DETECTING COGNITIVE IMPAIRMENT IN OLDER ADULTS:

A VALIDATION STUDY OF SELECTED

SCREENING INSTRUMENTS

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Patricia McBride-Houtz, B.S., M.S.

Denton, Texas

May, 1993

DETECTING COGNITIVE IMPAIRMENT IN OLDER ADULTS:

A VALIDATION STUDY OF SELECTED

SCREENING INSTRUMENTS

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Patricia McBride-Houtz, B.S., M.S.

Denton, Texas

May, 1993

McBride-Houtz, Patricia. <u>Detecting Cognitive Impair-</u>
<u>ment in Older Adults: A Validation Study of Selected</u>
<u>Screening Instruments</u>. Doctor of Philosophy (Counseling
Psychology), May, 1993, 147 pp., 12 tables, 6 illustrations,
references, 99 titles.

The present study investigated the criterion-based
validity of the Mini-Mental State Examination (MMSE), the
Cognitive Capacity Screening Examination (CCSE), and the
Neurobehavioral Cognitive Status Examination (NCSE) in a
sample of older adults with suspected cognitive impairment.
As cognitive screening tests, the MMSE, CCSE, and NCSE
should predict performance relative to a more thorough
testing procedure. In the present study, performance on the
Halstead-Reitan Neuropsychological Test Battery (HRNTB) was
employed as the criterion measure. Scores on the General
Neuropsychological Deficit Scale (G-NDS), a global per-
formance measure computed from the HRNTB, served as the
standard by which to judge the presence of cognitive
impairment. The sensitivity, specificity, and predictive
value of each screening test, as well as how well each
screening test correlated with the G-NDS, were investigated.

Results of this investigation found that, although the
MMSE, CCSE, and NCSE were all significantly correlated with
the G-NDS, only the NCSE demonstrated an appropriate balance

between high sensitivity and specificity. When a rigorous neuropsychological evaluation was employed as the criterion standard, the NCSE accurately detected the presence of cognitive impairment in 82% of the cases. The MMSE and CCSE, however, failed to detect cognitive deficits in approximately 80% of the cases. These findings strongly suggest that the MMSE and CCSE may have limited utility in the identification of cognitive impairment in older adults.

The heightened sensitivity of the NCSE appears to be the result of several unique features of the instrument, including a multidimensional scoring system and a graded series of increasingly difficult items within each ability area. Future studies need to examine the utility of the NCSE in other geriatric settings, as well as with more diverse populations suffering from a variety of organic mental syndromes.

ACKNOWLEDGEMENTS

I would like to thank several people for making this research possible. I am grateful to Dr. James Hall, my committee chair, for his support throughout the dissertation process. I would also like to thank my committee members, Drs. Harrell, Hipple, Overton, and Watkins. A special thanks goes to Ms. Phyllis Eccleston, whose hard work and flexibility allowed me to meet crucial deadlines.

I want to express my appreciation to the Dallas VA Medical Center and TCOM for allowing me to test patients for this research. I am deeply grateful to Dr. Lynnora Ratliff of the VA for her supervision and support. Thank you to the many older patients who were willing to participate. A special thanks to Mr. Roger Boatwright.

Finally, I am especially grateful for the support that my family provided throughout this project. Thank you to my husband, Andrew, for enduring long hours of supervising and nurturing our rambunctious two-year-old daughter, Keverly. Thank you to my in-laws, Mary and Bill Houtz, for your understanding and support. Thank you to my mother, June McBride, for your love and encouragement. A special thanks to my father, the late Earl McBride, for teaching me the importance of hard work, persistence, and a sense of humor in meeting life's challenges.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

Cognitive Screening Tests

In recent years there has been an exponential growth of interest in dementia and in brain-behavior relationships. A variety of cognitive disorders occurs with increasing frequency as people age, accompanying such progressive dementing conditions as Alzheimer's disease and multi-infarct dementia; other neurological diagnoses such as cerebrovascular accident, anoxia, and multiple sclerosis; and cognitive disorders secondary to psychiatric syndromes (e.g., pseudo-dementia). These cognitive disorders produce considerable morbidity and mortality, and appropriate recognition and management can substantially improve quality of life for both afflicted persons and their caregivers. Thus, it is in the best interest of the patient for health care profes-sionals to become increasingly attuned to the presence of cognitive dysