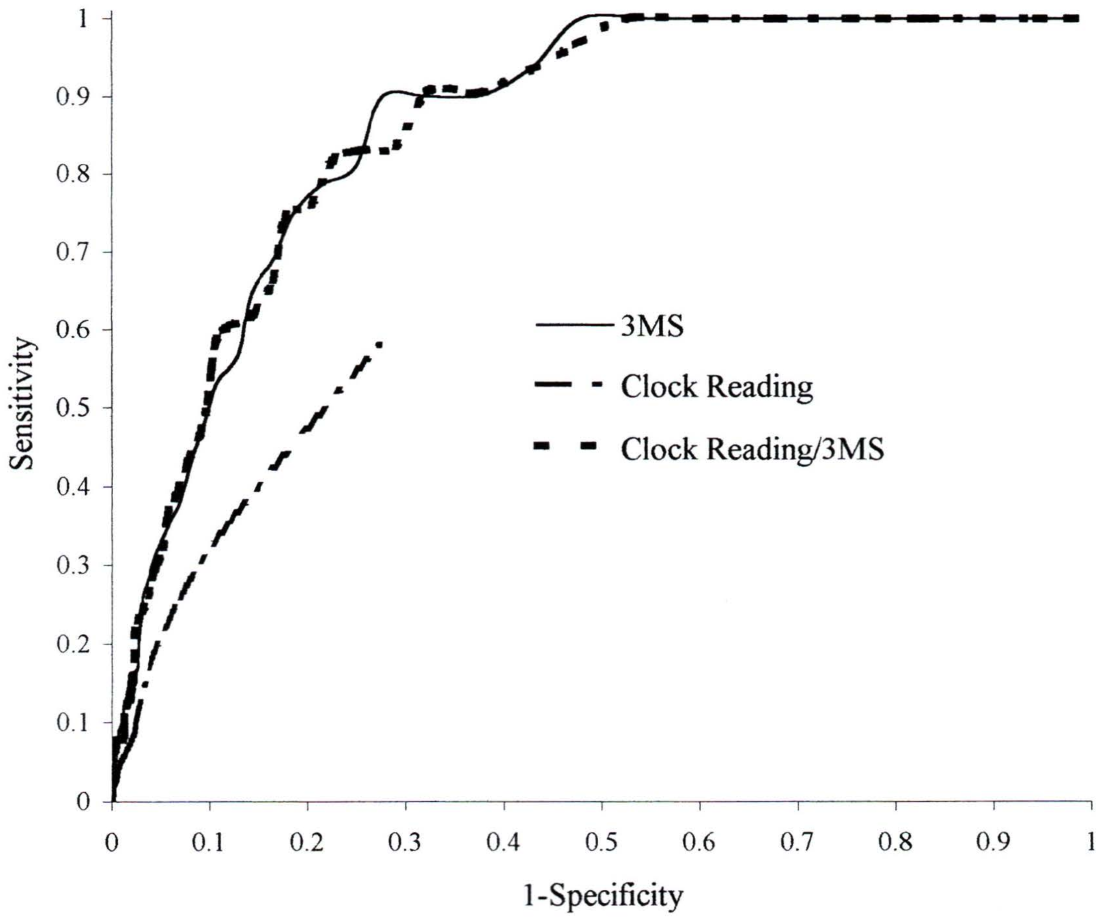


Figure 8. ROCs of the 3MS, Clock Reading, and Combined Clock Reading/3MS tasks.

Effects of Age, Education, and Sex on the Components of the Clock Test

Refer to Table 2 (p. 57) for a correlation of the components of the Clock Test with age, education, and sex. All components of the Clock Test were predicted by a combination of age, education, and sex (Clock Drawing – Reversed $F_{(3, 1979)} = 43.22, p < 0.001, R^2 = 0.61$; Clock Setting $F_{(3, 1979)} = 164.13, p < 0.001, R^2 = 0.199$; Clock Reading $F_{(3, 1979)} = 78.29, p < 0.001, R^2 = 0.106$). Clock Drawing – Reversed was uniquely predicted by age ($t_{1979} = -9.77, p < 0.001$) and education ($t_{1979} = 4.15, p < 0.001$). Sex did not predict Clock Drawing – Reversed scores ($t_{1979} = 0.25, p = 0.802$). All demographic variables uniquely predicted the Clock Setting score: age ($t_{1979} = -16.47, p < 0.001$), education ($t_{1979} = 11.15, p < 0.001$), and sex ($t_{1979} = -3.75, p < 0.001$) were significant. Similarly, for Clock Reading, all demographic variables uniquely predicted the total score: age ($t_{1979} = -11.38, p < 0.001$), education ($t_{1979} = 7.23, p < 0.001$), and sex ($t_{1979} = -3.73, p < 0.001$) were significant. Analysis of the plots of age (see Figure 9) and education (see Figure 10) versus components of the Clock Test revealed that the relations were relatively linear; consequently, pre-determined age and education groups were constructed. Two age groups were formed: persons aged 65 – 79 years old and persons 80 years and over. Two education groups were also formed: persons with 8 or fewer years of education and those with 9 or more years of education. Finally, the previously existing sex groups (i.e., males and females) were used for comparisons.

Accuracy of Screening Across Age Groups

The average 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading scores were significantly greater for participants aged 65 – 79 than for those aged 80 years and over (3MS $t_{307} = 16.55, p < 0.001$; Clock Drawing - Reversed $t_{307} = 8.18, p <$

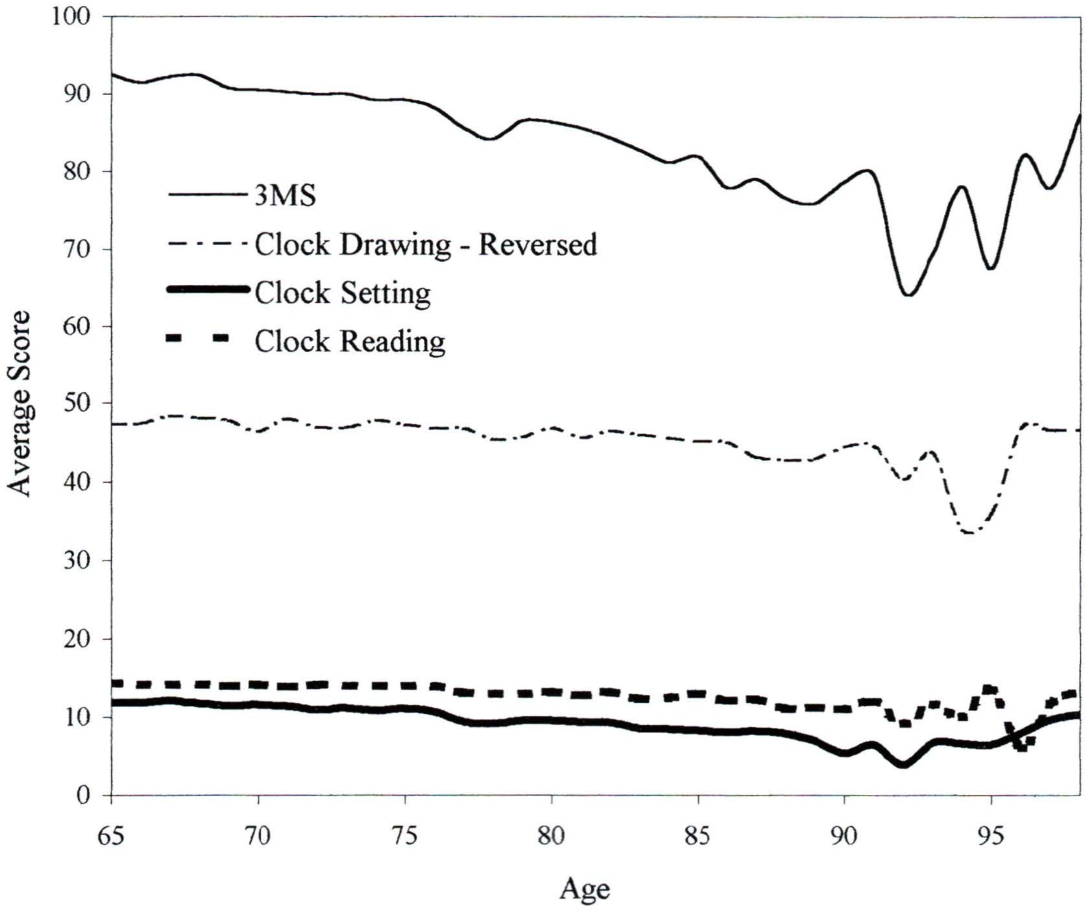


Figure 9. Linear plots of age versus the 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks.

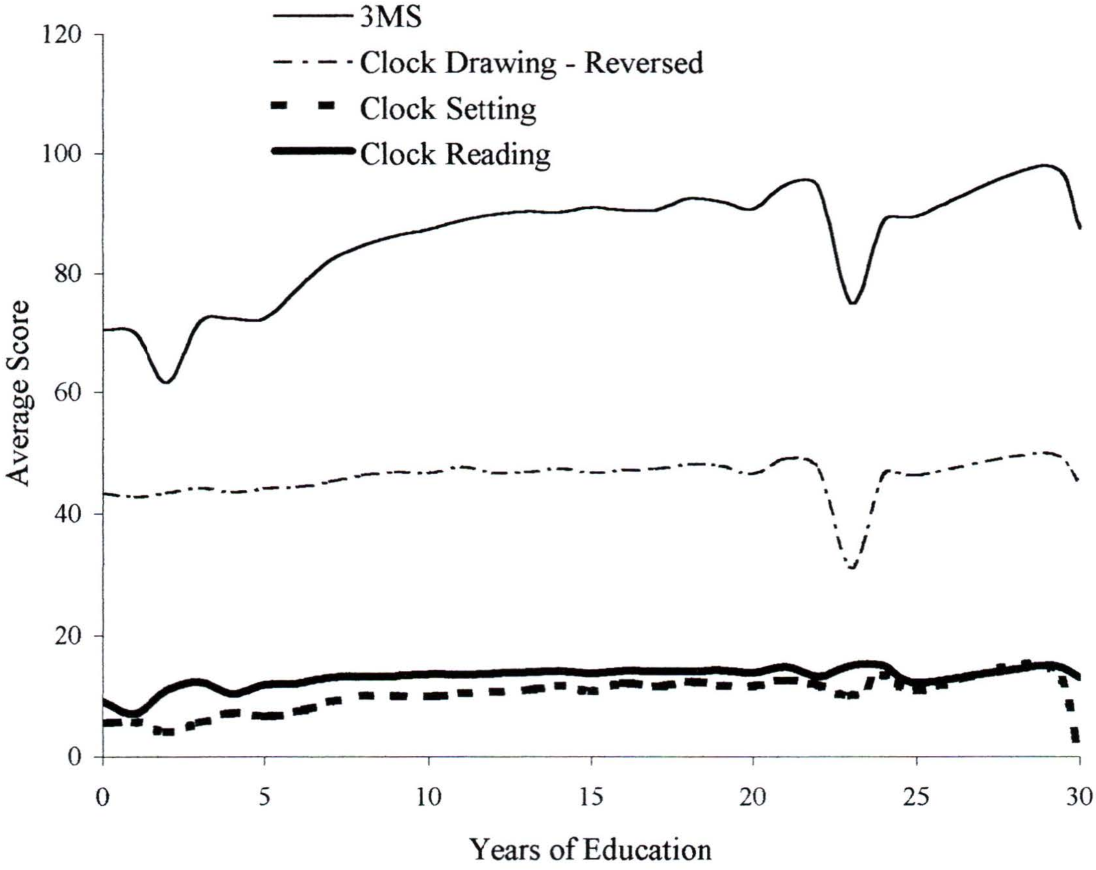


Figure 10. Linear plots of education versus the 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading tasks.

0.001; Clock Setting $t_{307} = 14.31, p < 0.001$; Clock Reading $t_{307} = 10.73, p < 0.001$). Similarly, the average combined scores were significantly greater for participants aged 65 – 79 years old than for those 80 years and over (Clock Drawing – Reversed/3MS $t_{307} = 16.571, p < 0.001$; Clock Setting/3MS $t_{307} = 17.90, p < 0.001$; Clock Reading/3MS $t_{307} = 17.13, p < 0.001$). The means and standard deviations of the performance of younger (i.e., 65 – 79 years old) and older (i.e., 80 and over years old) participants on each screening test are presented in Table 4.

The Clock Reading – Reversed, Clock Setting, and Clock Reading tasks were more accurate for older than for younger participants. However, the 3MS, and the combined screening tools were equally accurate for older and younger participants. The accuracy of each of the screening tests (i.e., AUCs with standard errors) for each age group, and the *z*-values and, if applicable, the difference in accuracy across age groups are summarized in Table 5.

Figure 11 illustrates the ROC plots of the 3MS for both age groups, and Figure 12 shows the ROC plots of the Clock Drawing – Reversed task for both age groups. Figure 13 illustrates the ROC plots for the combined Clock Drawing – Reversed/3MS tests for both age groups. The ROC plots of the Clock Setting task for each age group are shown in Figure 14. Figure 15 shows the ROC plots for the combined Clock Setting/3MS for younger and older age groups. Figure 16 illustrates for the ROC plots of Clock Reading for both age groups. Finally, Figure 17 shows the ROC plots of the combined Clock Reading/3MS for each age group.

Table 4. Mean Scores (*M*s) and Standard Deviations (*SD*s) of the Screening Tools for each Age Group.

	65 – 79 Years Old (*n* = 1466)		80+ Years Old (*n* = 517)	
	M	*SD*	*M*	*SD*
3MS	89.49	8.55	81.45	11.76
Clock Drawing – Reversed	47.14	4.40	45.12	5.85
Clock Drawing – Reversed/3MS	136.63	10.68	126.57	14.72
Clock Setting	10.97	3.26	8.38	4.24
Clock Setting/3MS	100.46	10.49	89.83	14.32
Clock Reading	13.82	2.24	12.39	3.43
Clock Reading/3MS	103.31	9.60	93.84	13.67

Table 5. The AUCs, Standard Errors (*SE*s), z-values, and the Difference in Percent (Diff in %) of Each Screening Tool for Each Age Group.

Screening Measure	AUC	*SE*	*z*-Value	Diff in %
3MS				
65-79 Years Old	0.8390	0.0366	0.09	N/A
80+ Years Old	0.8330	0.0304		
Clock Drawing – Reversed				
65-79 Years Old	0.4089	0.0338	-2.37*	12.6 %
80+ Years Old	0.5353	0.0413		
Clock Drawing – Reversed/3MS				
65-79 Years Old	0.8367	0.0401	-0.01	N/A
80+ Years Old	0.8370	0.0296		
Clock Setting				
65-79 Years Old	0.4965	0.0438	-2.38*	14.4 %
80+ Years Old	0.6403	0.0416		
Clock Setting/3MS				
65-79 Years Old	0.8455	0.0349	0.01	N/A
80+ Years Old	0.8451	0.0311		
Clock Reading				
65-79 Years Old	0.0831	0.0101	-3.47*	8.2 %
80+ Years Old	0.1651	0.0213		
Clock Reading/3MS				
65-79 Years Old	0.8397	0.0378	-0.01	N/A
80+ Years Old	0.8400	0.0266		

* $p < 0.02$ two-tailed.

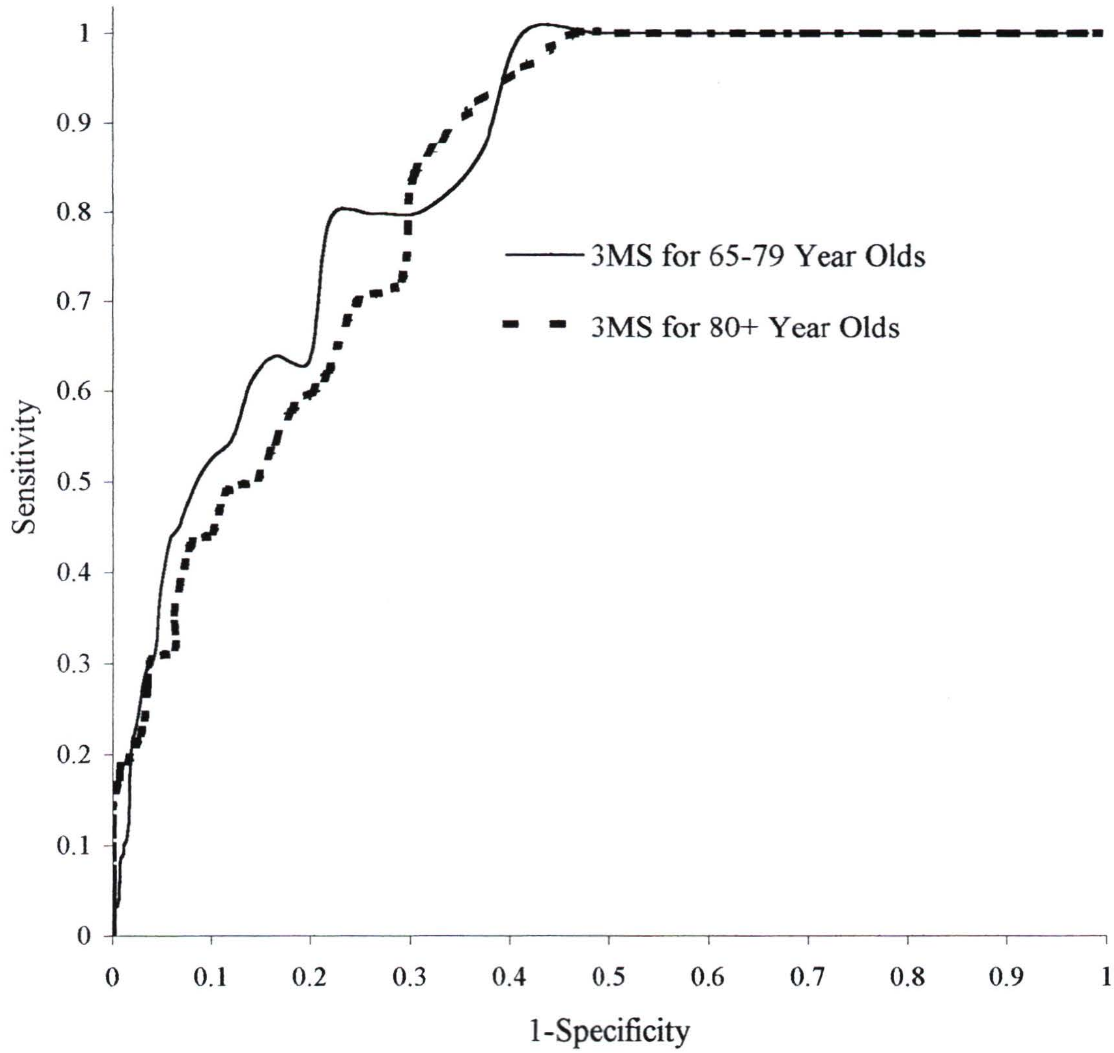


Figure 11. ROCs of the 3MS for each age group

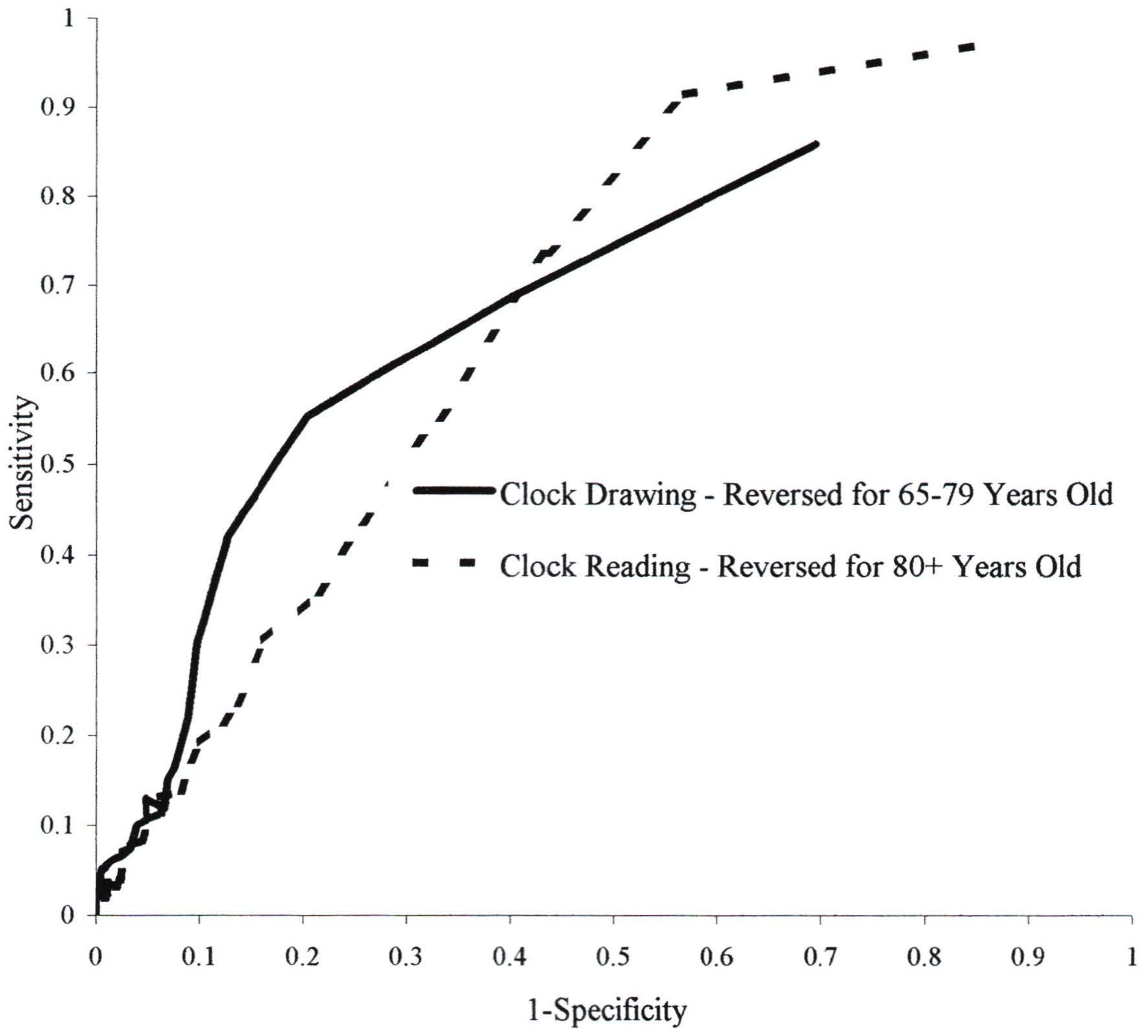


Figure 12. ROCs of the Clock Drawing – Reversed task for each age group.

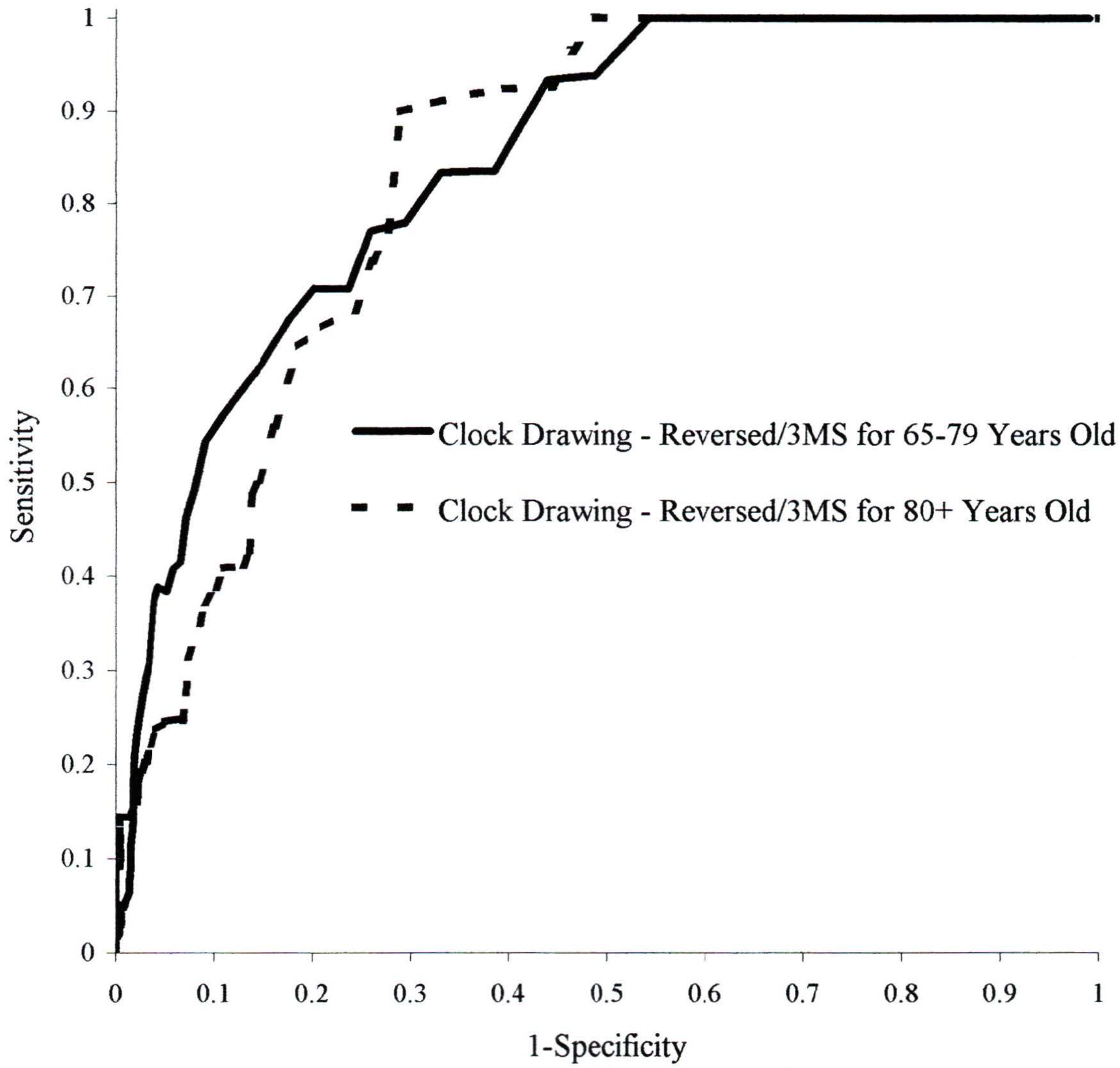


Figure 13. ROCs of the Combined Clock Drawing – Reversed/3MS task for each age group.

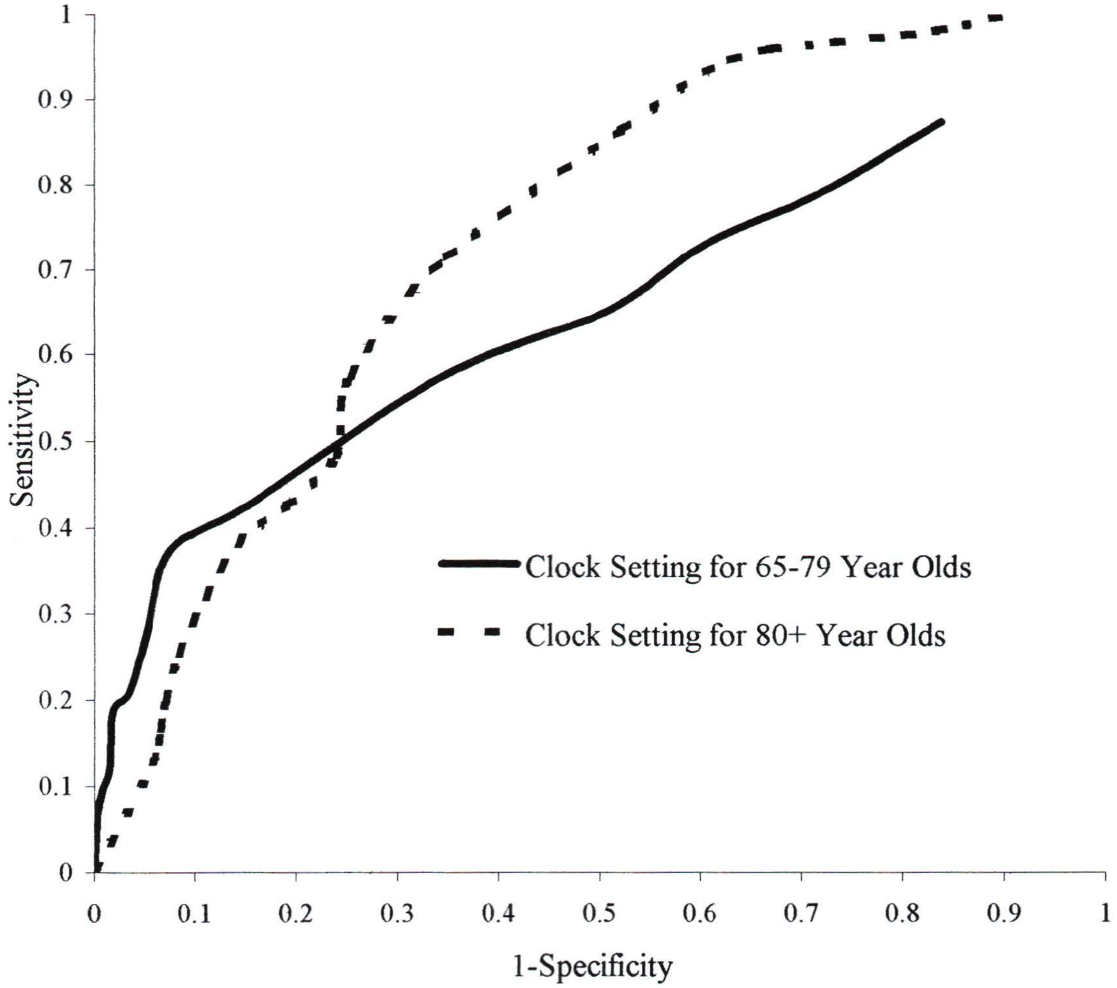


Figure 14. ROCs of the Clock Setting task for each age group.

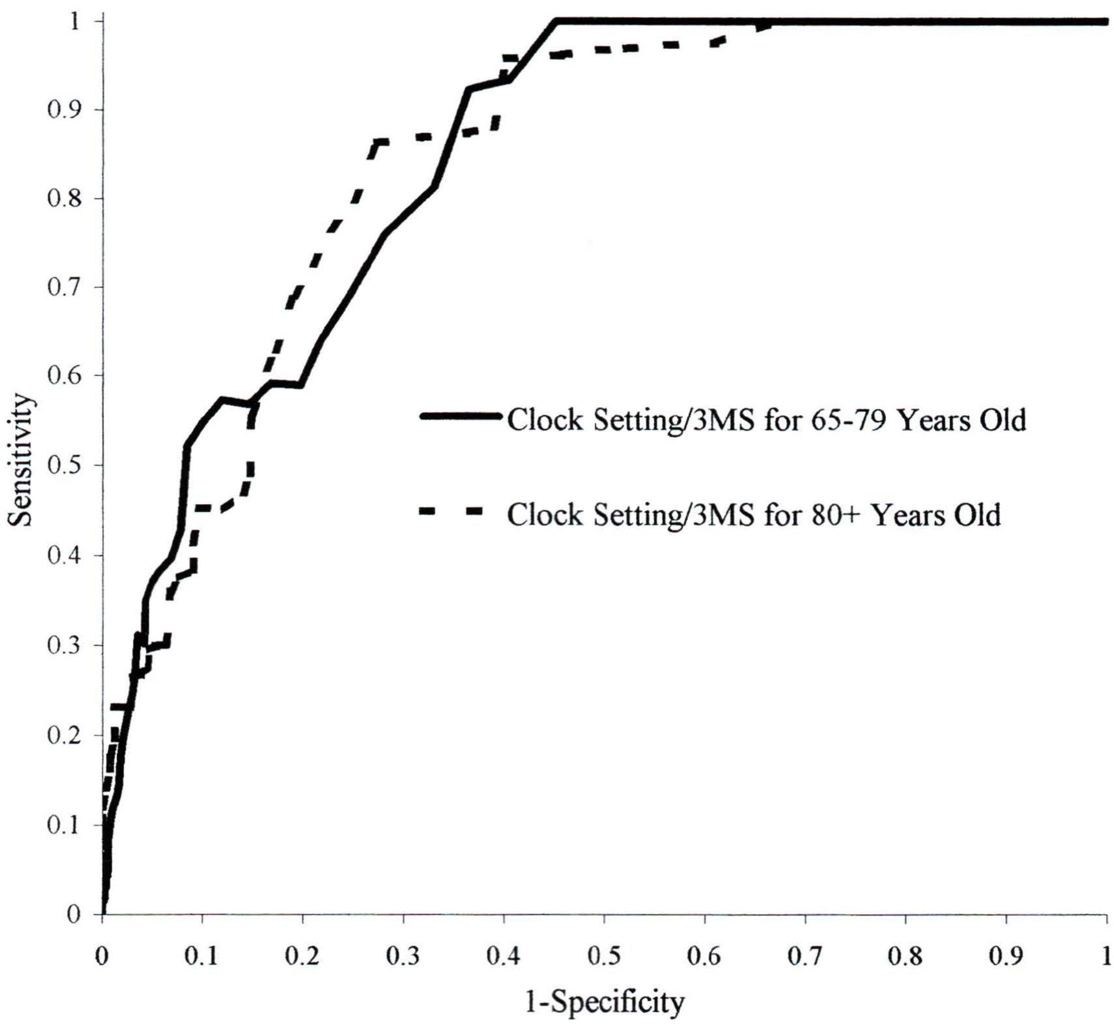


Figure 15. ROCs of the combined Clock Setting/3MS task for each age group.

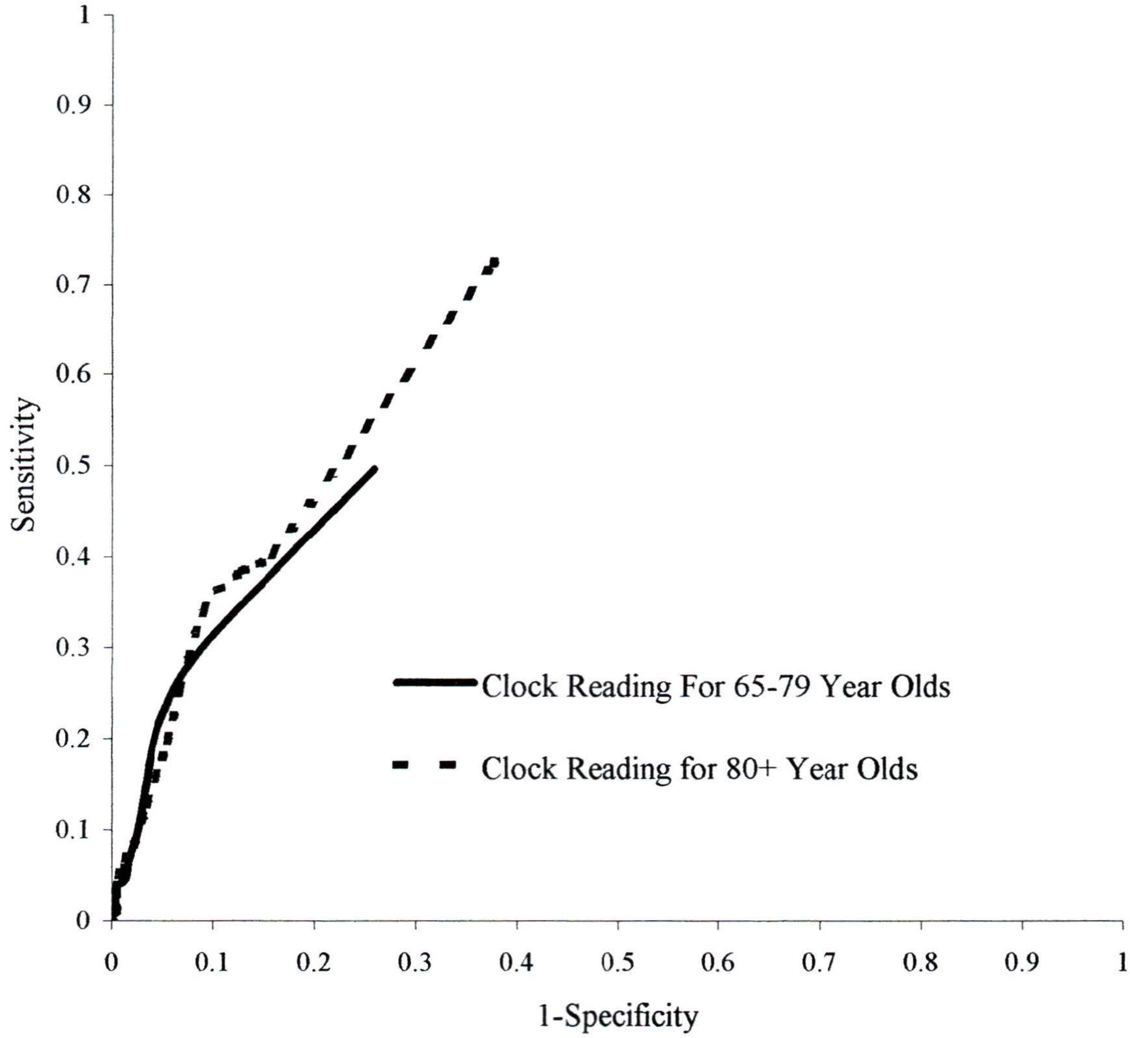


Figure 16. ROCs of the Clock Reading task for each age group.

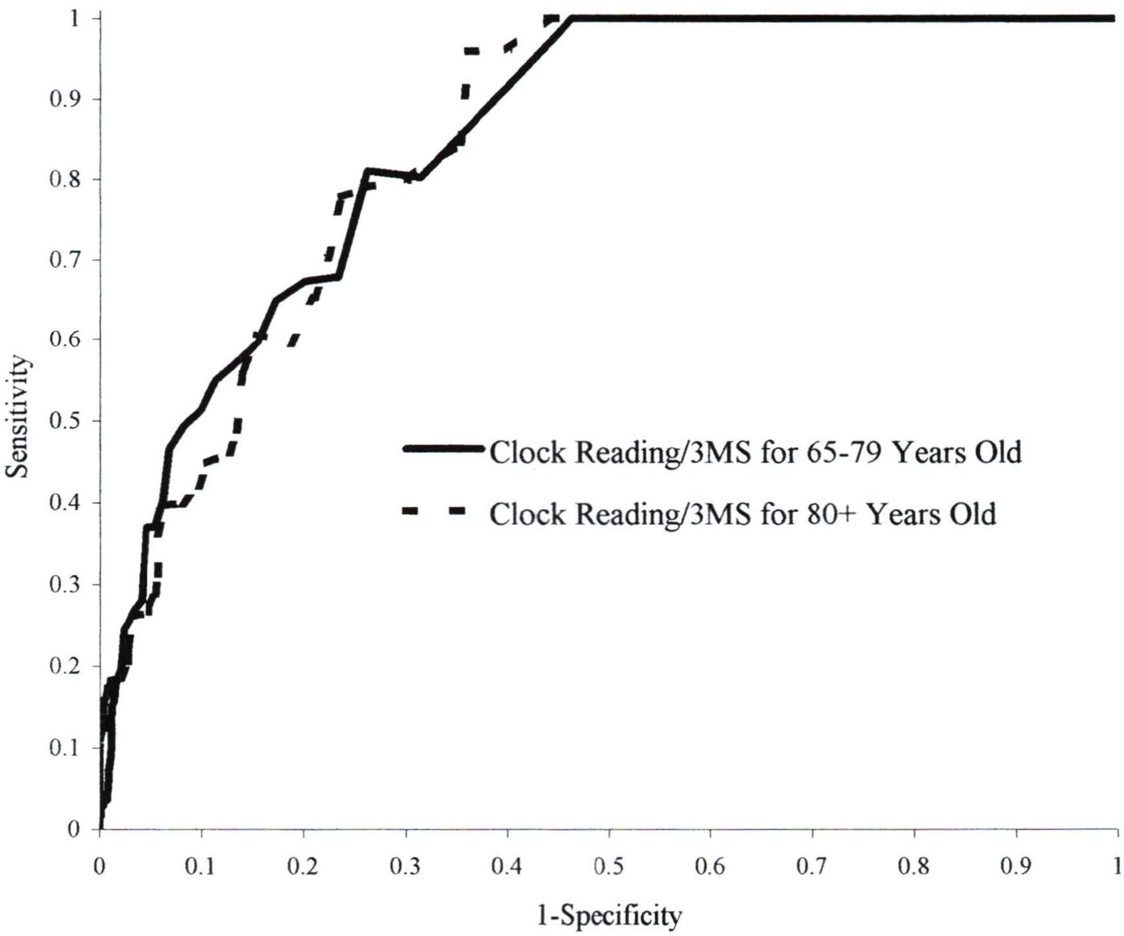


Figure 17. ROCs of the combined Clock Reading/3MS task for each age group.

Accuracy of Screening Across Education Groups

The average 3MS, Clock Drawing – Reversed, Clock Setting, and Clock Reading scores were significantly greater for participants with 9 or more years of education than for participants with 8 or fewer years of education (3MS t_{1979}= 16.18, $p < 0.001$; Clock Drawing - Reversed $t_{1979} = 5.51$, $p < 0.001$; Clock Setting $t_{1979} = 9.56$, $p < 0.001$; Clock Reading $t_{1979} = 7.67$, $p < 0.001$). Similarly, the average combined scores were significantly greater for participants with 9 or more years of education than for participants with 8 or fewer years of education (Clock Drawing – Reversed/3MS t_{1979} = 15.09, $p < 0.001$; Clock Setting/3MS $t_{1979} = 16.04$, $p < 0.001$; Clock Reading/3MS t_{1979} = 16.00, $p < 0.001$). The means and standard deviations of the performance on each screening test for participants with 8 or fewer years of education and those with 9 or more years of education are presented in Table 6.

The AUCs, standard errors (*SE*s), *z*-values, and if relevant, percent difference in accuracy of the screening tools for the two education groups are presented in Table 7. Clock Setting and Clock Reading were significantly more accurate for participants with eight or fewer years of education than for participants with nine or more years of education. All other measures were equally accurate for both education groups. Figure 18 illustrates the ROC plot of the 3MS for both education groups. Figure 19 shows ROC plots for the Clock Drawing – Reversed and Figure 20 shows the ROC plots of the combined Clock Drawing – Reversed/3MS tests for each education group. Figure 21 shows the Clock Setting ROC plots and Figure 22 shows the ROC plots of the combined Clock Setting/3MS for both education groups. Figure 23 illustrates the ROC plots of

Table 6. Means (*M*s) and Standard Deviations (*SD*s) of the Screening Tools for Each Education Group.

Screening Measure	0 - 8 Years Education (*n* = 451)		9+ Years Education (*n* = 1532)	
	M	*SD*	*M*	*SD*
3MS	81.02	12.22	89.27	8.57
Clock Drawing - Reversed	45.51	5.24	46.94	4.75
Clock Drawing – Reversed/3MS	126.52	15.09	136.21	10.91
Clock Setting	8.86	4.18	10.72	3.46
Clock Setting/3MS	89.87	14.94	99.99	10.67
Clock Reading	12.61	3.27	13.70	2.43
Clock Reading/3MS	93.63	14.12	102.97	9.75

Table 7. The AUCs, Standard Errors (*SE*s), z-values, and the Difference in Percent (Diff in %) of Each Screening Tool for Each Education Group.

Screening Measure	AUC	*SE*	*z*-Value	Diff in %
3MS				
0-8 Years Education	0.8524	0.0412	0.11	N/A
9+ Years Education	0.8469	0.02968		
Clock Drawing – Reversed				
0-8 Years Education	0.4864	0.0477	0.05	N/A
9+ Years Education	0.4413	0.0291		
Clock Drawing – Reversed/3MS				
0-8 Years Education	0.8416	0.0372	-0.19	N/A
9+ Years Education	0.8508	0.0320		
Clock Setting				
0-8 Years Education	0.6865	0.0424	3.21*	18.6 %
9+ Years Education	0.5001	0.0396		
Clock Setting/3MS				
0-8 Years Education	0.8485	0.0335	-0.26	N/A
9+ Years Education	0.8600	0.0300		
Clock Reading				
0-8 Years Education	0.1586	0.0229	2.91*	7.3 %
9+ Years Education	0.0858	0.0101		
Clock Reading/3MS				
0-8 Years Education	0.8364	0.0424	-0.41	N/A
9+ Years Education	0.8578	0.0306		

* $p < 0.005$ two-tailed.

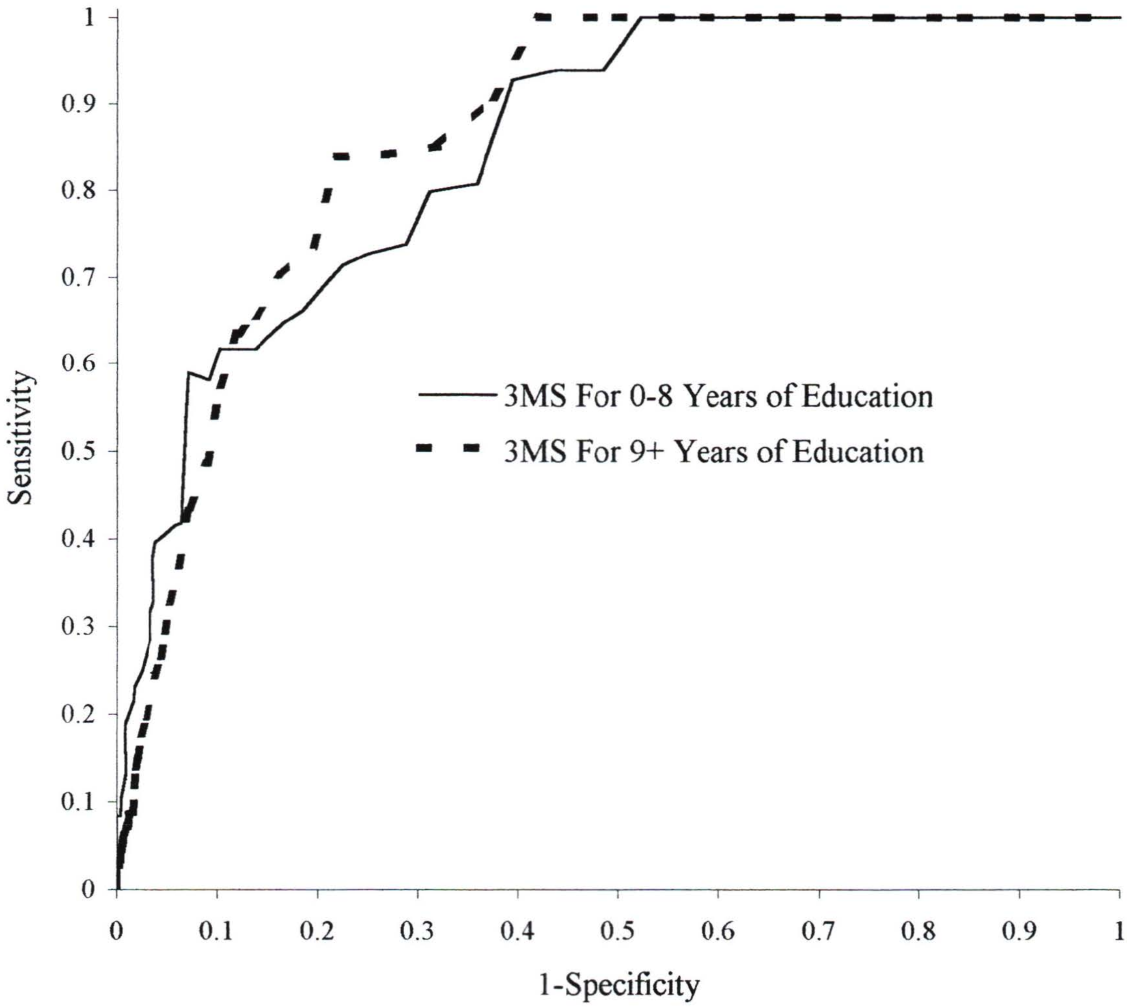


Figure 18. ROCs of the 3MS for each education group.

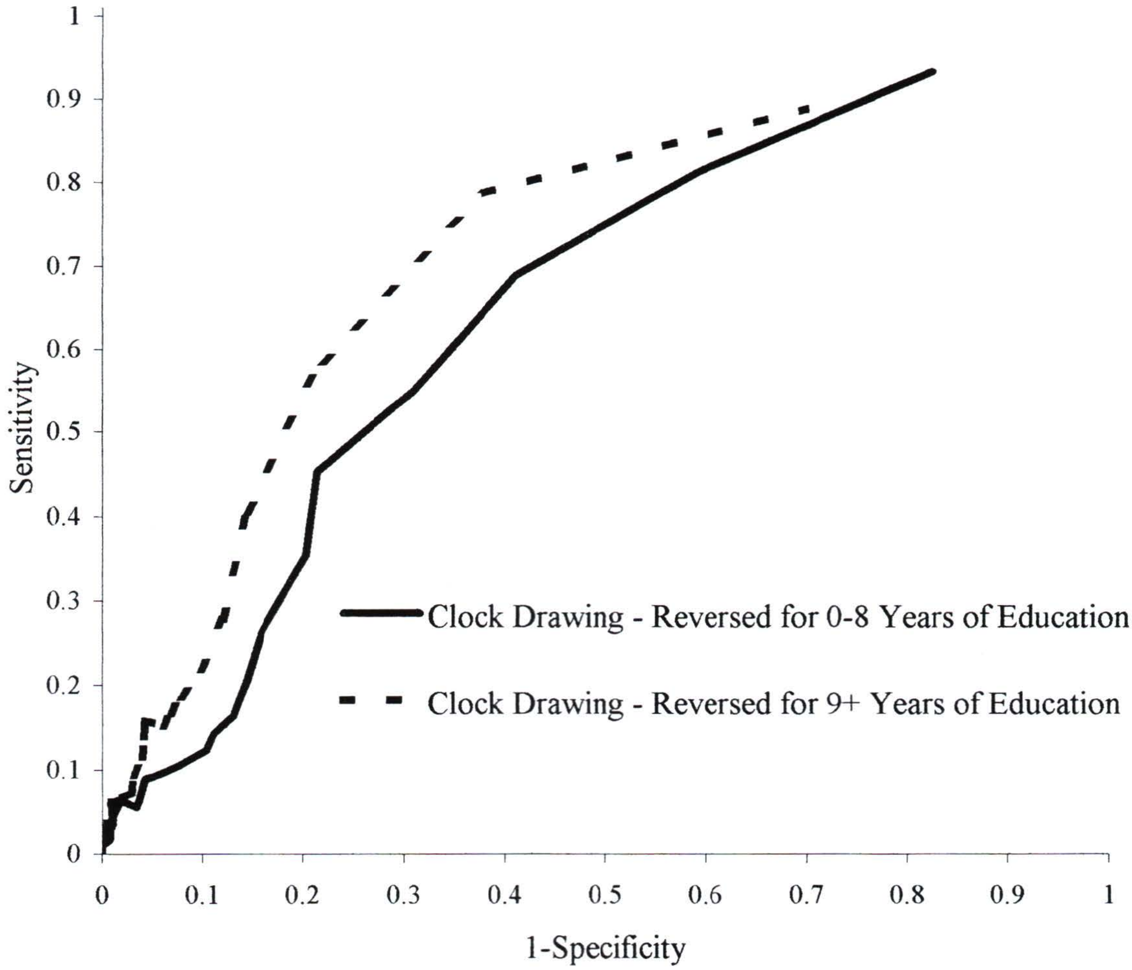


Figure 19. ROCs of the Clock Drawing – Reversed task for each education group

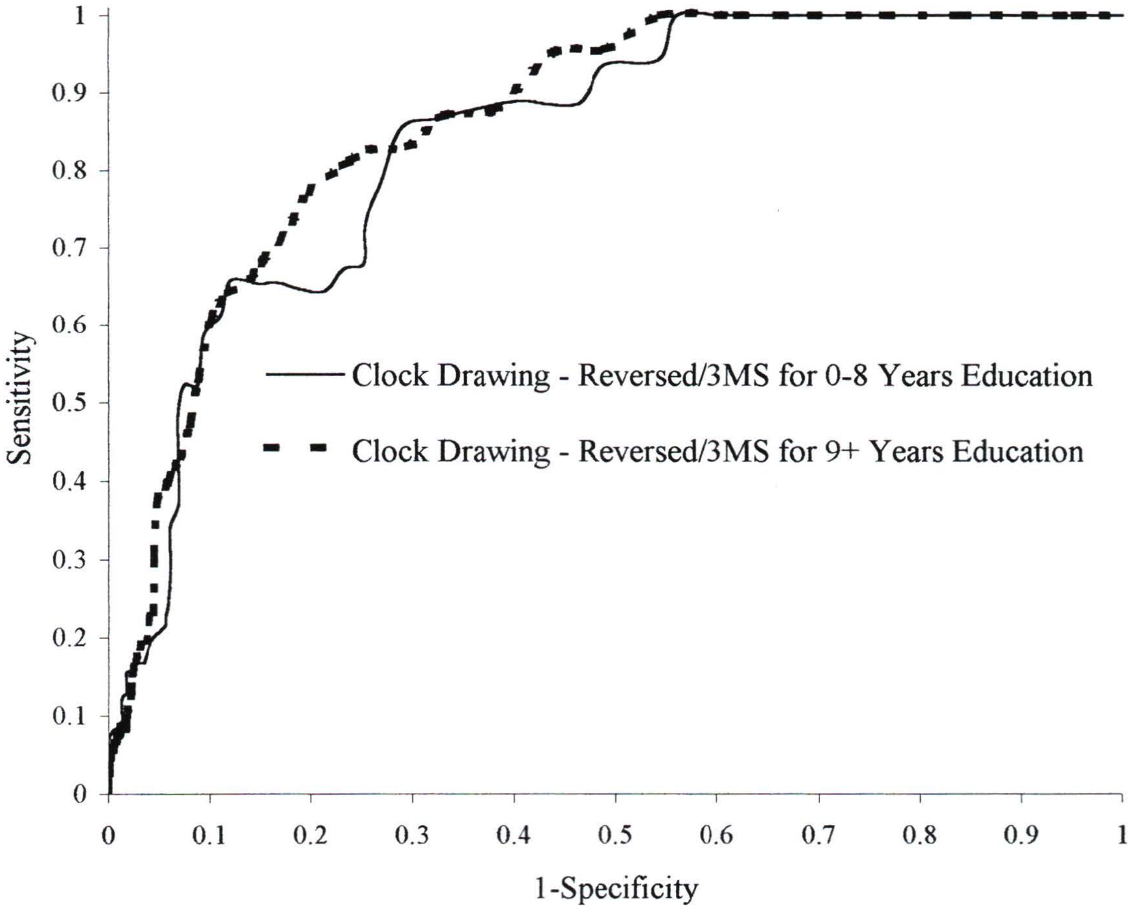


Figure 20. ROCs of the combined Clock Drawing – Reversed/3MS task for each education group.

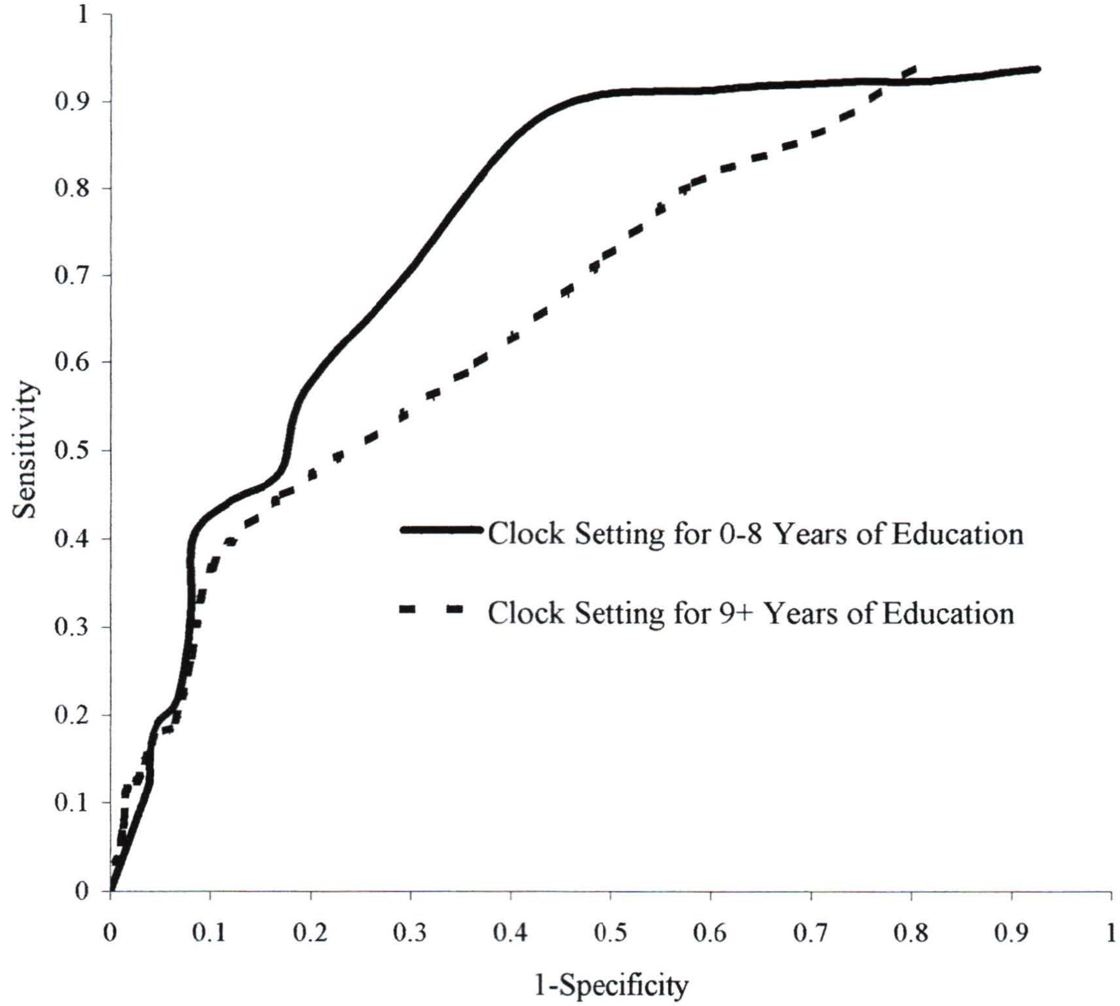


Figure 21. ROCs of the Clock Setting task for each education group.

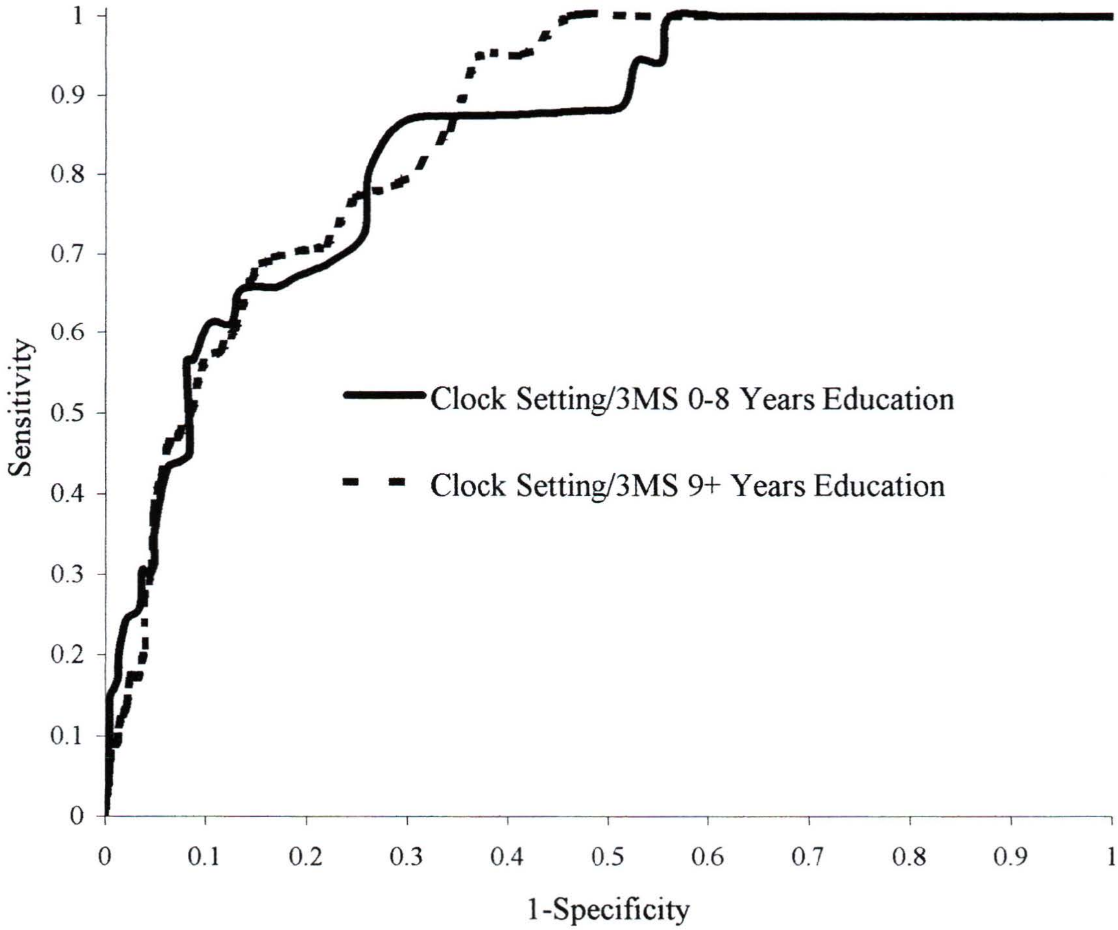


Figure 22. ROCs of the combined Clock Setting/3MS task for each education group.

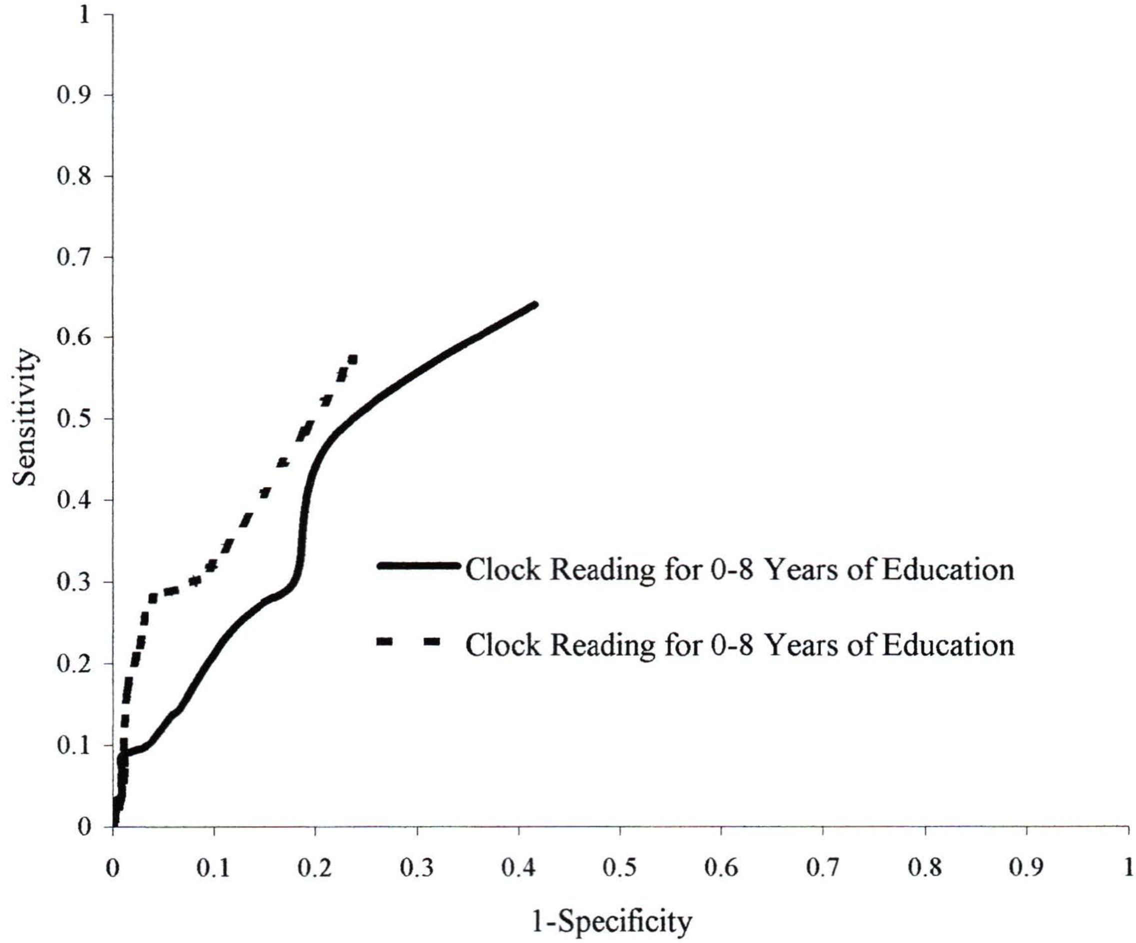


Figure 23. ROCs of the Clock Reading task for each education group.

Clock Reading and Figure 24 illustrates the ROC plots of the combined Clock Reading/3MS for both education groups.

Accuracy of Screening Across Sex Groups

The average 3MS, combined Clock Drawing – Reversed/3MS and combined Clock Reading/3MS scores were significantly greater for female participants than for male participants (3MS $t_{1979} = 3.47$, $p < 0.001$; Clock Drawing – Reversed/3MS $t_{1979} = 2.62$, $p = 0.009$, Clock Reading/3MS $t_{1979} = 2.06$, $p = 0.040$). However, the average Clock Setting and Clock Reading scores were significantly greater for males than for females (Clock Setting $t_{1979} = 4.40$, $p < 0.001$; Clock Reading $t_{1979} = 4.27$, $p < 0.001$). Finally, the average Clock Drawing – Reversed and combined Clock Setting/3MS scores were equivalent for males and females (Clock Drawing – Reversed $t_{1979} = 0.42$, $p = 0.672$; Clock Setting/3MS $t_{1979} = -1.52$, $p = 0.128$). The means and standard deviations of the performance on each screening test for males and females are presented in Table 8.

The Clock Drawing – Reversed was more accurate at identifying participants with and without cognitive impairment for females than males. The 3MS, combined Clock Drawing/3MS, and combined Clock Reading/3MS were marginally more accurate for females than for males. All other measures were equally accurate at identifying participants with and without cognitive impairment for males and females. The AUCs, *SE*s, *z*-values, and % differences (if applicable) for each screening measure are presented in Table 9. Figure 25 illustrates the ROC plots of the 3MS for males and females. Figure 26 shows ROC plots for the Clock Drawing – Reversed and Figure 27 shows ROC plots

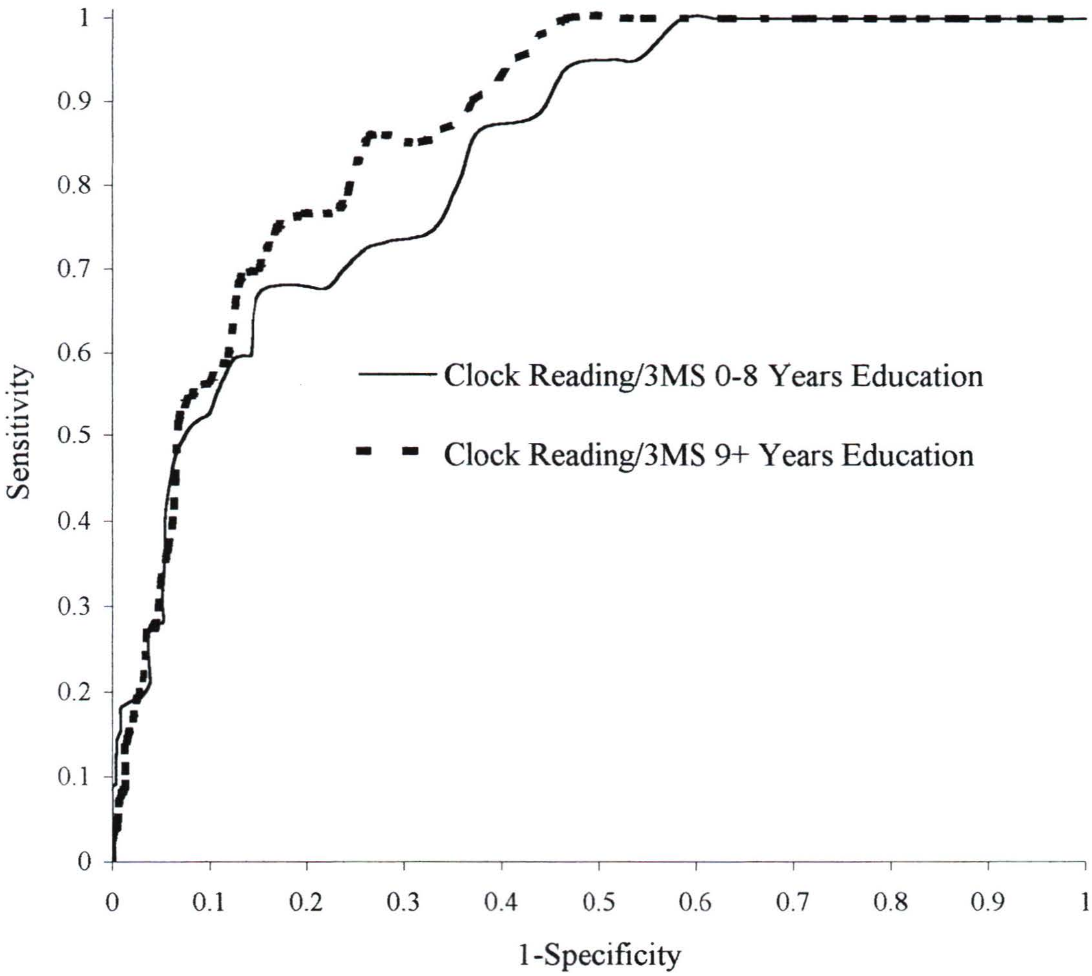


Figure 24. ROCs of the combined Clock Reading/3MS task for each education group.

Table 8. Mean Scores (*M*s) and Standard Deviations (*SD*s) of Screening Tools for Each Sex Group.

Screening Measure	Males (n = 883)		Females (n = 1100)	
	M	*SD*	*M*	*SD*
3MS	86.51	10.37	88.10	9.87
Clock Drawing - Reversed	46.67	5.06	46.57	4.77
Clock Drawing – Reversed/3MS	133.18	12.57	134.67	12.69
Clock Setting	10.70	3.57	9.97	3.80
Clock Setting/3MS	97.21	12.63	98.07	12.41
Clock Reading	13.741	2.33	13.22	2.91
Clock Reading/3MS	100.25	11.59	101.32	11.55

Table 9. The AUCs, Standard Errors (*SE*s), *z*-values, and the Difference in Percent (Diff in %) of Screening Tool for Each Sex Group.

Screening Measure	AUC	*SE*	*z*-Value	Diff in %
3MS				
Males	0.8107	0.0374	-1.76[a]	7.8%
Females	0.8885	0.0235		
Clock Drawing – Reversed				
Males	0.3925	0.0332	-2.29[b]	11.2%
Females	0.5045	0.0359		
Clock Drawing – Reversed/3MS				
Males	0.8109	0.04065	-1.73[c]	8.4%
Females	0.8946	0.0262		
Clock Setting				
Males	0.5449	0.0377	-0.31	N/A
Females	0.5642	0.0482		
Clock Setting/3MS				
Males	0.8273	0.0371	-1.36	N/A
Females	0.8929	0.0307		
Clock Reading				
Males	0.0842	0.0115	-1.36	N/A
Females	0.1083	0.0134		
Clock Reading/3MS				
Males	0.8163	0.0379	-1.70[d]	7.7%
9+ Years Education	0.8932	0.0247		

[a] $p = 0.078$ two-tailed. [b] $p = 0.022$ two-tailed. [c] $p = 0.084$ two-tailed. [d] $p = 0.092$ two-tailed.

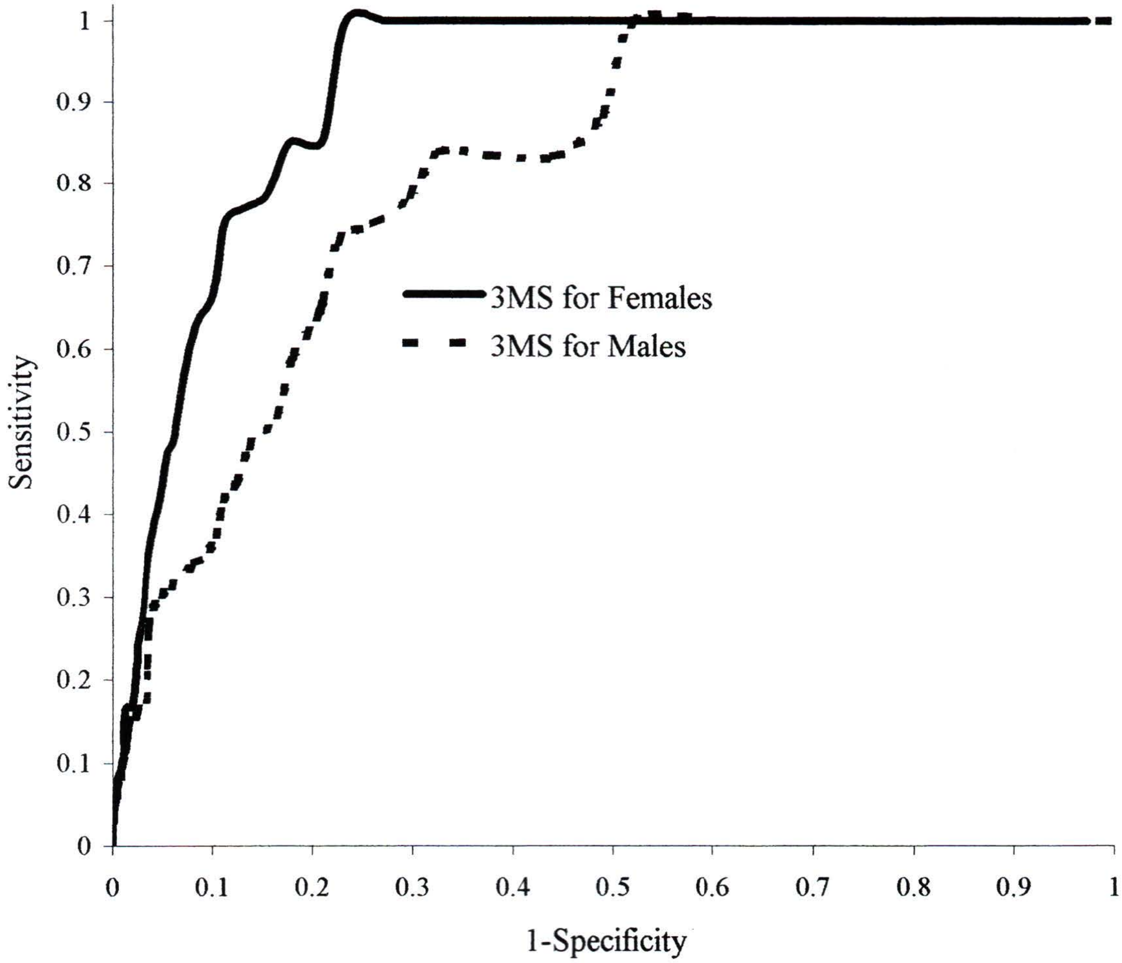


Figure 25. ROCs of the 3MS for each sex group.

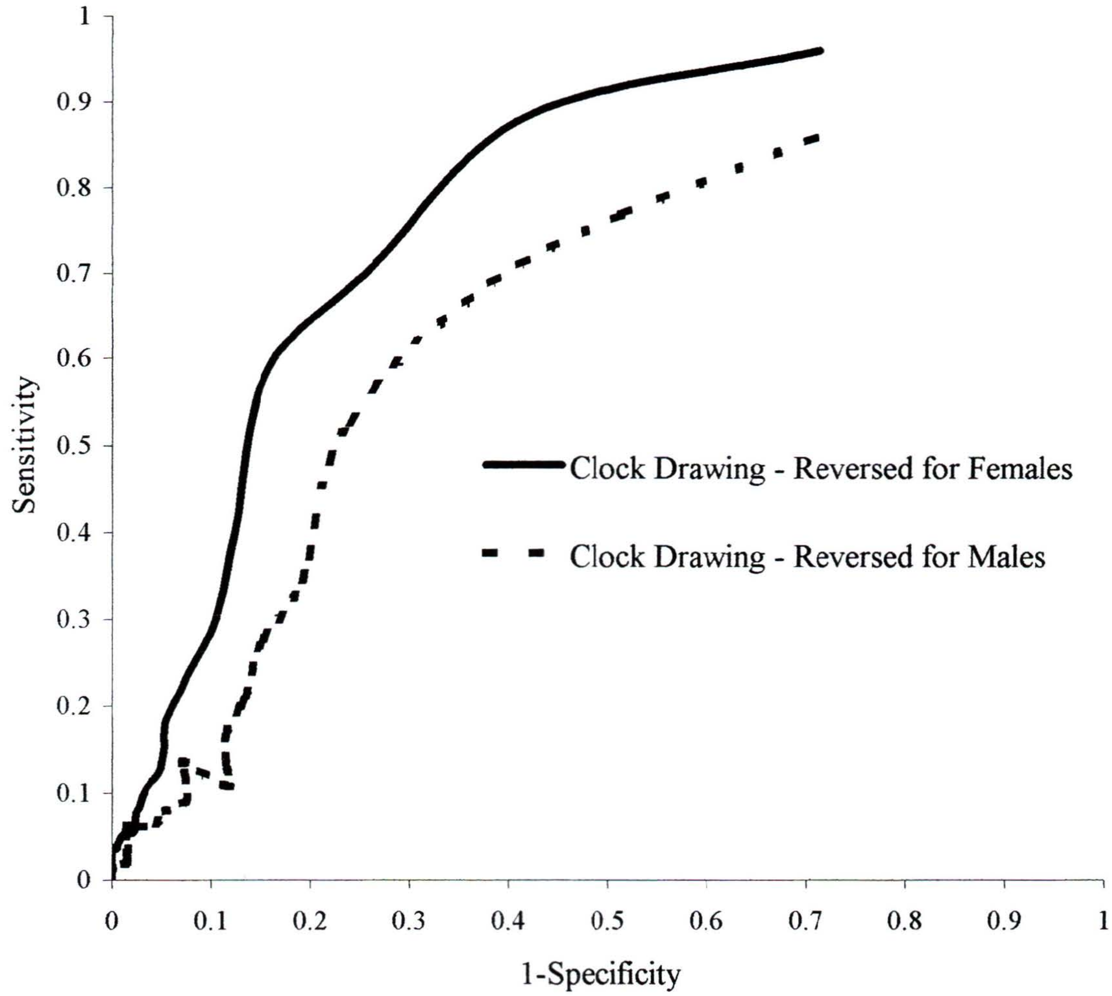


Figure 26. ROCs of the Clock Drawing – Reversed task for each sex group.

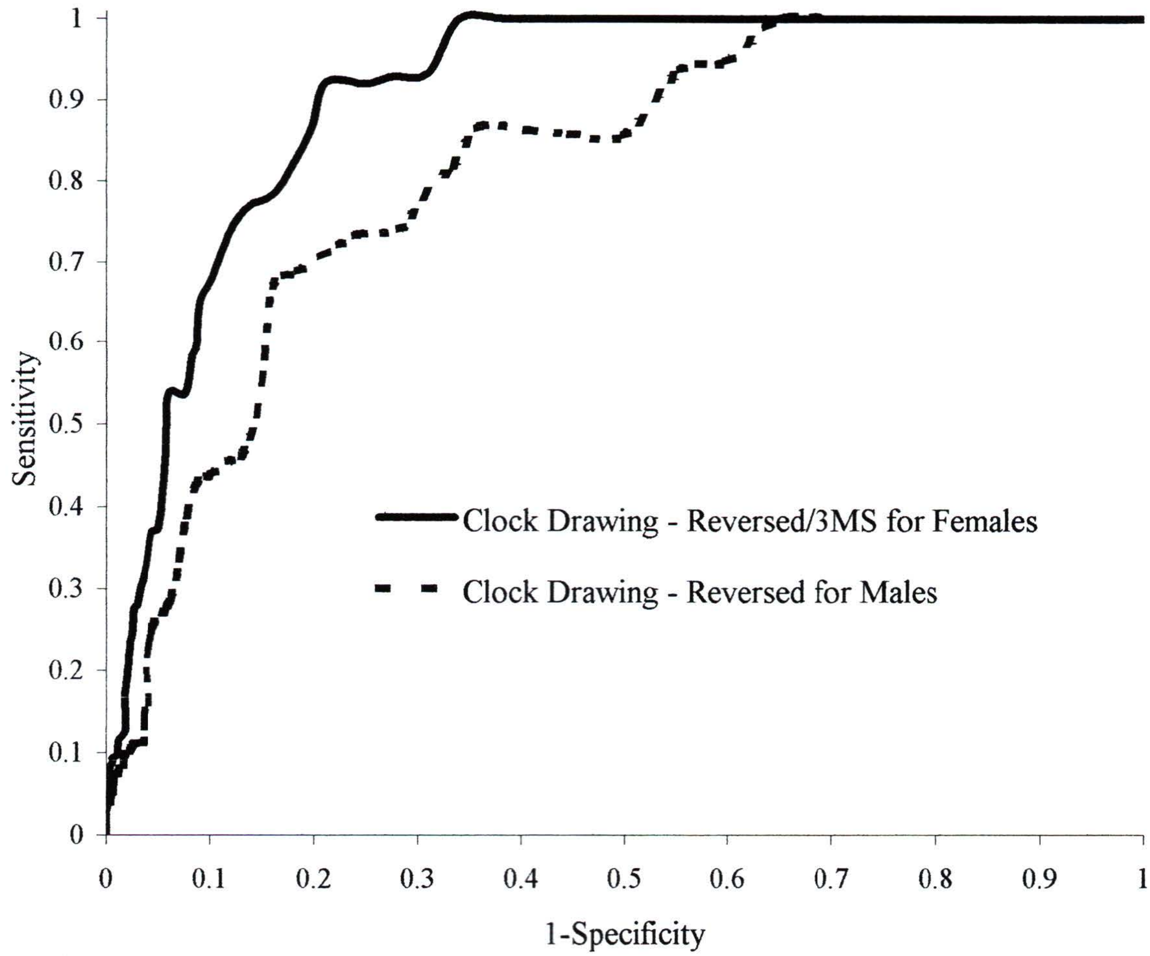


Figure 27. ROCs of the combined Clock Drawing - Reversed/3MS task for each sex group.

plots for Clock Setting and Figure 29 shows ROC plots for the combined Clock Setting/3MS for males and females. Figure 30 shows ROC plots for the Clock Reading task and Figure 31 shows ROC plots for the combined Clock Setting/3MS for both sex groups.

Combined Versus Single Screening Tools for Age Groups

I wanted to know if components of the Clock Test added to the accuracy of the 3MS for persons from different age groups. However, the results were similar for both age groups: the combined Clock Drawing – Reversed/3MS was equally accurate as the 3MS task alone but more accurate than the Clock Drawing task alone. Similarly, the combined Clock Setting/3MS was equally accurate to the 3MS alone but more accurate than the Clock Setting task alone for both the younger and the older age groups. Finally, the combined Clock Reading/3MS was more accurate than the Clock Reading task alone but equally accurate to the 3MS alone. The AUCs, *SE*s, *z*-values, and percent difference (if applicable) of the combined versus single screening tools are shown in Table 10. Figure 32 shows the ROC plots for the younger age group and Figure 33 shows the ROC plots for the older age group for the 3MS, the Clock Drawing – Reversed task, and the combined Clock Drawing – Reversed/3MS tasks. Figure 34 shows the ROC plots for the younger age group and Figure 35 shows the ROC plots for the older age group for the 3MS, the Clock Setting task, and the combined Clock Setting/3MS tasks. Finally, Figure 36 shows the ROC plots for the younger age group and Figure 37 shows the ROC plots for the older age group for the 3MS, the Clock Reading task, and the combined Clock Reading/3MS.

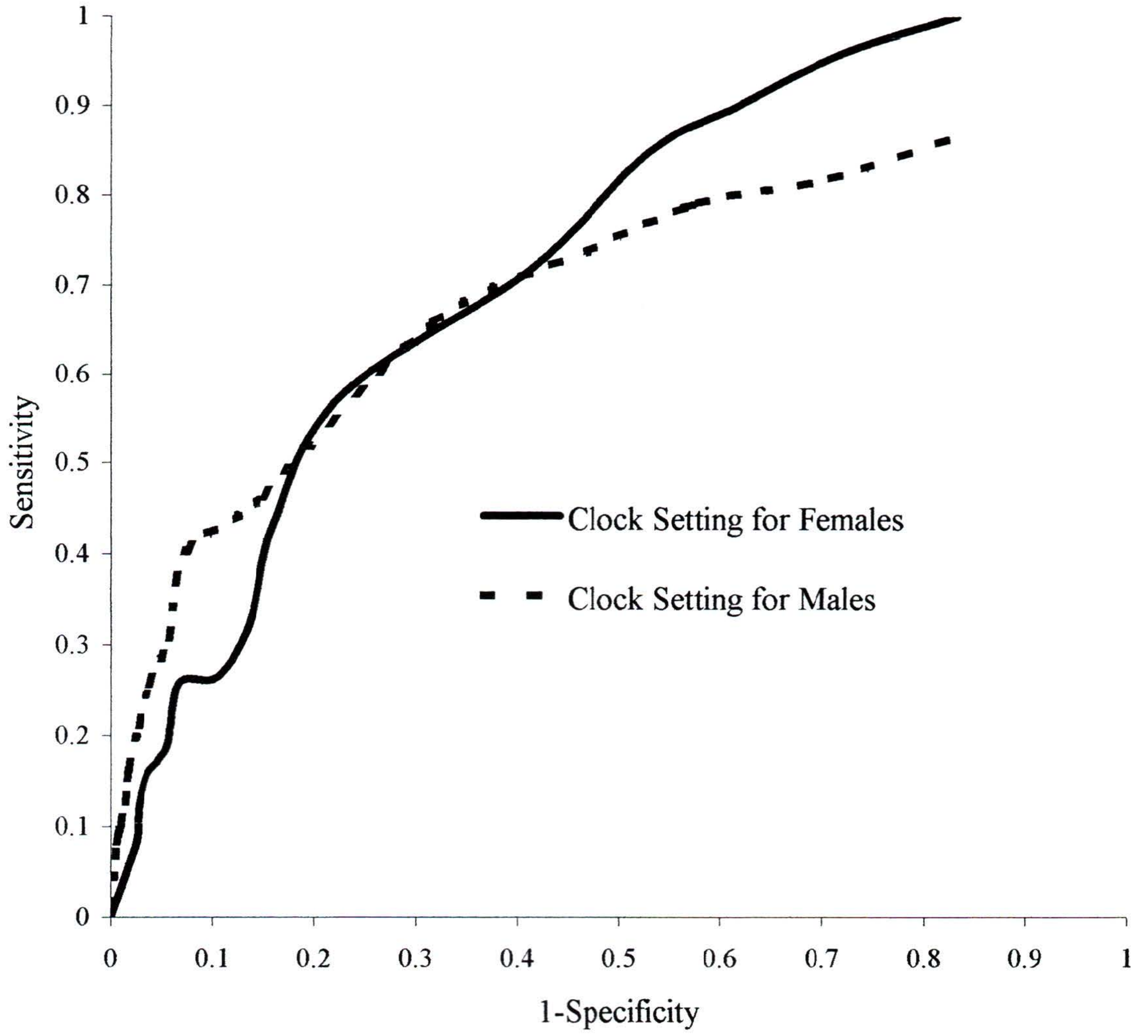


Figure 28. ROCs of the Clock Setting task for each sex group.

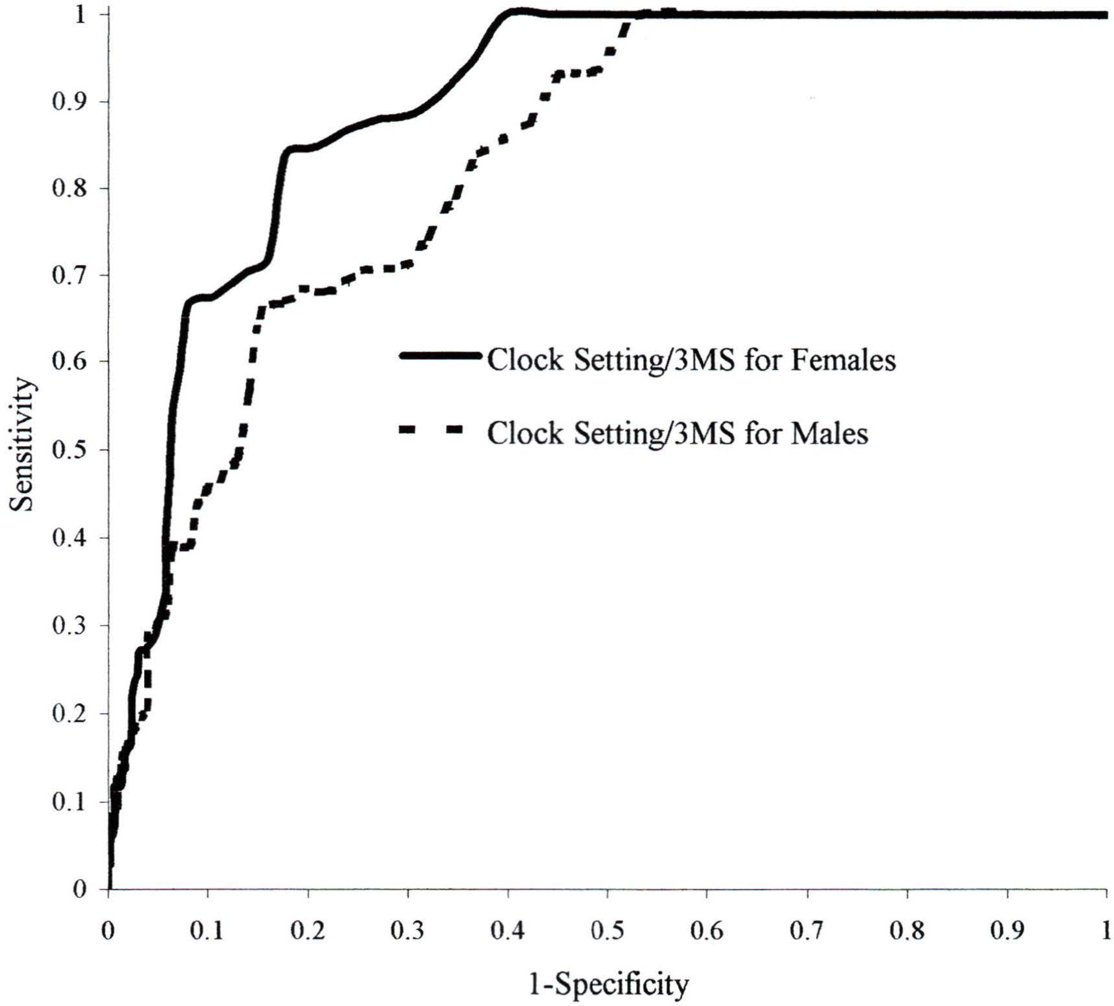


Figure 29. ROCs of the combined Clock Setting/3MS task for each sex group.

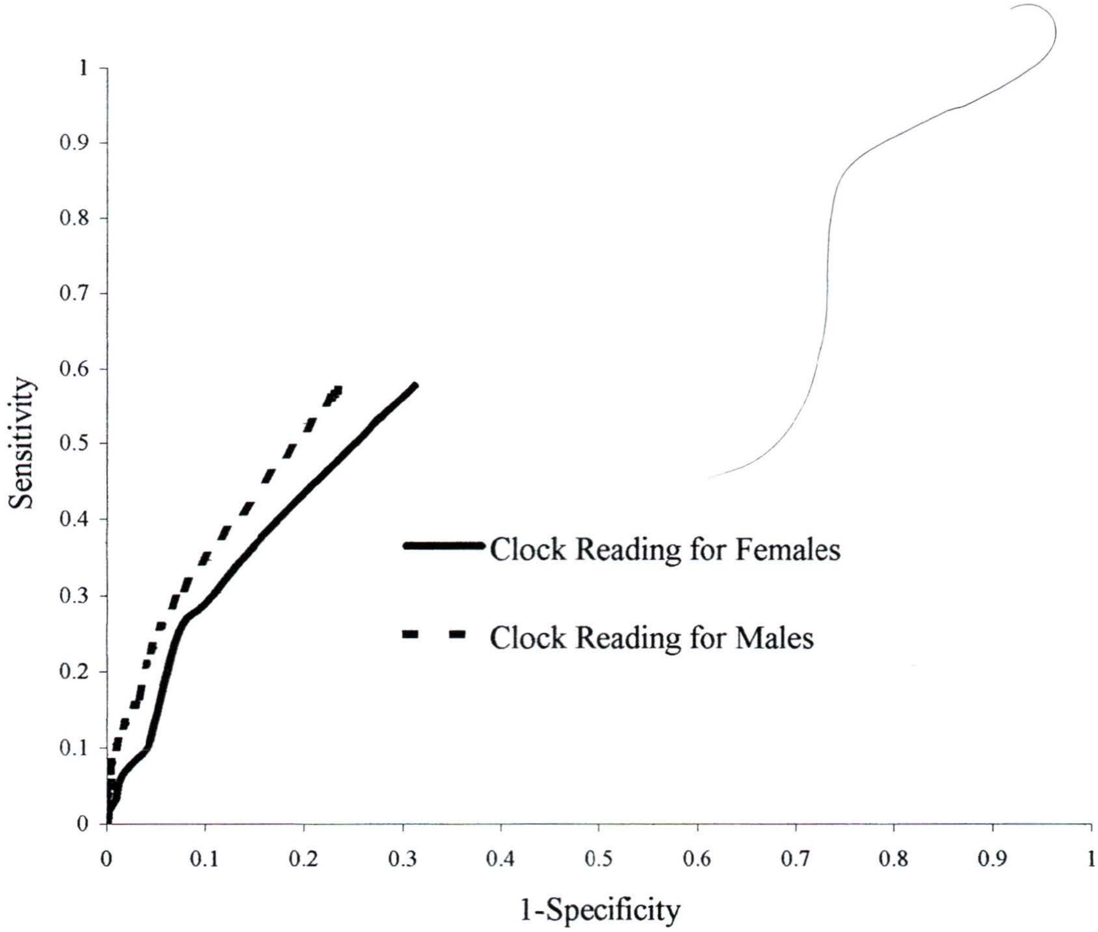


Figure 30. ROCs of the Clock Reading task for each sex group.

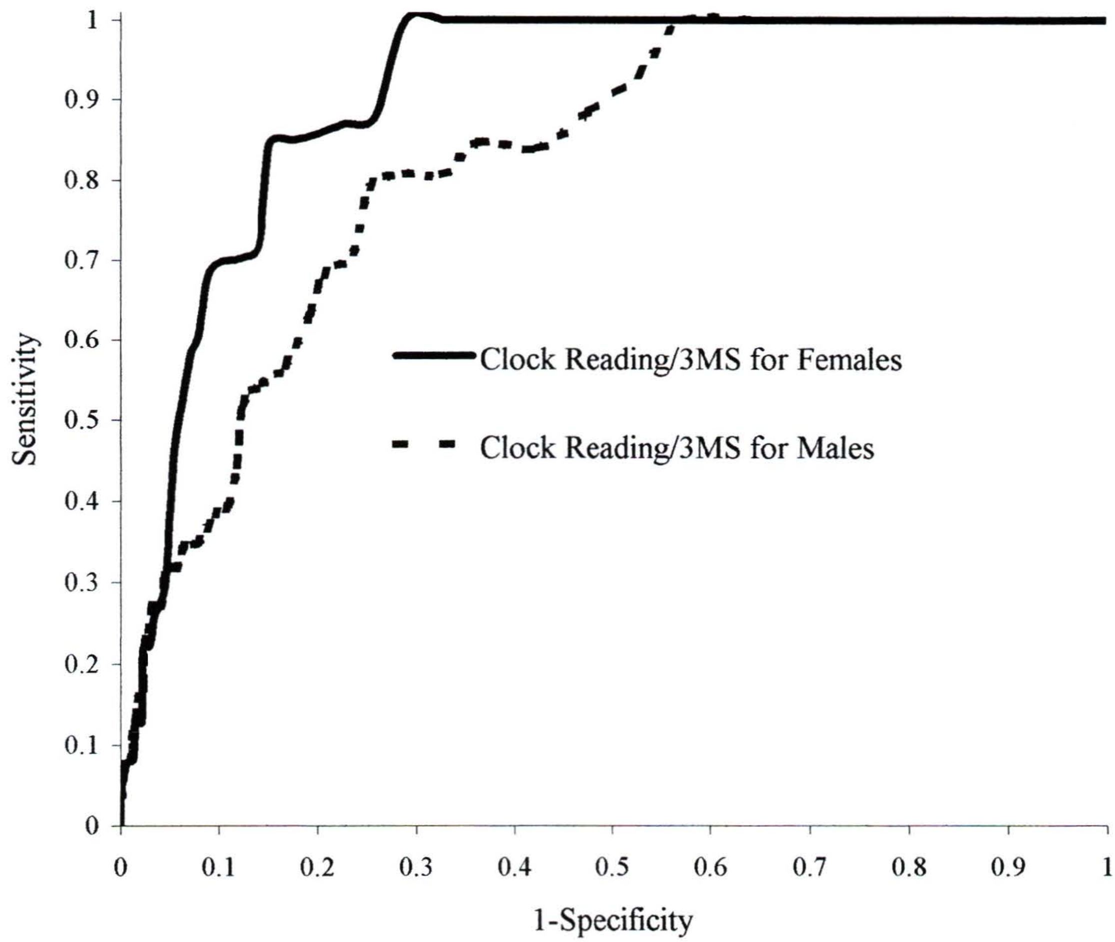


Figure 31. ROCs of the combined Clock Reading/3MS task for each sex group.

Table 10. The AUCs, Standard Errors (*SE*s), *z*-values, and, Difference in Percent (Diff %) of the Screening Tools for Each Age Group.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
65 – 79 Years Old				
3MS	0.8372	0.0366		
			0.03	N/A
Clock Drawing – Reversed/3MS	0.8367	0.0401		
			8.17*	42.7%
Clock Drawing - Reversed	0.4089	0.0338		
80+ Years Old				
3MS	0.8330	0.0304		
			-0.26	N/A
Clock Drawing – Reversed/3MS	0.8370	0.0296		
			5.94*	30.1%
Clock Drawing - Reversed	0.5353	0.0413		
65 – 79 Years Old				
3MS	0.8372	0.0366		
			-0.45	N/A
Clock Setting/3MS	0.8455	0.0350		
			6.23*	34.9%
Clock Setting	0.4965	0.0438		
80+ Years Old				
3MS	0.8330	0.0304		
			-0.77	N/A
Clock Setting/3MS	0.8451	0.0311		
			6.51*	20.5%
Clock Setting	0.6404	0.0416		

Table continues. * $p < 0.001$ two-tailed.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
65 – 79 Years Old				
3MS	0.8372	0.0366		
			-0.13	N/A
Clock Reading/3MS	0.8398	0.0378		
			19.35*	75.7%
Clock Reading	0.0831	0.0101		
80+ Years Old				
3MS	0.8330	0.0304		
			-0.47	N/A
Clock Reading/3MS	0.8400	0.0266		
			19.80*	67.5%
Clock Reading	0.1651	0.0213		

* $p < 0.001$ two-tailed.

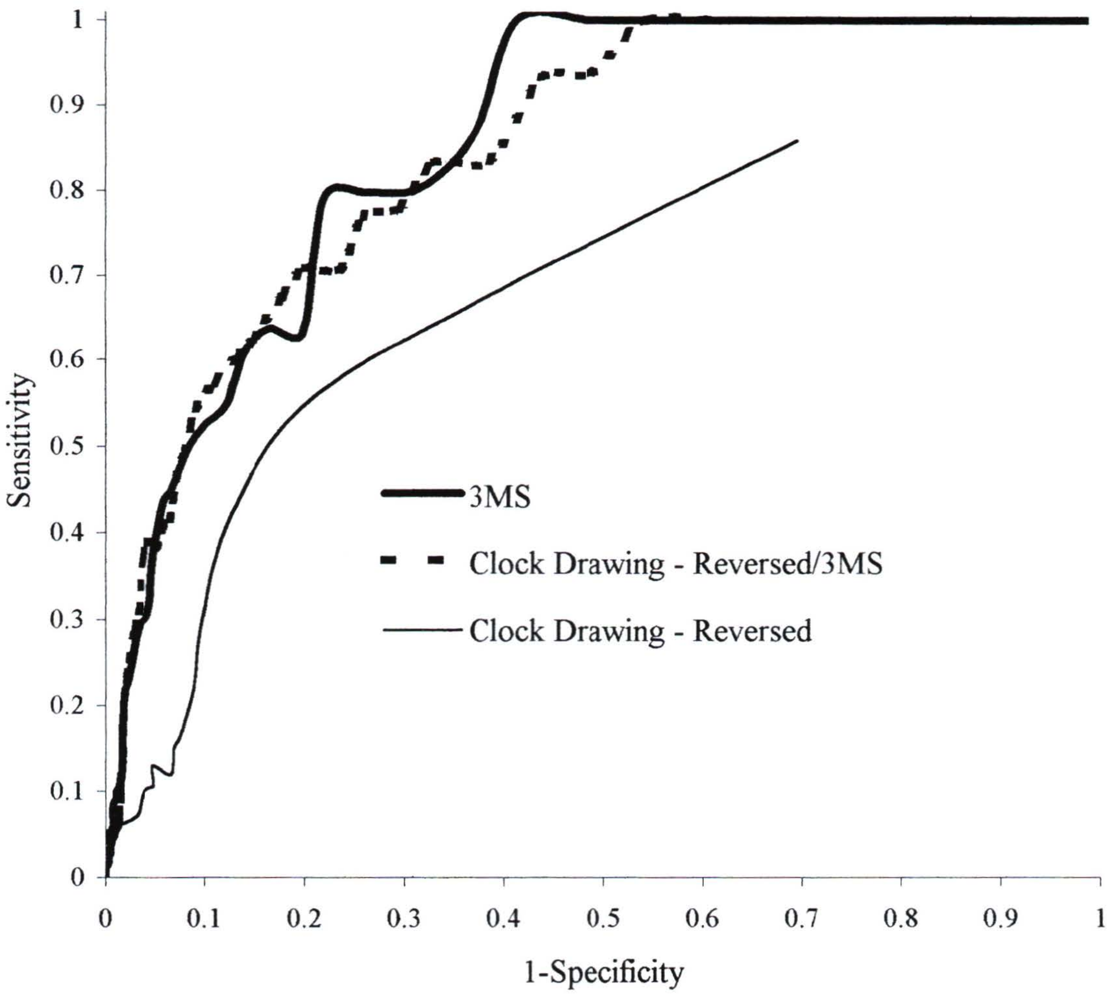


Figure 32. ROCs of the 3MS, Clock Drawing – Reversed, and Combined Clock Drawing – Reversed/3MS tasks for 65-79 year olds.

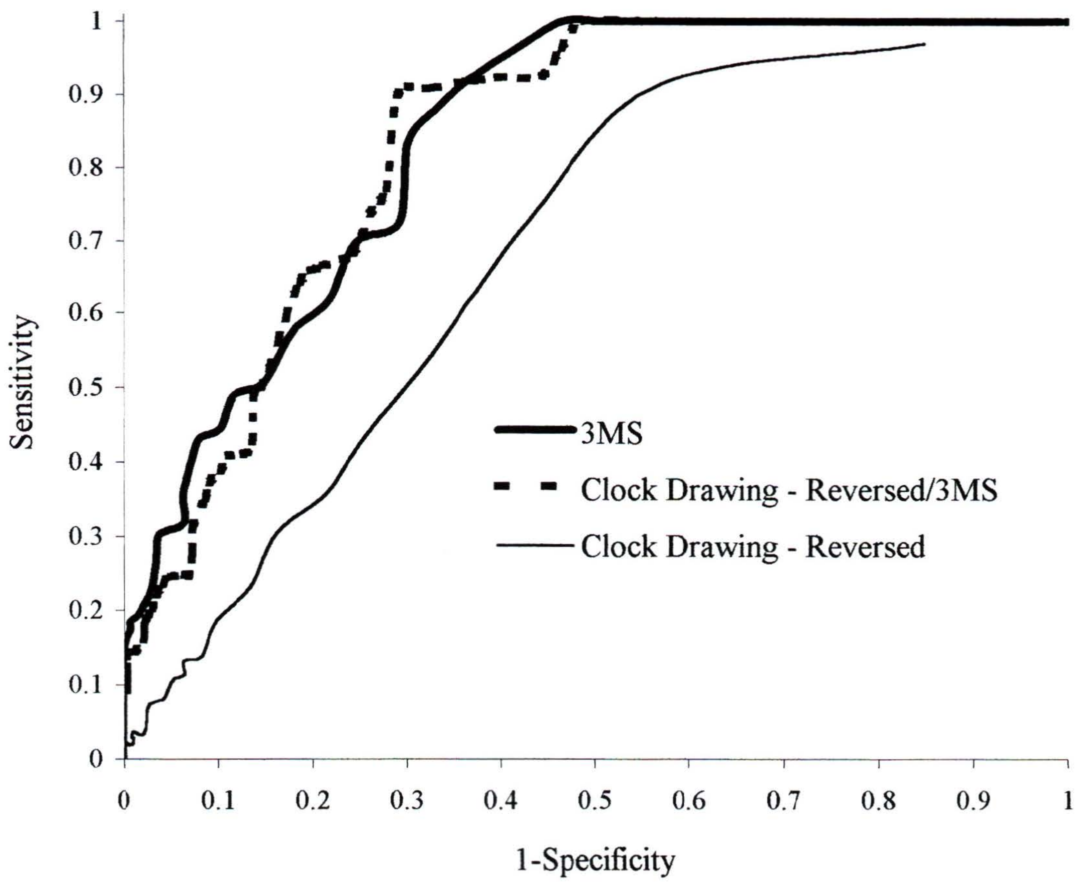


Figure 33. ROCs of the 3MS, Clock Drawing–Reversed, and Combined Clock Drawing–Reversed/3MS tasks for 80+ year olds.

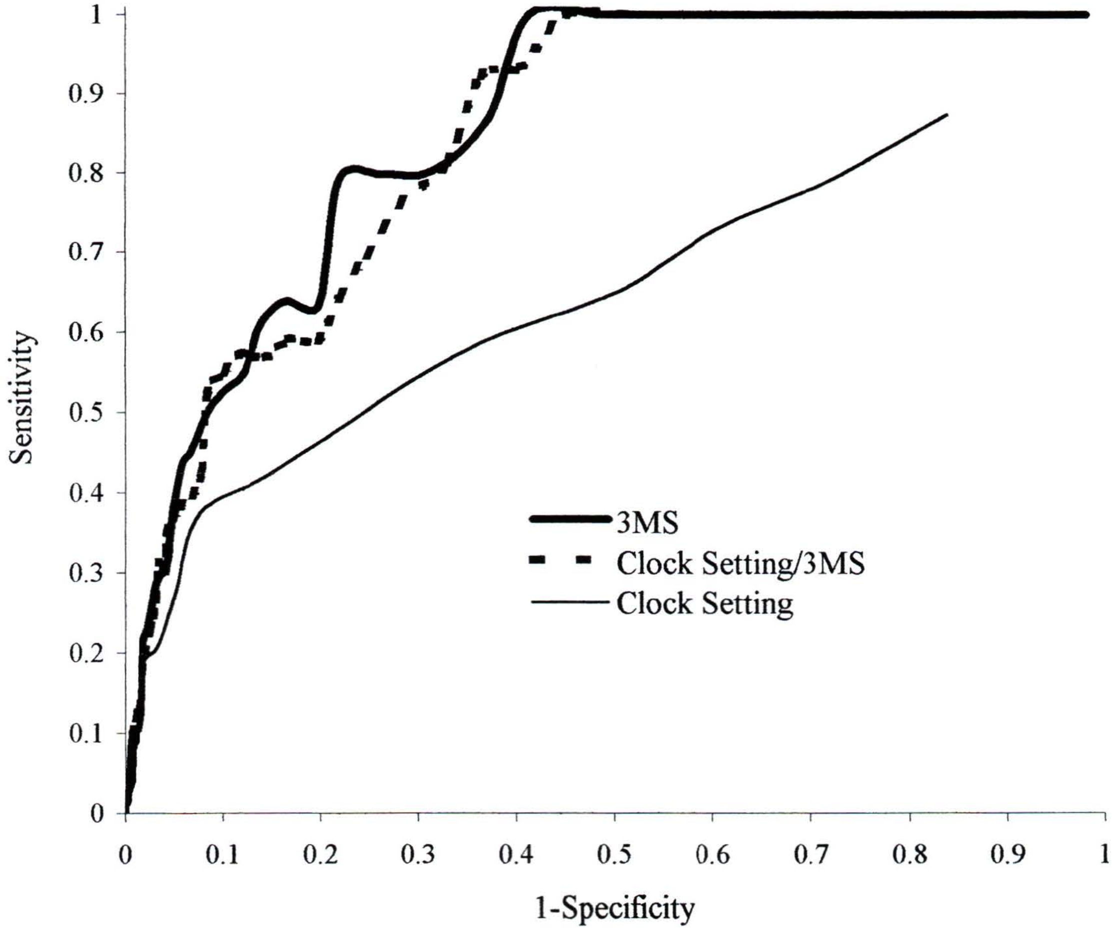


Figure 34. ROCs of the 3MS, Clock Setting, and Combined Clock Setting/3MS tasks for 65-79 year olds.

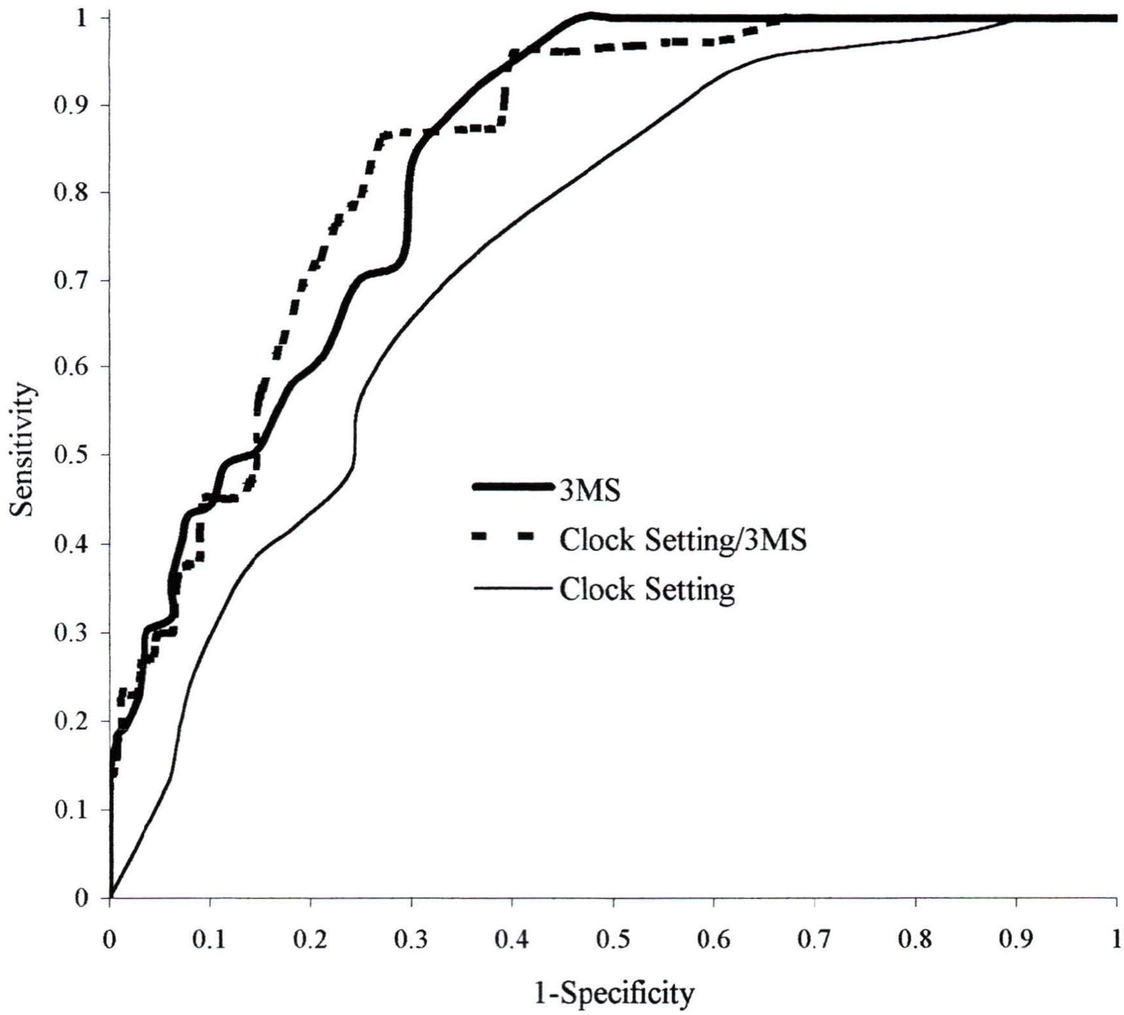


Figure 35. ROCs of the 3MS, Clock Setting, and Combined Clock Setting/3MS tasks for 80+ year olds.

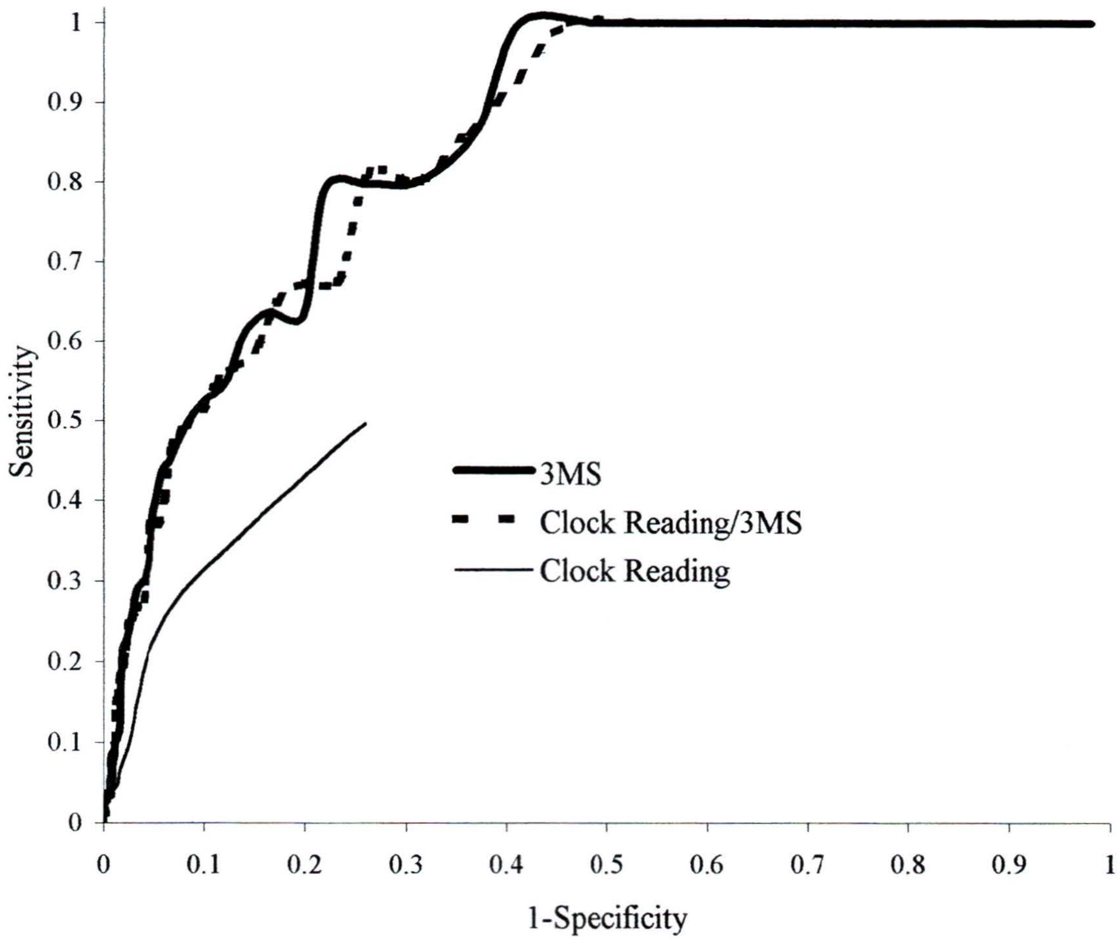


Figure 36. ROCs of the 3MS, Clock Reading, and Combined Clock Reading/3MS tasks for 65-79 year olds.

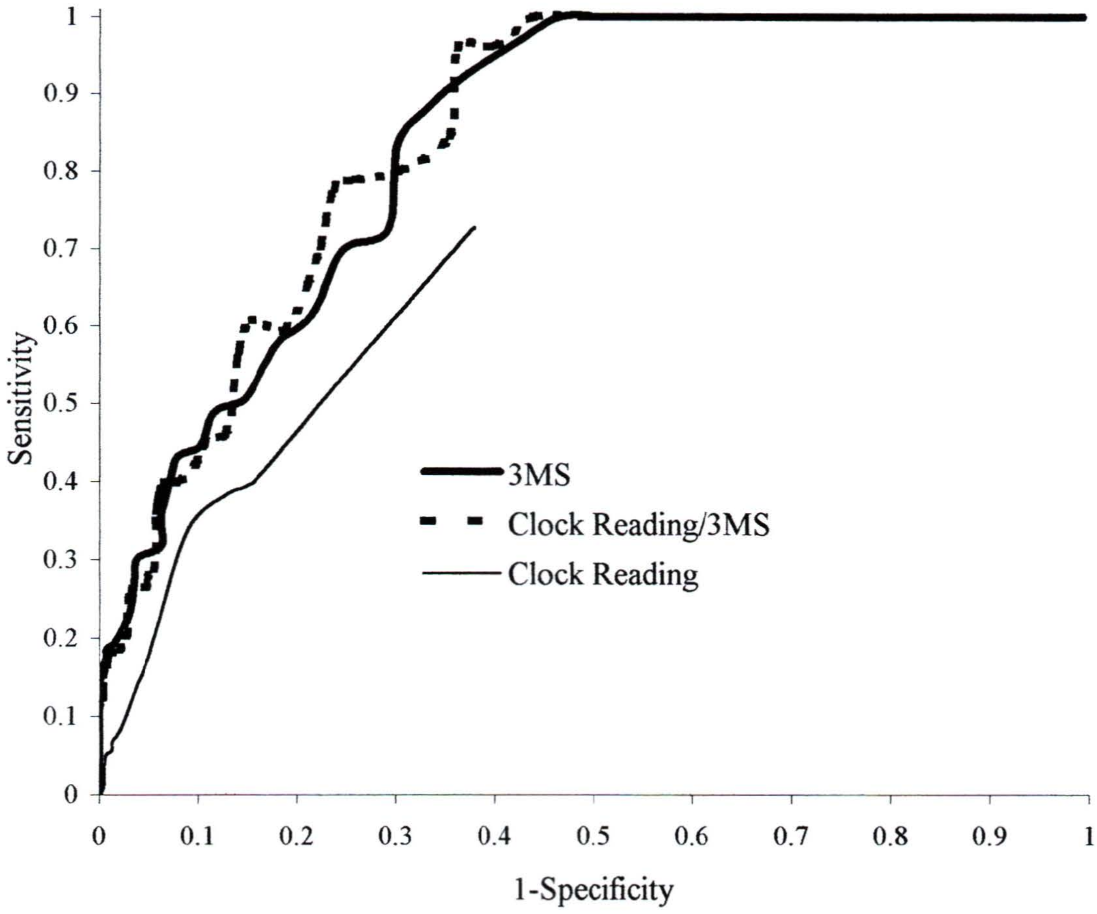


Figure 37. ROCs of the 3MS, Clock Reading, and Combined Clock Reading/3MS tasks for 80+ year olds.

Combined Versus Single Screening Tools for Education Groups

I wanted to know if components of the Clock Test added to the accuracy of the 3MS for persons with different levels of education. However, the results were similar for both education groups: the combined Clock Drawing – Reversed/3MS was equally accurate as the 3MS task alone but more accurate than the Clock Drawing task alone. Similarly, the combined Clock Setting/3MS was equally accurate to the 3MS alone but more accurate than the Clock Setting task alone for both education groups. Finally, the combined Clock Reading/3MS was equally accurate to the 3MS alone but more accurate than the Clock Reading task alone. The AUCs, SEs, *z*-values, and percent difference (if applicable) of the combined versus single screening tools are shown in Table 11. Figure 38 shows the ROC plots for the group with eight or fewer years of education and Figure 39 shows the ROC plots for the groups with nine or more years of education for the combined Clock Drawing – Reversed/3MS versus the Clock Drawing – Reversed and 3MS tasks. Figure 40 shows the ROC plots for the group with fewer years of education and Figure 41 shows the ROC plots for the group with more years of education of the 3MS, the Clock Setting task, and the combined Clock Setting/3MS tasks. Figure 42 shows the ROC plots for the group with few years of education and Figure 43 shows the ROC plots for the group with more years of education for the 3MS, the Clock Reading task, and the combined Clock Reading/3MS tasks.

Combined Versus Single Screening Tools for Sex Groups

I was interested in whether components of the Clock Test added to the accuracy of the 3MS for males or females. However, the results were similar for males and females: the combined Clock Drawing – Reversed/3MS was equally accurate as the 3MS

Table 11. The AUCs, Standard Errors (SEs), *z*-values, and, Difference in Percent (Diff %) of the Screening Tools for Each Education Group.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
0 – 8 Years Education				
3MS	0.8524	0.0412		
			0.01	N/A
Clock Drawing – Reversed/3MS	0.8416	0.0342		
			5.87*	35.5%
Clock Drawing - Reversed	0.4864	0.0477		
9+ Years Education				
3MS	0.8469	0.0297		
			-0.00	N/A
Clock Drawing – Reversed/3MS	0.8508	0.0320		
			9.48*	41.0%
Clock Drawing - Reversed	0.4413	0.0291		
0 – 8 Years Education				
3MS	0.8524	0.0412		
			0.00	N/A
Clock Setting/3MS	0.8485	0.0335		
			5.48*	16.2%
Clock Setting	0.6865	0.0424		
9+ Years Education				
3MS	0.8469	0.0297		
			-0.86	N/A
Clock Setting/3MS	0.8600	0.0300		
			7.25*	36.0%
Clock Setting	0.5001	0.0396		

Table continues. * $p < 0.001$ two-tailed.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
0 – 8 Years Education				
3MS	0.8524	0.0412		
			0.75	N/A
Clock Reading/3MS	0.8364	0.0424		
			14.07*	67.8%
Clock Reading	0.1586	0.0229		
9+ Years Education				
3MS	0.8469	0.0297		
			-0.01	N/A
Clock Reading/3MS	0.8578	0.0306		
			23.92*	77.2%
Clock Reading	0.0858	0.0101		

* $p < 0.001$ two-tailed.

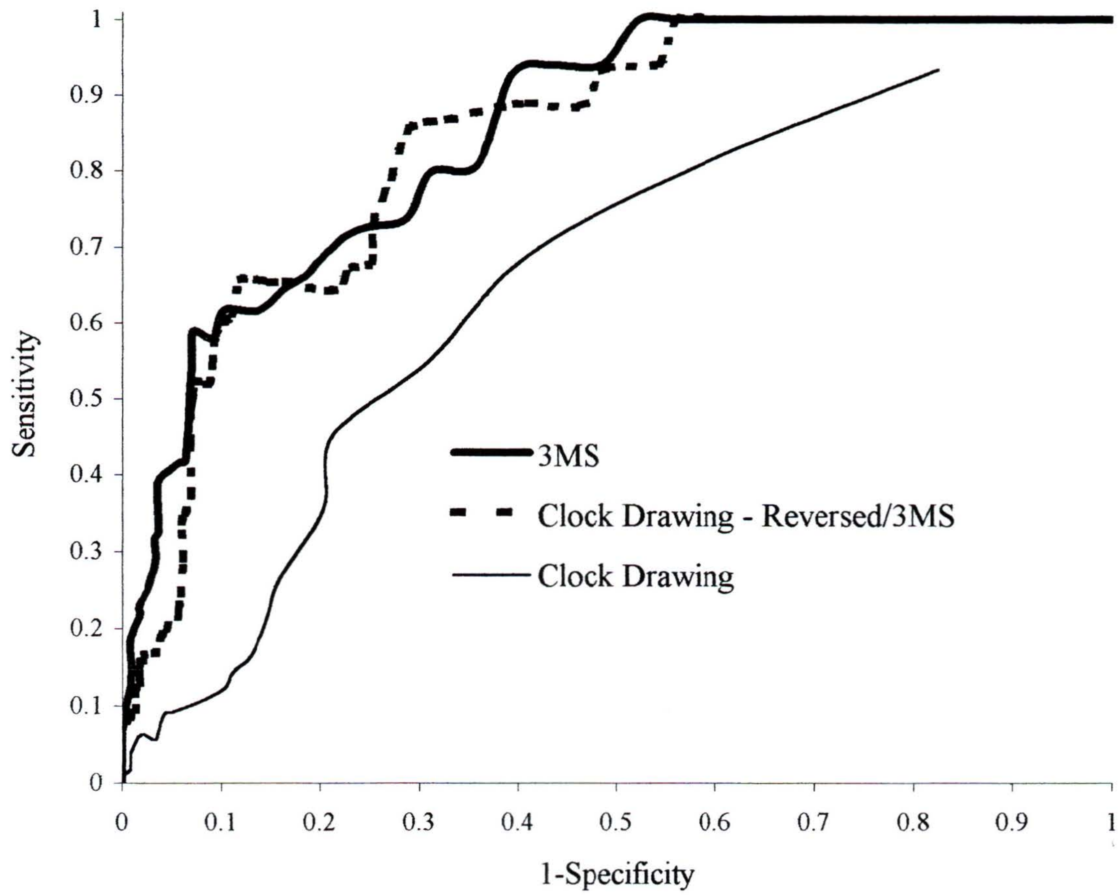


Figure 38. ROCs of the 3MS, Clock Drawing – Reversed, and combined Clock Drawing – Reversed/3MS tasks for individuals with 0-8 years of education.

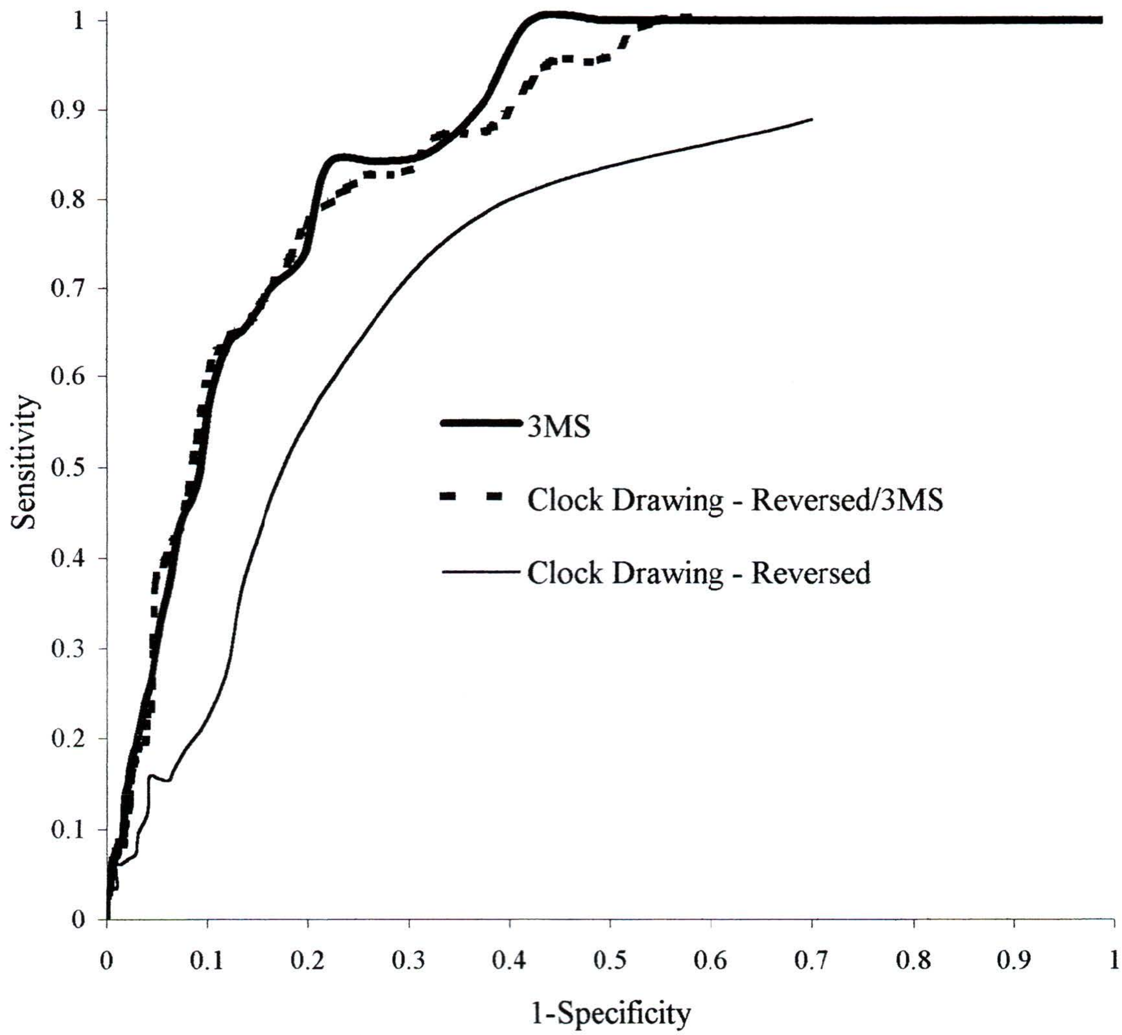


Figure 39. ROCs of the 3MS, Clock Drawing – Reversed, and combined Clock Drawing – Reversed/3MS tasks for individuals with 9+ years of education.

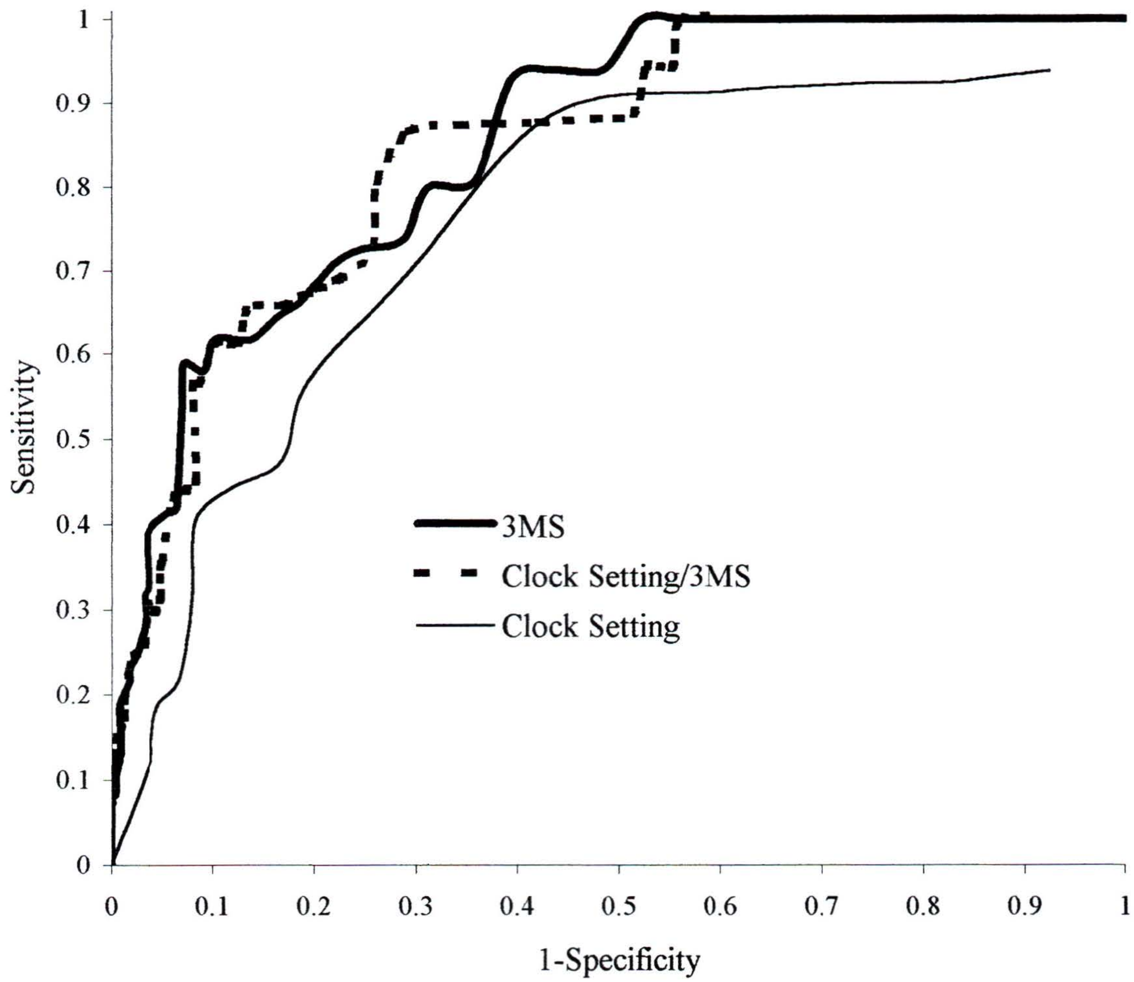


Figure 40. ROCs of the 3MS, Clock Setting, and combined Clock Setting/3MS tasks for individuals with 0-8 years of education.

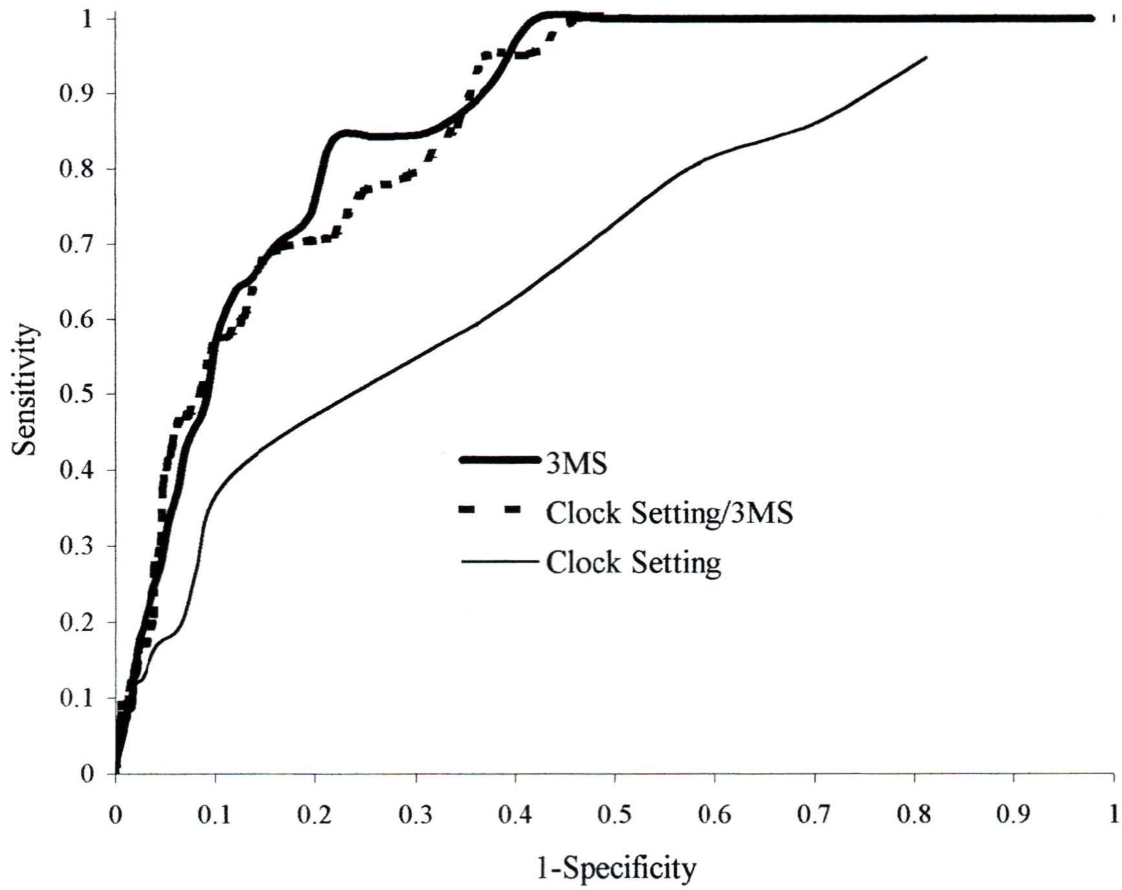


Figure 41. ROCs of the 3MS, Clock Setting, and combined Clock Setting/3MS tasks for iIndividuals with 9+ years of education.

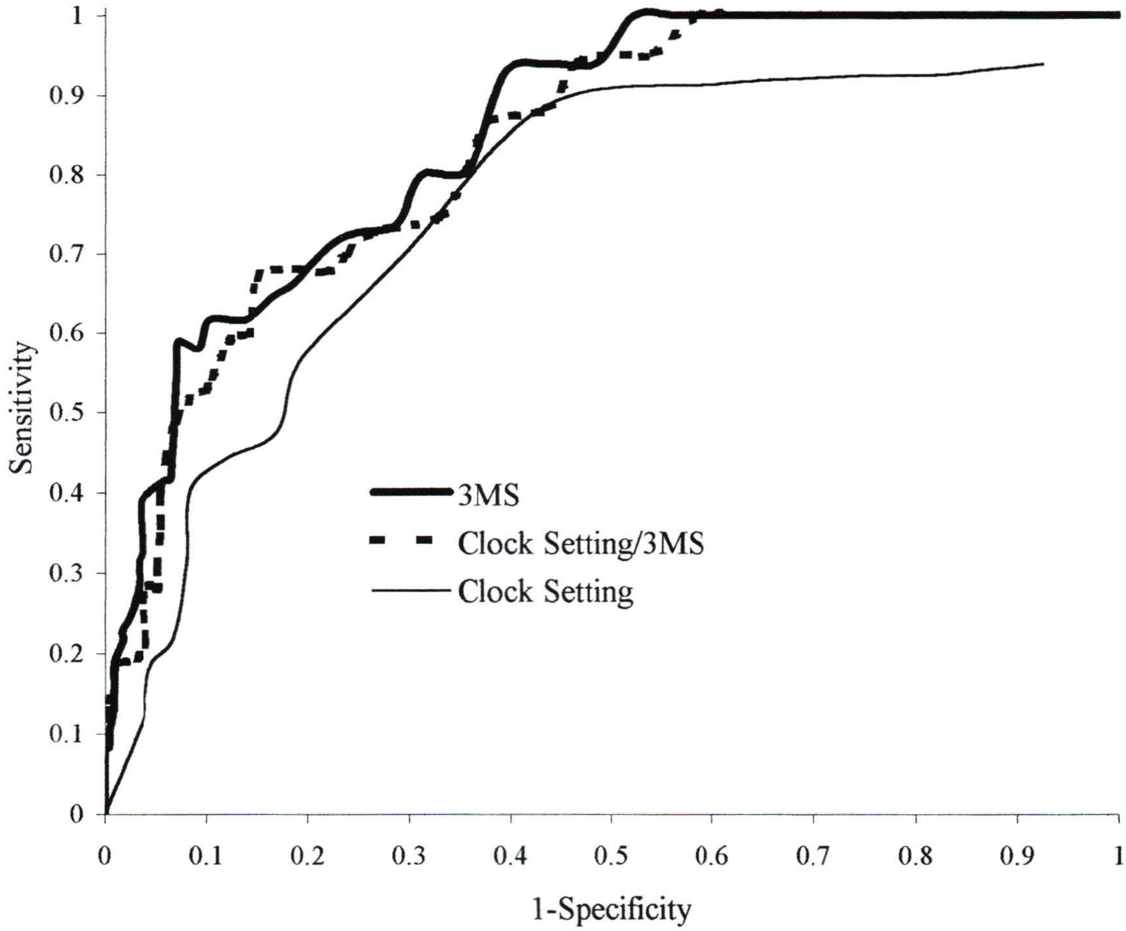


Figure 42. ROCs of the 3MS, Clock Reading, and combined Clock Reading/3MS tTasks for individuals with 0-8 years of education.

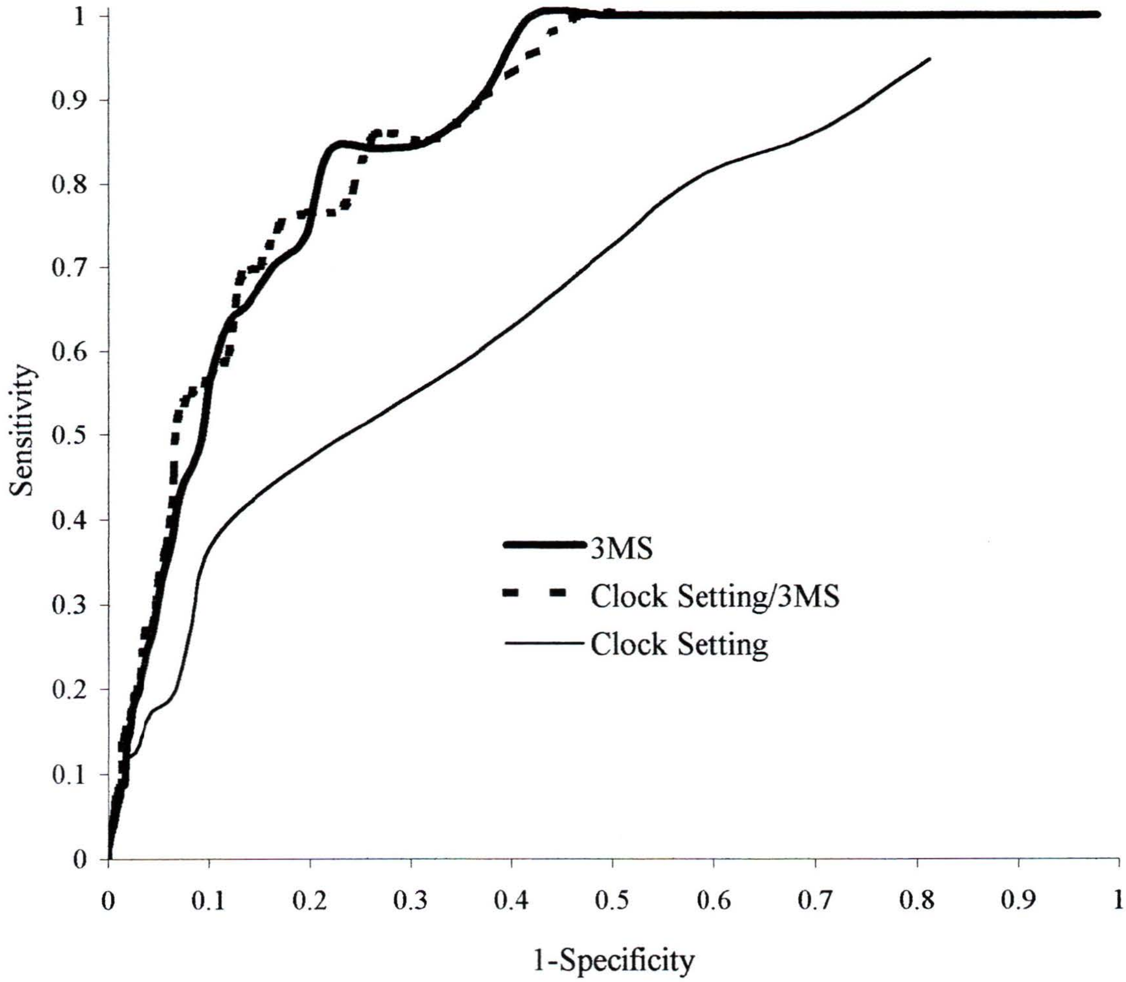


Figure 43. ROCs of the 3MS, Clock Reading, and combined Clock Reading/3MS tasks for individuals with 9+ years of education.

task alone but more accurate than the Clock Drawing task alone. Similarly, the combined Clock Setting/3MS was equally accurate to the 3MS alone but more accurate than the Clock Setting task alone for both education groups. Finally, the combined Clock Reading/3MS was more accurate than the Clock Reading task alone but equally accurate to the 3MS alone. Table 12 shows the AUCs, *SE*s, *z*-values, and, if applicable, percent difference between the combined and single screening tools for males and females. Figure 44 shows the ROC plots for males and Figure 45 shows the ROC plots for females of the 3MS, the Clock Drawing – Reversed task, and the combined Clock Drawing – Reversed/3MS tasks. Figure 46 shows the ROC plots for males and Figure 47 shows the ROC plot for females of the 3MS, the Clock Setting task, and the combined Clock Setting/3MS tasks. Figure 48 shows the ROC plots for males and Figure 49 shows the ROC plots for females of the 3MS, the Clock Reading task, and the combined Clock Reading/3MS tasks.

Table 12. The AUCs, Standard Errors (*SE*s), *z*-values, and Differences in Percent (Diff %) of the Screening Tools for Each Sex Group.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
Males				
3MS	0.8107	0.0374		
			-0.00	N/A
Clock Drawing – Reversed/3MS	0.8109	0.0407		
			7.97*	41.8%
Clock Drawing - Reversed	0.3025	0.0332		
Females				
3MS	0.8885	0.0236		
			-0.01	N/A
Clock Drawing – Reversed/3MS	0.8946	0.0262		
			8.79*	39.0%
Clock Drawing - Reversed	0.5045	0.0359		
Males				
3MS	0.8107	0.0374		
			-0.87	N/A
Clock Setting/3MS	0.8273	0.0371		
			5.34*	28.2%
Clock Setting	0.5449	0.0377		
Females				
3MS	0.8885	0.0236		
			-0.00	N/A
Clock Setting/3MS	0.8929	0.0307		
			5.75*	32.9%
Clock Setting	0.5642	0.0482		

Table continues. * $p < 0.001$ two-tailed.

Screening Tool	AUC	*SE*	*z*-Value	Diff %
Males				
3MS	0.8107	0.0374		
			-0.29	N/A
Clock Reading/3MS	0.8163	0.0379		
			18.49*	73.2%
Clock Reading	0.0842	0.0115		
Females				
3MS	0.8885	0.0236		
			-0.32	N/A
Clock Reading/3MS	0.8932	0.0247		
			27.91*	78.5%
Clock Reading	0.1083	0.0134		

* $p < 0.001$ two-tailed.

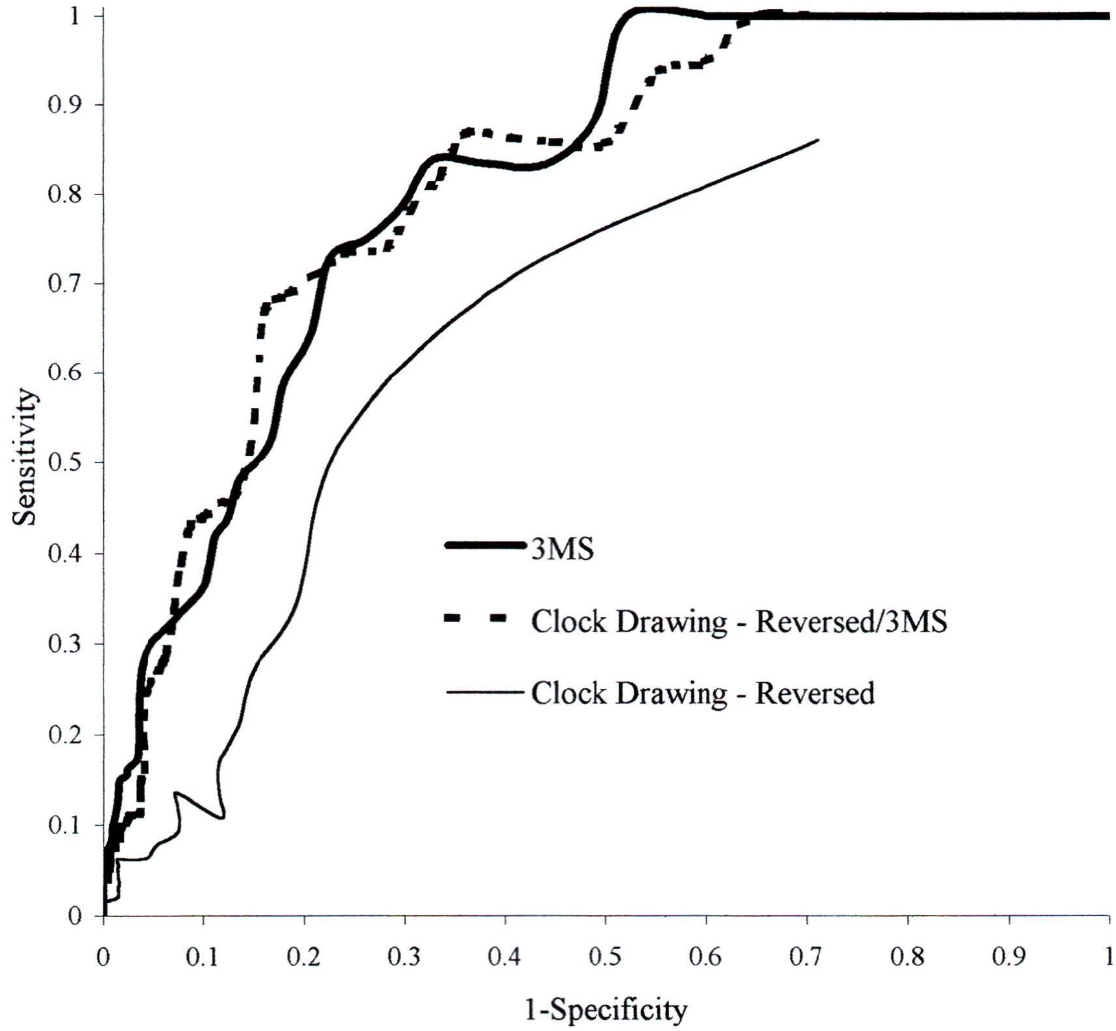


Figure 44. ROCs of the 3MS, Clock Drawing – Reversed, and combined Clock Drawing – Reversed/3MS tasks for males.

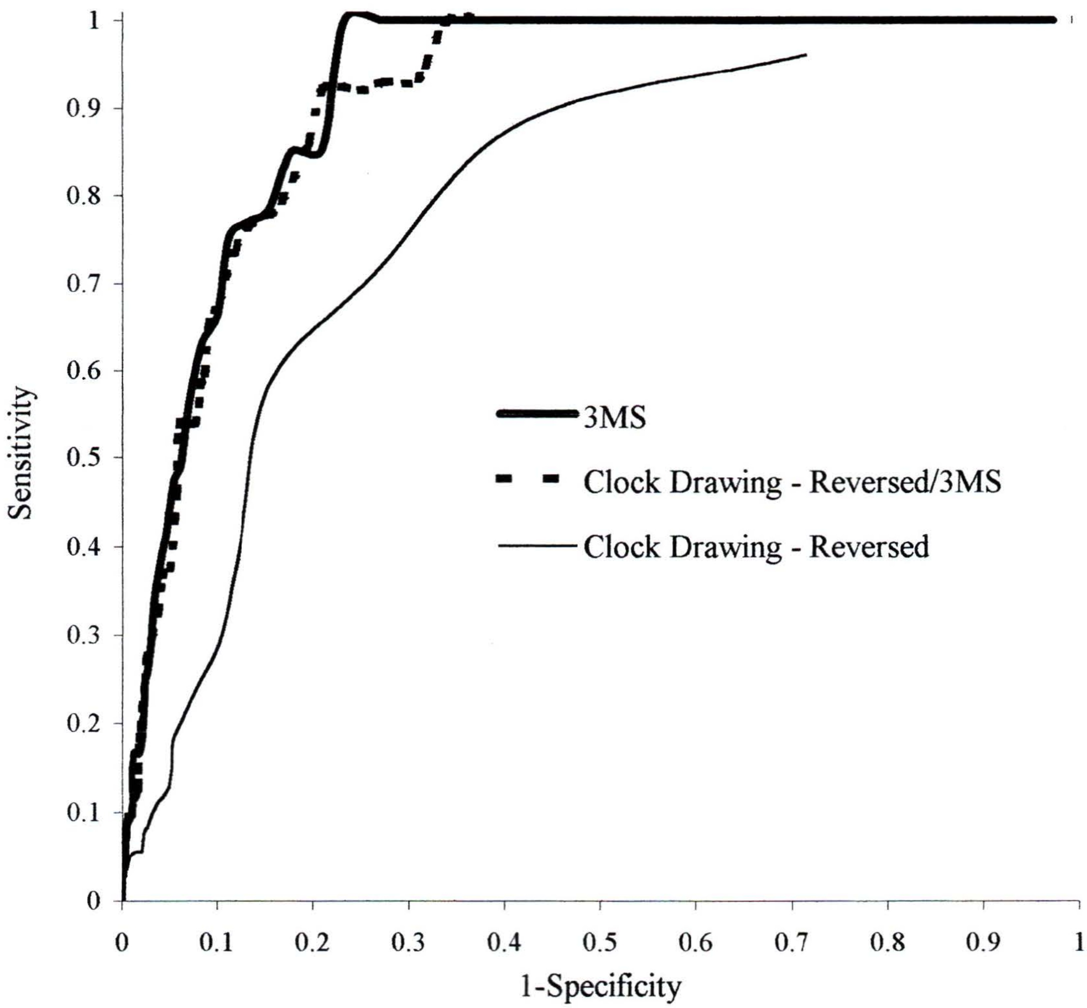


Figure 45. ROCs of the 3MS, Clock Drawing – Reversed, and combined Clock Drawing – Reversed/3MS tasks for females.

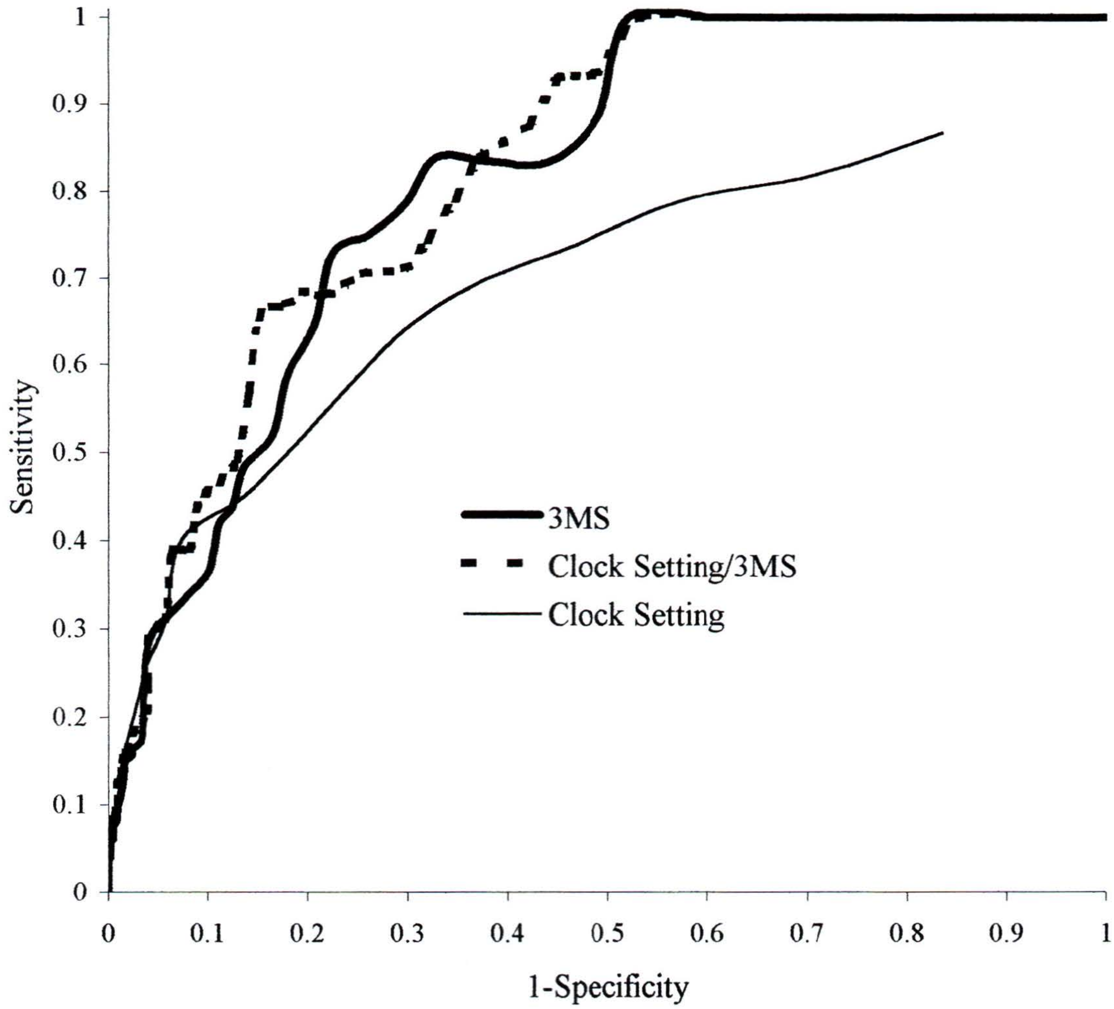


Figure 46. ROCs of the 3MS, Clock Setting, and combined Clock Setting/3MS tasks for males.

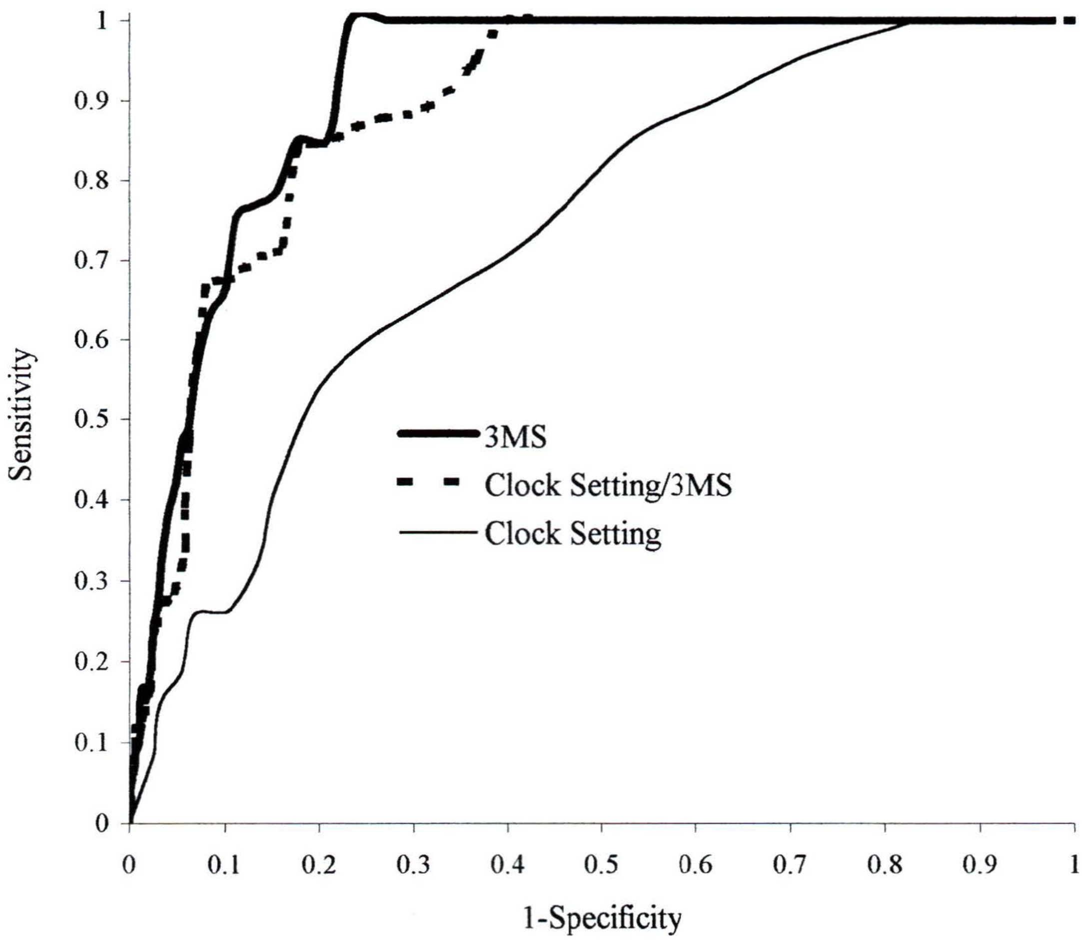


Figure 47. ROCs of the 3MS, Clock Setting, and combined Clock Setting/3MS tasks for females.

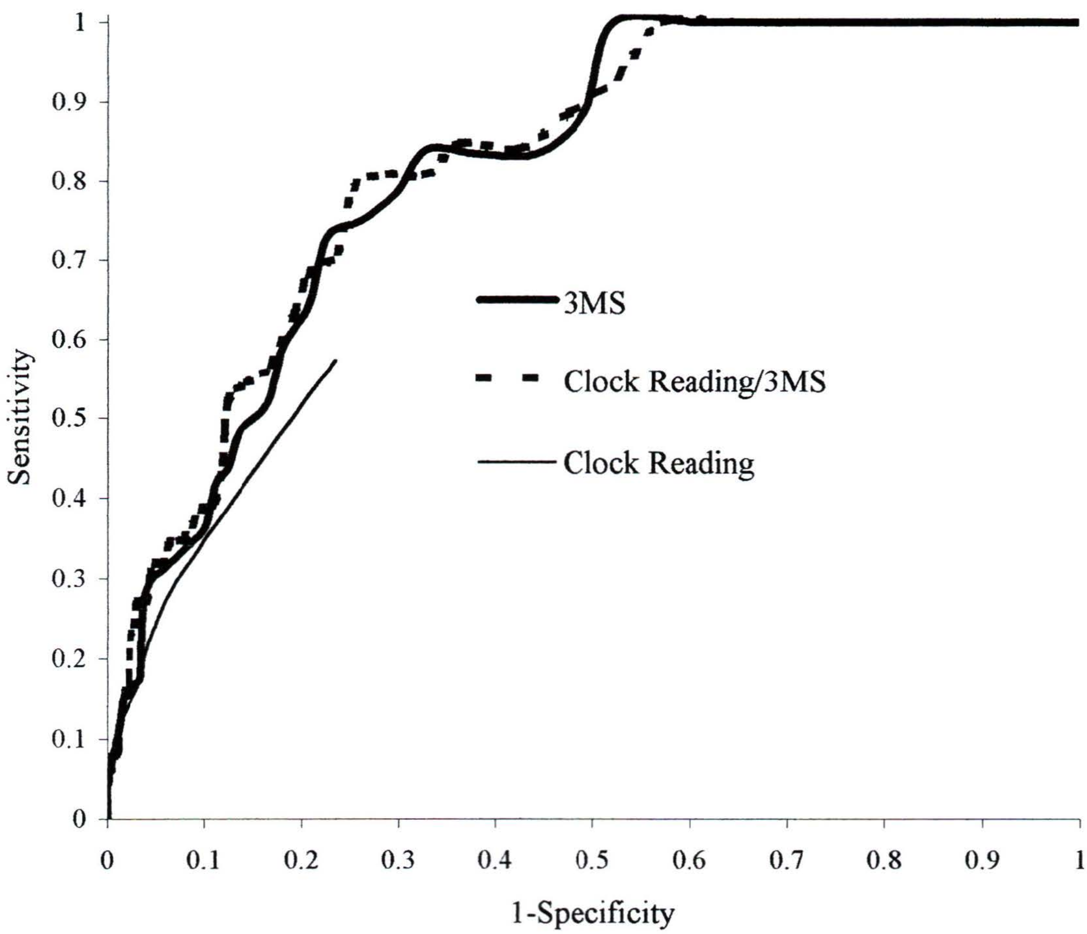


Figure 48. ROCs of the 3MS, Clock Reading, and combined Clock Reading/3MS tasks for males.

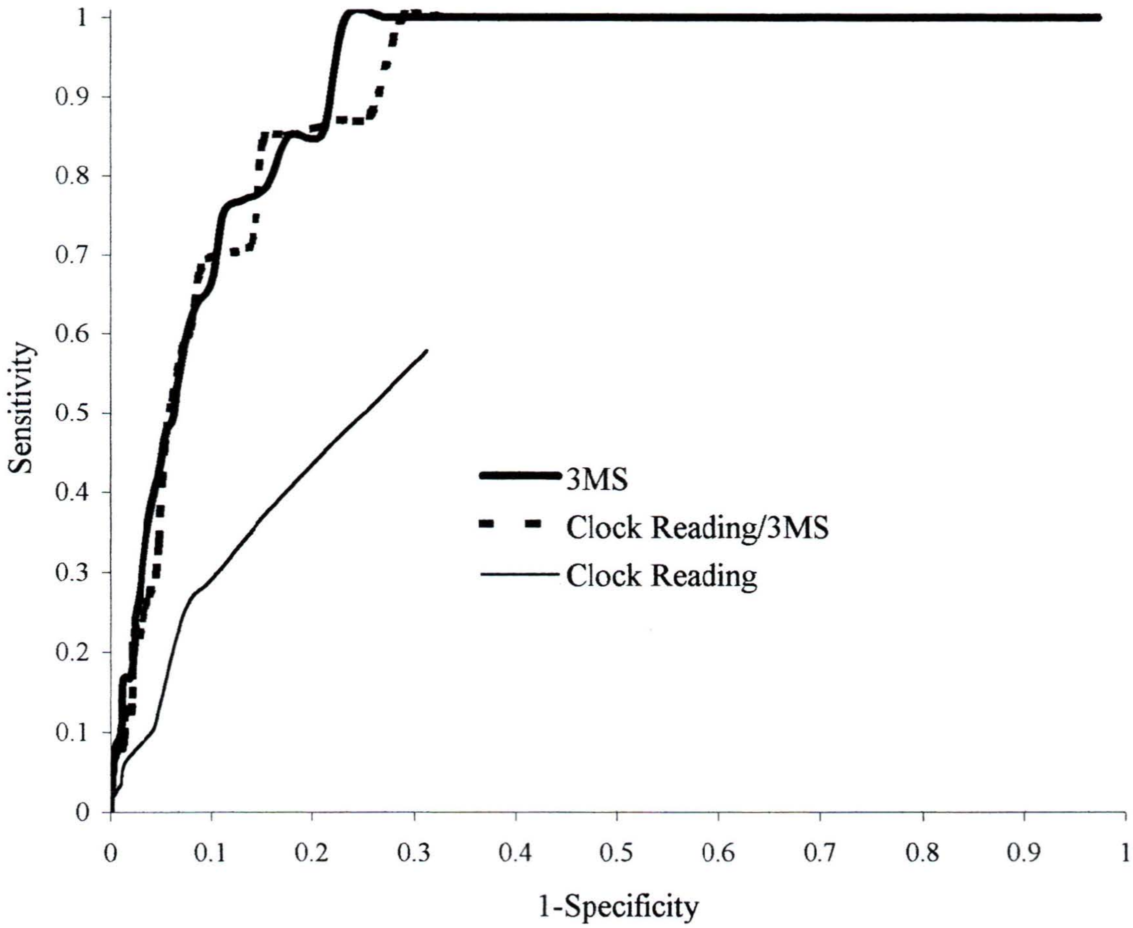


Figure 49. ROCs of the 3MS, Clock Reading, and combined Clock Reading/3MS tasks for females.

CHAPTER 5

Discussion

The aim of this study was to find the most accurate screening tool for cognitive impairment. To this end, I attempted to improve the accuracy of the 3MS by adjusting it for the influence of age, education, and sex. In addition, I combined the 3MS with other screening tools for cognitive impairment to see if a combined test increased the accuracy (i.e., increased the sensitivity and/or specificity) of screening for cognitive impairment. The effects of demographic influences on the accuracy of screening for cognitive impairment were also investigated.

Review of Results

3MS Versus Adjusted-3MS

Consistent with previous findings, the 3MS was significantly influenced by age, education (Bravo & Hebert, 1997a; Tombaugh et al., 1996), and sex (Bravo & Hebert, 1997a). In this study, the 3MS was adjusted for the influence of these demographic influences in an attempt to make a more accurate mental status screening tool for cognitive impairment. When the original 3MS and the adjusted-3MS scores were compared, the AUC analysis indicated that the original 3MS was marginally more accurate than the adjusted-3MS in identifying participants with and without cognitive impairment. This finding was contrary to my hypothesis, because, based on previous research, I expected that validity (i.e., as measured by the sensitivity and specificity) would be compromised when factors other than cognitive impairment influenced the score an individual obtains on the screening tool (Retzlaff & Gibertini, 1994). However, the use of regression adjusted scores for the 3MS was less accurate than the original 3MS scores. Previous research found a similar result when using simultaneous multiple regression to adjust MMSE scores for the effects of age and education: the adjusted MMSE was less accurate than the original MMSE (Kraemer, Moritz, & Yesavage, 1998).

From these data, it appears that the portion of the 3MS score that predicts cognitive impairment is not predicted by age, education, and sex; consequently, adjustment for these demographic influences does not increase the predictive validity (i.e., accuracy) of the 3MS. Alternatively, the use of regression weights, which are based on the average influence of each demographic variable on the 3MS, are too gross of an adjustment for each individual's score. Therefore, the adjustment based on the average of the degree of influence marginally diminishes the predictive value of the 3MS for each individual. However, the averages are the best estimates available for the whole sample. The use of age and education corrected cut-off scores, a commonly used procedure in clinical practice, is also based on the average performance. The implications that these findings have on the use of age and education corrected cut-off scores for the 3MS are discussed later in the Future Research section of the discussion. Finally, it is possible that the adjusted-3MS was marginally less accurate than the original 3MS because the relation between demographic influences and validity in screening is contrary to the assumption that demographic influences reduce validity in screening, which is discussed further in the next section.

Effects of Age, Education, and Sex on Screening Tools

The relation between the influence of demographic variables on screening tools and the effect on the accuracy of these tools was investigated as a corollary of pursuing the main thesis, which was to attempt to create a more accurate screening tool for cognitive impairment. Previous research assumed that lower scores for a particular demographic group translated into a less accurate screening tool for that group, in that persons who obtained lower scores were more likely to be misclassified as cognitively impaired (Jagger et al., 1992). For instance, if education were to influence scores on the screening tool such that persons with eight or fewer years of education would obtain significantly lower scores than persons with nine or more years of education, the assumption is that persons with eight or fewer years of education would be more likely to appear spuriously impaired on the screening tool. This has been shown with the MMSE (e.g., Anthony, LeResche, Niaz, von Korff, & Folstein, 1982; Marcopulos, McLain, & Giuliano, 1997) and the 3MS (Tombaugh et al. 1996).

Consistent with the hypotheses of this study, the 3MS was greater for individuals with nine or more year of education; however, contrary to the hypotheses, the 3MS was equally accurate for both education groups. Consequently, participants with fewer years of education were not more likely to be misclassified by the 3MS as cognitively impaired than were participants with more education. I also hypothesized that the 3MS would be greater, and more accurate, for younger than older participants, and these data suggest that, although the average 3MS raw score was greater for younger than for older participants, the 3MS was equally accurate for each age group. In this study, females obtained greater scores on the 3MS, which was consistent with my hypothesis and previous research (Bravo & Hebert, 1997a). Additionally, the results suggested that the 3MS was marginally more accurate for females than for males, which was also consistent with my hypothesis that the 3MS would be more accurate for females than for males because males were more likely to obtain lower scores on the 3MS. These data suggest that lower scores by a older participants or participants with few years of education did not translate into diminished accuracy when screening for cognitive impairment with the 3MS for these demographic groups, which raises questions about assumptions that particular demographic groups are more likely to be misclassified as impaired because they tend to obtain lower scores on the screening tool for cognitive impairment.

In fact, examples from components of the Clock Test suggested that lower scores by a particular demographic group translated into increased accuracy when identifying cognitive impairment for that demographic group. All scores on components of the Clock Test were significantly higher for younger participants (i.e., 65-79 years old) than for older participants (i.e., 80+ years old), which was consistent with the hypotheses of this study and previous research (Tuokko et al., 1995). Specifically, Clock Setting scores were higher for younger than for older persons, but the Clock Setting task was significantly more accurate for the older age group than for the younger age group. Similarly, the Clock Reading scores were significantly higher for younger than for older persons; however, the Clock Reading task was significantly more accurate for older persons than for younger persons.

Similar results were seen with education groups and components of the Clock Test. Contrary to the hypotheses and previous research (Tuokko et al., 1995), all components of the Clock Task were affected by education level. For example, average scores on the Clock Setting task were significantly higher for participants with nine or more years of education than for participants with eight or fewer years of education. However, when the accuracy of the Clock Setting task was examined, Clock Setting was more accurate for participants with eight or fewer years of education than for participants with nine or more years of education. Similarly, the average Clock Reading score was higher for participants with nine or more years of education than for individuals with eight or fewer years of education, but the Clock Reading task was more accurate for participants with eight or fewer years of education than for participants with nine or more years of education.

These results raise questions not only about the assumption that persons of a particular demographic group who score lower on a screening tool are more likely to be mislabelled as cognitively impaired, but also about the use of demographically corrected cut-off scores for screening tools. The rationale behind the use of demographically corrected cut-off scores is based on the above-mentioned assumption that persons who tend to obtain lower scores on the screening tool are more likely to appear impaired when compared to persons who tend to obtain higher scores on the screening tool. These data indicated that screening tools for cognitive impairment could, in fact, be more accurate for individuals who tend to obtain lower scores on the screening tool. Therefore, the use of demographically corrected cut-off scores could, potentially, reduce the accuracy in some instances. For example, the use of age corrected cut-off scores for the Clock Setting task, which was more accurate for older than for younger participants might diminish its superior accuracy at identifying cognitive impairment in older persons. Future research could help elucidate the relation between lower scores on a screening tool by particular demographic groups and the screening tool's accuracy (i.e., validity) for that particular demographic group. Future research could also address if the use of demographically corrected cut-off scores either reduce validity or have no influence on the validity of screening for cognitive impairment.

Combining Components of the Clock Test with the 3MS

Components of the Clock Test (i.e., Clock Drawing – Reversed, Clock Setting, and Clock Reading) were added to the 3MS in an attempt to increase the discriminability of persons with and without cognitive impairment (i.e., increase the accuracy). Forward stepwise logistic regression suggested that Clock Drawing – Reversed and Clock Setting predicted disease status above the prediction provided by the 3MS. Consequently, these three tasks were added together (i.e., Clock Drawing – Reversed with Clock Setting and the 3MS) and compared to each screening tool alone. Although the combined Clock Drawing – Reversed/Clock Setting/3MS was more accurate when compared to the Clock Drawing – Reversed and Clock Setting tasks alone, it was less accurate when compared to the 3MS alone. Finally, a combined total Clock Test/3MS score was computed by summing the three components of the Clock Test and the 3MS together. The combined total Clock Test/3MS was equal in accuracy when compared to the 3MS alone, but the combined score was more accurate than each of the components of the Clock Test alone.

Previous research with simultaneous multiple regression (Tuokko, 1993 as cited in Tuokko et al., 1995) suggested that only the Clock Setting task predicted cognitive impairment over and above the 3MS. In this study, forward stepwise logistic regression suggested that Clock Setting and Clock Drawing – Reversed were found to predict cognitive impairment over and above the 3MS. However, when compared using AUC analysis, the addition of Clock Setting and Clock Drawing – Reversed to the 3MS did not improve prediction of participants with and without cognitive impairment. The disparate findings from the regression analyses is likely due to the fact that the regressions were based on different samples. Tuokko (1993 as cited in Tuokko et al., 1995) investigated normal elderly and those diagnosed with dementia, whereas this study investigated cognitively normal elderly participants and participants with cognitive impairment, which was either dementia or cognitive impairment without dementia. It is plausible, therefore, that in a sample of cognitively normal versus demented participants, only the Clock Setting task would have added uniquely to the 3MS. Further, the types of regression differed (i.e., simultaneous multiple regression versus forward stepwise logistic regression). Finally, in this study a constant was not used in the regression equation because I was interested only in whether components of the Clock Test added to the 3MS in terms of identifying participants with and without cognitive impairment. I did not use the constant (which would have been based on the chance levels or average of the number of participants who were normal and cognitively impaired) because I was not interested in whether or not components of the Clock Test added to the 3MS and to chance levels of predicting cognitive impairment.

In addition, this study found that, although forward stepwise multiple regression suggested that the Clock Drawing – Reversed and Clock Setting tasks added unique prediction of participants with and without cognitive impairment above the 3MS, the AUC analysis did not find any improved prediction from adding the Clock Drawing – Reversed and Clock Setting tasks to the 3MS. The AUC and logistic regression both indicate the correct classification rate of a screening tool and AUC analysis and stepwise logistic regression both indicate if the correct classification rate for two screening tools is significantly different. However, this study found that the results of the forward stepwise logistic regression differed from the results of the AUC analysis. These differing results could have occurred for a variety of reasons, the fact that the logistic regression model did not include a constant. In addition, the difference in the results from the logistic regression and the AUC analysis likely occurred, at least in part, because the two analyses were computed on different samples. The logistic regression was, by necessity, only computed on the sub-sample with a verified diagnosis, and the AUC analysis was based on the whole sample. Further, the AUCs used in the AUC analysis had the verification bias corrected for, which increased the specificity and decreased the sensitivity. It is possible that the changes in sensitivity and specificity that resulted from the correction for the verification bias eliminated the predictive advantage, which appeared in the logistic regression analysis, of the components of the Clock Test above the 3MS. Finally, logistic regression did not combine the screening tools in an unweighted manner, as was done in the AUC analysis. The use of the logistic regression weights may have altered the contribution of the components of the Clock Test to the 3MS in such a way that led to a more accurate composite score.

Combining Single Components of the Clock Test with the 3MS

I combined single components of the Clock Test with the 3MS to see if these combined tests had increased sensitivity and/or specificity when identifying cognitive impairment. I found that when the components of the Clock Test were combined with the 3MS, the combined score was equally accurate at identifying cognitive impairment when compared to the 3MS alone, but the combined score was more accurate than each component of the Clock Test alone. The finding that the combined Clock Setting/3MS was equally accurate to the 3MS was contrary to my hypothesis. I had hypothesized that the combined Clock Drawing – Reversed/3MS and combined Clock Reading/3MS would not be more accurate than the 3MS alone, which was consistent with the results.

I also hypothesized that the components of the Clock Test combined with the 3MS would be predominantly helpful at identifying cognitive impairment in participants of particular demographic groups. However, the results did not change when the combined screening tools were compared to the 3MS for different age, education, and sex groups. The combined screening tools were not more accurate than the 3MS alone for any demographic group.

I hypothesized that the components of the Clock test when combined with the 3MS would be particularly helpful at identifying cognitive impairment in persons who were poorly educated. These hypotheses were partially based on previous findings with a study of a clinic and population based sample, which showed that the Clock Test was not affected by education level (Tuokko et al., 1995). In addition, although recent research suggested that the Clock Test may be affected by education (Corney, 2000), other research suggested that a clock drawing task was not as affected by education as was the MMSE, which is a screening tool on which the 3MS is based (Borson et al., 1999). Consequently, I hypothesized that the components of the Clock Test would not be as affected by education level, if affected at all, as is the 3MS. However, in this study the components of the Clock Test were affected by education level Although, the Clock Setting and Clock Reading tasks were more accurate for participants with eight or fewer years of education than for participants with nine or more years of education, this advantage for the poorly educated group apparently diminished when the screening tools were combined with the 3MS.

I also hypothesized that persons with eight or fewer years of education would perform more poorly on the 3MS but assumed that this would mean that the 3MS would be less accurate for persons with eight or fewer years of education than for persons with nine or more years of education. However, this assumption proved to be wrong. Although persons with eight or fewer years of education obtained lower average scores on the 3MS than persons with nine or more years of education, the 3MS was no less accurate for persons with eight or fewer years of education.

Limitations

This study attempted to provide useful information for clinical practice in screening for cognitive impairment by endeavoring to find the most accurate screening tool for cognitive impairment. However, the methods used in this study did not mirror those used in clinical practice; consequently, the results of this study are limited in their generalizability to clinical practice. In this study, the raw unweighted scores were combined and then these combined scores were analyzed to determine how well they were able to differentiate cognitively normal participants from cognitively impaired participants, whereas in clinical practice, the results of the two screening tools are typically assessed separately. This study did not assess the utility of the clinical information obtained from each screening measure and how this information corresponded with the final diagnosis. Future research could address the usefulness of adding components of the Clock Test to the 3MS when the 3MS score was apparently normal (at a particular cut-off point) but one or more of the components of the Clock Test were indicative of cognitive impairment. This would more closely resemble clinical practice and, therefore, would be generalizable for clinical practice when screening for cognitive impairment.

Another limitation of this study also impacts the ability to generalize these results in a way that may be clinically useful. The sample used in this study was population-based. Although this population based sample included persons who would have been seen in clinical practice, it differs greatly from the samples commonly seen by clinicians. Typically, clinicians see people who are having some sort of difficulties. This population-based sample included healthy older adults who were randomly selected (but restricted by self-selection bias). It is unclear how this population-based sample would differ from a typical clinical sample, and these results may vary greatly if the procedure were applied to a clinic-based sample of older adults.

Another limitation of this the fact that the AUC analysis as performed in this study, which does not appear to be wholly appropriate for all types of screening tools. AUC analysis involves comparing the AUCs of two screening tools to determine whether one screening tool is more accurate over all possible cut-off points. Because the upper left hand corner of the ROC space represents 100% specificity and 100% sensitivity (i.e., 0% for 1-specificity), it is assumed that the greater the AUC, the higher the plot is in the left-hand corner, and the more accurate the screening tool (Zweig & Campbell, 1993). This assumption appears to be problematic, at least for the AUC calculated with the trapezoidal method over all possible cut-off points (e.g., Bamber, 1975; Hanley & McNeil, 1982).

Potential problems appear to exist especially when the AUC analysis, as performed in this study at least, is applied to screening tools whose ROC plots do not converge on the coordinate of 1,1 (i.e., 1.0 for sensitivity and 1.0 for 1-specificity). This lack of convergence occurs when there is a ceiling effect for the screening tool;

consequently, the sensitivity and specificity do not vary between 0 and 100%. In these situations, the value of the AUC may be misleading. For example, the AUC analysis of the Clock Drawing – Reversed task for each education group indicated that the test was more accurate for persons with fewer years of education (i.e., AUC = 0.4864) than for persons with more years of education (AUC = 0.4413). Figure 19 (on page 86) clearly shows that the Clock Drawing – Reversed task for persons with nine or more years of education approaches the upper left hand corner of the plot, however, the screening tool with the highest area was not the screening tool that extended further toward the upper left hand corner of the plot. The AUC for Clock Drawing – Reversed was larger for the group with eight or fewer years of education because this plot extended further to the right (i.e., its range extended to lower levels of specificity and higher levels of sensitivity). Similar results occurred with Clock Drawing - Reversed for age groups (see Figure 12) and with Clock Reading for education groups (see Figure 23). In these cases the screening tool with the higher AUC did not represent the most accurate in terms of maximum sensitivity and specificity. In these three cases (i.e., Clock Drawing – Reversed for age groups, Clock Drawing – Reversed for education groups, and Clock Reading for education groups), the curve with the highest AUC was not the curve closest to the point of 100% sensitivity and 100% specificity. However, the curve with the highest AUC did have higher sensitivity, which may be clinically meaningful, and it could be argued that the difference in AUCs reflects this clinically useful information.

These results highlight a limitation of AUC analysis when used for screening tools that have a ceiling effect. Zweig and Campbell (1993) warn that, although the AUC is a useful way to communicate the accuracy of a screening tool, it can be misleading when interpreted in the absence of the ROC plot. Zweig and Campbell cite an example of two ROC plots that have the same AUC but are different in shape. Clearly the results from this study show that the AUC can be misleading if the ROC plot were not investigated, particularly if the screening tools have a ceiling effect as is shown in the example discussed of Clock Drawing – Reversed for the two education groups.

Future Research

Future research could address the limitations of this study. For example, the results of the addition of components of the Clock Test to the 3MS could be replicated with different samples. A clinic-based sample may differ from a population-based sample; consequently, I cannot generalize these results to a clinic sample. Future research could investigate whether the addition of components of the Clock Test add to the 3MS in terms of correctly identifying persons with and without cognitive impairment in a clinical sample. However, clinic-based samples differ depending on the referral base and referral sources; therefore, many types of clinic-based samples are required to address whether the results found here are generalizable to clinical practice.

In addition, future research could add the components of the Clock Test to the 3MS in a different manner than was done in this study to see if it resulted in a more accurate screening tool for cognitive impairment. Specifically, future research could address if the addition of components of the Clock Test to the 3MS is particularly useful when the 3MS score was normal at a particular cut-off point but when one or more of the components of the Clock Test were indicative of cognitive impairment. Further, Tuokko et al. (1995) found that the Clock Test was most accurate at identifying persons with and without dementia when using a criterion of two out of three the components indicate impairment, based on age-corrected cut-off scores. Future research could impose cut-off scores for the 3MS and the Clock Test to see if positive results on the 3MS and two components of the Clock Test are better able to identify persons with and without cognitive impairment than solely using the cut-off score for the 3MS only.

These results would have likely been different had I investigated only cognitively normal versus demented older adults. The variability added by including the cognitively impaired but not demented sample likely decreased the sensitivity and the specificity of the screening tools in this study. It is unclear whether the addition of components of the Clock Test to the 3MS would have resulted in different findings when trying to identify cognitively normal versus demented older adults. Future research could address this question. Further, future research could investigate whether the addition of components of the Clock Test added to the 3MS when screening for Alzheimer's disease. Research suggests that clock tasks might be particularly sensitive to Alzheimer's disease (Rouleau et al., 1992; Rouleau et al., 1996; Tuokko et al., 1992); therefore, the Clock Test may be particularly useful when added to the 3MS when screening for Alzheimer's disease.

Future research could investigate different scoring techniques for clock drawing to see whether the 3MS when combined with different versions of the clock drawing task result in greater identification of cognitively normal and cognitively impaired older adults. Based on research by Tuokko et al. (2000) which suggested the Tuokko et al., (1995) and Shulman et al., (1986) scoring techniques for clock drawing were most specific and sensitive than to other scoring techniques for clock drawing, future research could investigate whether different results are found when combining clock drawing scored using the Shulman et al. (1986) scoring technique combined with the 3MS.

Future research could also investigate if components of the Clock Test add to the MMSE when identifying persons with and without cognitive impairment because the MMSE the most widely used mental status screening tool for cognitive impairment.

Future research could address the implications of correcting the 3MS for age, education, and sex. The use of simultaneous multiple regression to adjust the 3MS for the effects of age, education, and sex resulted in significantly lower predictive validity than the original 'biased' 3MS. This may have implications for using age and education corrected cut-off scores for the 3MS. This empirical question could be investigated by comparing the sensitivity and specificity of percentile scores obtained from age and education cut-offs with percentile scores based on the overall average, providing the result obtained by using percentiles was equivalent to using raw scores.

Finally, future research might help elucidate the relation between lower scores for a particular demographic group and accuracy for that demographic group. The results from this study raise questions about the assumption that a person from a particular demographic group who scores lower on a screening tool is more likely to be misclassified as cognitively impaired. In addition, the results from this study raise questions about the use of demographically corrected cut-off scores because the use of these corrected cut-off scores are based on the above-mentioned assumption. Future research could address whether or not the use of demographically corrected cut-off scores are appropriate when screening for cognitive impairment.

Conclusions

In conclusion, adjusting the 3MS to correct for the effects of age, education, and sex did not result in a more accurate screening tool for cognitive impairment, possibly, because the portion of the 3MS score that predicts cognitive impairment is not affected by age, education, and sex. Consequently, adjustment for these demographic influences did not alter the predictive validity (i.e., sensitivity and or specificity) of the 3MS. This has implications for the use of age and education corrected cut-off scores for the 3MS.

Another important finding, which was a corollary to the main thesis, was that lower scores for a particular demographic group did not always translate into diminished accuracy for that demographic group. For instance, participants with eight or fewer years of education obtained lower average scores on the Clock Setting task than participants with nine or more years of education but the Clock Setting task was more accurate for participants with fewer years of education. Previous assumptions held that cognitively normal persons with eight or fewer years of education would be more likely to appear impaired when compared to cognitively normal persons with nine or more years of education. However, these data found that for a few measures, like Clock Setting for example, the measure was more accurate for persons with eight or fewer years of education than for persons with nine or more years of education.

Finally, the results of this study suggested that the addition of components of the Clock Test to the 3MS did not result in increased accuracy when identifying persons with and without cognitive impairment when compared to the use of the 3MS alone. Equivalent results were found when comparing across age, education, and sex groups: the components of the Clock Test when combined with the 3MS did not add predicative accuracy better than the 3MS alone. Based on the results of adding the unweighted components of the Clock Test to the unweighted 3MS, I recommend against administering the components of the Clock Test in combination with the 3MS, because the additional time and effort required to administer the components of the components of the Clock Test did not appear improve the predictive value. However, this recommendation is moderated by the fact that the methods used in this study do not mirror the methods used in clinical practice.

References

Abraham, I. L., Manning, C. A., Boyd, M. R., Neese, J. B., Newman, M. C., Plowfield, L. A., Reel, S. J. (1993). Cognitive screening of nursing home residents: Factor structure of the Modified Mini-Mental State (3MS) examination. *International Journal of Geriatric Psychiatry, 8,* 133-138.

Allen, M. S. & Yen, W. M. (1979). *Introduction to measurement theory.* Brooks/Cole Publishing Company: Monterey, CA.

Ainslie, N. K. & Murden, R. A. (1993). Effect of education on the clock-drawing dementia screen in non-demented elderly persons. *Journal of the American Geriatric Society, 41,* 249-252.

Anthony, J. C., LeResche, L., Niaz, U., von Korff, M. R., & Folstein, M. F. (1982). Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine, 12,* 297-408.

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

Baker, F. M. (1989). Screening tests for cognitive impairment. *Hospital and Community Psychiatry, 40,* 339-340.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology, 12,* 387-415.

Begg, C. B. & Greenes, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics, 39,* 207-215.

Bloch, D. A. (1997). Comparing two diagnostic tests against the same "gold standard" in the same sample. *Biometrics, 53,* 73-85.

Borson, S., Brush, M., Gil, E., Scanlan, J., Vitaliano, P., Chen, J., Cashman, J., Sta Maria, M. M., Barnhart, R., & Roques, J. (1999). The clock drawing test: Utility for dementia detection in multiethnic elders. *Journal of Gerontology, 54A,* 534-540.

Bravo, G. & Hebert, R. (1997a). Age- and education-specific reference values for the Mini-Mental and Modified Mini-Mental State examinations derived from a non-demented elderly population. *International Journal of Geriatrics Psychiatry, 12,* 1008-1018.

Bravo, G. & Hebert, R. (1997b). Reliability of the Modified Mini-Mental State examination in the context of a two-phase community prevalence study. *Neuroepidemiology, 16,* 141-148.

Cahn, D. A. & Kaplan, E. (1997). Clock drawing in the oldest old. *The Clinical Neuropsychologist, 11,* 96-100.

Cahn, D. A., Salmon, D. P., Monsch, A. U., Butters, N., Weiderholt, W. C., Corey-Bloom, J., & Barrett-Connor, E. (1996). Screening for dementia of the Alzheimer's type in the community: The utility of the clock drawing test. *Archives of Clinical Neuropsychology, 11,* 529-539.

Cahn-Weiner, D., Sullivan, E., Shear, P. K., Fama, R., Lim, K. O., Yesavage, J. A., Tinklenberg, J. R., & Pfefferbaum, A. (1999). Brain structural and cognitive correlates of clock drawing performance in Alzheimer's disease. *Journal of the International Neuropsychological Society, 5,* 502-509.

Canadian Study of Health and Aging Working Group. (1994). Canadian Study of Health and Aging: Study methods and prevalence of dementia. *Canadian Medical Association Journal, 6,* 433-440.

Corney, P. J. (2000). *The Clock Test: What does poor performance mean?* Unpublished honours thesis, University of Victoria, Victoria, British Columbia, Canada.

Critchley, M. (1966). *The parietal lobes.* Hafner Publishing Company: New York.

Death, J., Douglas, A., & Kenny, R. A (1993). Comparison of clock drawing with Mini Mental State Examination as a screening test in elderly acute hospital admissions. *Postgraduate Medical Journal, 69,* 696-700.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L., (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44,* 837-845.

Donald, A. (1996). Verification bias: A pitfall in evaluating screening tests. *Nursing Research, 45,* 350-352.

Dwyer, C. A. (1996). Cut scores and testing: Statistics, judgement, truth, and error. *Psychological Assessment, 8,* 360-362.

Esteban-Santillan, C., Praditsuwan, R., Ueda, H., & Geldmacher, D. S. (1998). Clock drawing in very mild Alzheimer's disease. *Journal of the American Geriatric Society, 46,* 1266-1269.

Essex-Sorlie, D. (1995). *Medical Biostatistics & Epidemiology.* Appleton & Lange: Norwalk, CN.

Evans, L. D. (1996). A two-score composite program for combining standard scores. *Behavior Research Methods, Instruments, and Computers, 28,* 209-213.

Ferrer, H. P. (1968). *Screening for Health: Theory and Practice.* Butterworth & Co.: Toronto.

Ferrucci, L., Cecchi, F., Guralnik, J. M., Giampaoli, S., Noce, C. L., Salani, B., Bandinelli, S., & Baroni, A. (1996). Does clock drawing predict cognitive decline in older persons independent of the Mini-Mental State Examination? *Journal of the American Geriatrics Society, 44,* 1326-1331.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State" a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research, 12,* 189-198.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. I., Delis, D. C. (1994). *Clock drawing: A neuropsychological analysis.* New York: Oxford University Press.

Ganguli, M. (1997). The use of screening instruments for the detection of dementia. *Neuroepidemiology, 16,* 270-280.

Grace, J., Nadler, J. D., White, D. A., Guilmette, T. J., Giuliano, A. J., Monsch, A. U., & Snow, M. G. (1995). Folstein vs Modified Mini-Mental State examination in geriatric stroke. *Archives of Neurology, 52,* 477-484.

Hadjistavropoulos, Tuokko, & Beattie. (1991). The Clock Test: Construct validity for a multidimensional assessment method [Abstract]. *Canadian Psychology, 32,* 219.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29-36.

Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839-843.

Jagger, C., Clarke, M., Anderson, J., & Battcock, T. (1992). Misclassification of dementia by the Mini-Mental State Examination: Are education and social class the only factors? *Age and Ageing, 21,* 404-411.

Juby, A. (1999). Correlation between the Folstein Mini-Mental State Examination and three methods of clock drawing scoring. *Journal of Geriatric Psychiatry and Neurology, 12,* 87-91.

Kozora, E. & Cullum, C. M. (1994). Qualitative features of clock drawing in normal aging and Alzheimer's disease. *Assessment, 1,* 179-187.

Kraemer, H. C., Moritz, D. J., & Yesavage, J. (1998). Adjusting Mini-Mental State Examination scores for age and education level to screen for dementia: Correcting bias or reducing validity. *International Psychogeriatrics, 10,* 43-51.

Kristjansson, B. (1990). *Administration and scoring manual for the Modified Mini-Mental State examination as used in the Canadian Study of Health and Aging.* University of Ottawa.

Kurzman, D. (1992). *The construct validity of the Clock Test in normal and demented adults.* Unpublished master's thesis, University of Victoria, Victoria, British Columbia, Canada.

Lam, L. C., Chiu, H. F., Ng, K. O., Chan, C., Chan, W. F., Li., S. W., & Wong, M. (1998). Clock-face drawing, reading and setting tests in the screening of dementia in Chinese elderly adults. *Journal of Gerontology, 53B,* 353-357.

Lamarre, C. J. & Patten, S. B. (1991). Evaluation of the Modified Mini-Mental State examination in a general psychiatric population. *Canadian Journal of Psychiatry, 36,* 507-511.

Lee, H., Seanwick, G. R., Coen, R. F., & Lawlor, B. A. (1996). Use of the clock drawing task in the diagnosis of mild and very mild Alzheimer's disease. *International Psychogeriatrics, 8,* 469-476.

Libon, D. J., Malamut, B. L., Swenson, R., Sands, L. P., & Cloud, B. S. (1996). Further analyses of clock drawings among demented and nondemented older subjects. *Archives of Clinical Neuropsychology, 11,* 193-205.

Libon, D. J., Swenson, R. A., Barnoski, E. J., & Sands, L. P. (1993). Clock drawing as an assessment tool for dementia. *Archives of Clinical Neuropsychology, 8,* 405-415.

Lin, S. C. (1999). Some results on combinations of two binary screening tests. *Journal of Biopharmaceutical Statistics, 9,* 81-88.

Malm, H. M. (1999). Medical screening and the value of early detection: When unwarranted faith leads to unethical recommendations. *Hastings Center Report, 29,* 26-37.

Manos, P. J. (1999). Ten-point clock test sensitivity for Alzheimer's disease in patients with MMSE scores greater than 23. *International Journal of Geriatric Psychiatry, 14,* 454-458.

Marcopulos, B. A., McLain, C. A., & Giuliano, A. J. (1997). Cognitive impairment or inadequate norms? A study of healthy, rural, older adults with limited education. *The Clinical Neuropsychologist, 11,* 111-131.

McDowell, I., Kristjansson, B., Hill, G. B., & Hebert, R. (1997). Community screening for dementia: The Mini Mental State Exam (MMSE) and Modified Mini-Mental State Exam (3MS) compared. *Journal of Clinical Epidemiology, 50,* 377-383.

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

Mitrushina, M. & Fuld, P. A. (1996). Cognitive Screening Methods. In I. Grand & K. M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders.* (pp. 118-138). New York: Oxford University Press.

O'Rourke, N., Tuokko, H., Hayden, S., & Beattie, B. L. (1997). Early identification of dementia: Predictive validity of the Clock Test. *Archives of Clinical Neuropsychology, 12,* 257-267.

Paolo, A. M. (1998). Psychometric issues in the clinical assessment of memory in aging and neurodegenerative disease. In A. I. Troester (Ed.), *Neurobiological and Cognitive and Clinical Perspectives on Memory Disorders, Patients with Dementia or Other Neurodegenerative Diseases.* (pp. 262-277). New York: Cambridge University Press.

Retzlaff, P. D. & Gibertini, M. (1994). Neuropsychometric Issues and Problems. In R. D. Vanderploeg (Ed.), *Clinicians Guide to Neuropsychological Assessment.* (pp. 185-209). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ritchie, K. (1988). The screening of cognitive impairment in the elderly: A critical review of current methods. *Journal of Clinical Epidemiology, 41,* 635-643.

Rouleau, I., Salmon, D. P., & Butters, N. (1996). Longitudinal analysis of clock drawing in Alzheimer's disease patients. *Brain and Cognition, 31,* 17-34.

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

Royall, D. R. & Polk, M. (1998). Dementias that present with and without posterior cortical features: An important clinical distinction. *Journal Of the American Geriatric Society, 46,* 95-105.

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

Statistics Canada (1999a). *Population by sex and age.* Retrieved July 1, 2000, from http://www.statcan.ca/english/Pgdb/People/Population/demo10a.htm.

Statistics Canada (1999b). *Population projections for 2001, 2006, 2011, 2021, and 2026.* Retrieved July 1, 2000, from http://www.statcan.ca/english/Pgdb/ People/ Population/ demo23.htm.

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 Geriatrics Society, 37,* 725-729.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285-1293.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-states diagnostics. *American Psychologist, 47,* 522-532.

Teng, E. L. & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry, 48,* 314-318.

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment, 8,* 48-59.

Tuokko, H. & Hadjistavropoulos, T. (1998). *An Assessment Guide to Geriatric Neuropsychology.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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's disease. *Journal of the American Geriatrics Society, 40,* 579-584.

Tuokko, H., Hadjistavropoulos, T., Miller, J. A., Horton, A., & Beattie, B. L. (1995). *The Clock Test.* Toronto: Multi-Health Systems.

Tuokko, H., Hadjistavropoulos, T., Rae, S., & O'Rourke, N. (2000). A comparison of alternative approaches to the scoring of clock drawing. *Archives of Clinical Neuropsychology, 15,* 137-148.

van Reekum, R., Simard, M., & Farcnik, K. (1999). Diagnosis of dementia and treatment of Alzheimer's disease: Pharmacologic management of disease progression and cognitive impairment. *Canadian Family Physician, 45,* 945-952.

Watson, Y. I., Arfken, C. L., & Birge, S. J. (1993). Clock completion: An objective screening test for dementia. *Journal of the American Geriatrics Society, 41,* 1235-1240.

Wilson, J. M. & Jungner, G. (1968). *Principles and Practice of Screening for Disease.* World Health Organization: Geneva.

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.

Zhou, X & Higgs, R. E. (2000). Assessing the relative accuracies of two screening tests in the presence of verification bias. *Statistics in Medicine, 19,* 1697-1705.

Zweig, M. H. & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39,* 561-577.

Appendix

The Modified Mini-Mental State (3MS)

3MS MMS

1. __/5 DATE AND PLACE OF BIRTH

"What is your birthdate?" (1 point for each)

Day ___, Month _______, Year ___

"Where were you born?" (1 point for each to maximum of 2 if the same as given for question 16)

Town _______, Province _______
Country _______ (if born outside of Canada)

2. __/3 __/3 REGISTRATION

"I am going to say 3 words for you to remember. Repeat them after I have said all 3 words: SHIRT...BROWN...HONESTY.

Say the words at 1 per second. Give 1 point for each correct answer on the first recall attempt. Repeat the words up to 3 more times until the client has learned all three.

SHIRT ___, BROWN ___, HONESTY ___

Number of presentations 1 2 3 4

3. __/7 __/5 MENTAL REVERSAL

"Please count from 1 to 5" Count forward Can __ Cannot __

If the client cannot coach once then ask them to count backward.

"Now count backward from 5 to 1" (Score for 3MS only)

Accurate	2
1 or 2 errors/misses	1
More than 2 errors/misses	0

"Now I want you to spell the word WORLD." Spell forward: Can __ Can't __
If the client cannot coach once then ask them to spell backward.

"Spell WORLD backward." Score the # of elements in the correct position.

Response _____________
Response in reverse _____________

Score 1 2 3 4 5

	3MS	MMS	
4.	__ / 9	__ / 3	FIRST RECALL

"What are the three words I asked you to remember?"

If client gives the word "world" as a response say "It was before world". Allow 3 seconds for a new response before proceding with cues.

For each word not recalled provide a category cue (e.g., "something to wear") and the multiple choice cue if category response is incorrect. If client can still not provide the correct answer give the correct answer.

	3MS	MMS
Spontaneous recall	3	1
Another word for "something to wear"	2	
Was it SHOES, SHIRT, or SOCKS?	0 or 1	
Spontaneous recall	3	1
Another word for "A colour"	2	
Was it BLUE, BLACK, or BROWN?	0 or 1	
Spontaneous recall	3	1
Another word for "A good personal quality"	2	
Was it HONESTY, CHARITY, or MODESTY?	0 or 1	

5.	__ / 15	__ / 5	TEMPORAL ORIENTATION

"What year is it?"

	3MS	MMS
Accurate	8	1
Missed by 1 year	4	
Missed by 2-5 years	1	
Missed by more than 5 years	0	

"What is the season?"

Accurate or within one month	0 or 1	1

"What month is this?"

Accurate or within 5 days	2	1
Missed by 1 month	1	
Missed by more than one month	0	

"What day of the month is this?"

Accurate	3	1
Missed by 1 - 2 days	2	
Missed by 3 - 5 days	1	
Missed by more than 5 days	0	

"What day of the week is this?"

Accurate	0 or 1	1

	3MS	MMS			
6.	__ 5	__ 5	SPATIAL ORIENTATION	3MS	MMS
			"What is the address here?"		
			Street number		1
			Street name		1
			City or town	0 or 1	1
			Province	0 or 1	1
			Country	0 or 2	1
			"Are we in a hospital (clinic), store, or home?"	0 or 1	

7. __ 5 __ 2

NAMING

Point to a part of your body and ask "What is the name of this?"

Score 0 if the client cannot name it within 2 seconds or gives an incorrect name. Do not repeat the question.

	3MS	MMS
Forehead	0 or 1	
Chin	0 or 1	
Shoulder	0 or 1	
Elbow	0 or 1	
Knuckle	0 or 1	
Show the client a pencil and ask "What is this?"		1
Point to a watch and ask "What is this?"		1

8. __ 10

FOUR-LEGGED ANIMALS

Ask "What animals have four legs?"

Allow 30 seconds for the response. If the subject gives no response in 10 seconds, repeat the question once. The first time an incorrect answer is provided, say, "I want four-legged animals." If the client has started to aswer and stopped before 10 animals, and 10 seconds have passed since the last name and 10 seconds remain ask, "What other animals have four legs." Do not correct subsequent errors. Score for the numer of correct responses, up to 10.

________	________	________
________	________	________
________	________	________
________	________	________
________	________	________

	3MS	MMS	
9.	__ 6		SIMILARITIES

Say "In what way are an arm and a leg alike?"
If the client fails to give an answer that is worth 2 points on the 1st part (arm & leg) assign the appropriate score and coach them by providing the 2 point answer. Do not coach for other items.

"Arm and Leg"	
Correct concept (i.e., limb, extremities)	2
Less correct answer (i.e., body parts, muscles)	1
Incorrect answer	0

"Laughing and Crying"	
Correct concept (i.e., feeling, emotion)	2
Less correct answer (i.e., use your mouth)	1
Incorrect answer	0

"Eating and Sleeping"	
Correct concept (i.e., necessary for life)	2
Less correct answer (i.e., body needs, relaxing)	1
Incorrect answer	0

10.	__ 5	__ 1	REPETITION

Say "Repeat what I say". "I would like to go home/out."
Use "out" in the sentence if the client if living at home.

	3MS	MMS
Accurate	2	
1 or 2 missed/wrong words	1	
> 2 missed/wrong words	0	

Say "Now repeat: no ifs, ands, or buts."

Entire phrase accurate	3	1
1 error	2	0
2 errors	1	0
3 errors	0	0

11.	__ 3	__ 1	READ AND OBEY (last page)

Hold up the paper on which the command is printed and say "Please do this."

If the client does not respond by closing his/her eyes in 5 seconds prompt him or her by pointing to the sentence and saying "Read this and do what it says". Allow 5 seconds for the response. As soon as the client closes his/her eyes say, "Please open them now."

	3MS	MMS
Obeys without prompting	3	1
Obeys after prompting	2	0
Reads aloud only	0 or 1	0

	3MS	MMS	
12.	__ / 5	__ / 1	WRITING (back of last page)

Provide the client with a blank sheet of paper, and a pencil with an eraser. Tell the client, **"I would like to have a sample of your handwriting." "Please write, I would like to go home/out."**

Repeat the sentence, but allow a maximum of 1 minute after the first reading for a response. The word "I" is excluded from the scoring.

	3MS	MMS
Entire phrase correct	5	1
1 word missed/wrong	4	0
2 words missed/wrong	3	0
3 words missed/wrong	2	0
4 or 5 words missed/wrong	0 or 1	0

	3MS	MMS	
13.	__ / 10	__ / 1	COPYING TWO PENTAGONS (last page)

Give the client the page with the pentagon drawings and say, **"Please copy this figure exactly as it is".** Allow 1 minute. MMS is scored as 1 if the whole figure is correct.

	3MS Each Pentagon	
5 approximately equal sides	4	4
5 unequal (≥ 2:1 sides)	3	3
Other enclosed figure	2	2
2 or more lines	1	1
1 line only	0	0
	Intersection	
4 corner enclosure	2	
Enclosure with < 4 corners	1	
Intersection lacking enclosure or no intersection	0	

	3MS	MMS	
14.	__ / 3	__ / 3	THREE-STAGE COMMAND

Hold a piece of paper in view of the client but out of his/her reach and say, **"Take this paper in your left hand"** (for left-handed persons say **"right hand"**), **"fold it in half, and hand it back to me."** After saying the whole command, hold the paper within reach of the client. Do not repeat any part of the command. Do not give visual cues (i.e., moving the paper toward a specific hand or reach for it before it is offered).

Take this paper with your left/right (nondominate) hand	0 or 1
Fold it in half	0 or 1
Hand it back to me	0 or 1

15. 3MS ___/9

SECOND RECALL

"What are the three words I asked you to remember earlier?"

If client gives the word "world" as a response say "It was before world". Allow 3 seconds for a new response before proceding with cues.

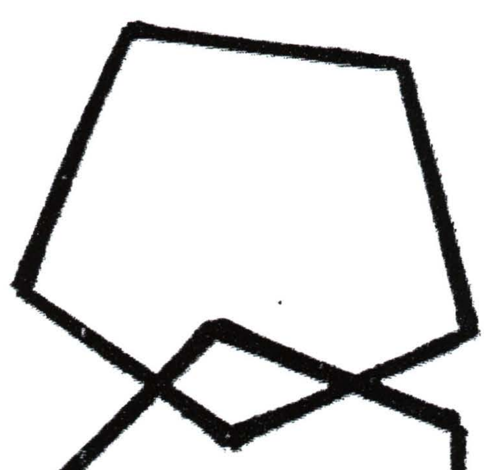
For each word not recalled provide a category cue (e.g., "something to wear") and the multiple choice cue if category response is incorrect.

Spontaneous recall	3
Another word for "something to wear"	2
Was it SHOES, SHIRT, or SOCKS?	0 or 1
Spontaneous recall	3
Another word for "A colour"	2
Was it BLUE, BLACK, or BROWN?	0 or 1
Spontaneous recall	3
Another word for "A good personal quality"	2
Was it HONESTY, CHARITY, or MODESTY?	0 or 1

Total score 3MS: ____/100

Total score MMS: ____/30

CLOSE YOUR EYES



VITA

Surname: O'Connell Given Names: Megan Eleine

Place of Birth: Thompson, Manitoba, Canada

Educational Institutions Attended:

University of Victoria	1999 to 2001
University of Saskatchewan	1995 to 1998
Simon Fraser University	1992 to 1994

Degrees Awarded:

B.A. (Honours)	University of Saskatchewan	1998

Honours and Awards:

B.A. with High Honours, University of Saskatchewan	1998

Publications:

O'Connell, M. E., Crossley, M., & Shaw, M. F. (1999). Study of age differences in working memory capacity [Abstract]. Journal of the International Neuropsychological Society, 5, 144.

Title of Thesis:

Two Step Screening for Cognitive Impairment: The Clock Test and the 3MS

Author [redacted]

Megan E. O'Connell

October 19, 2001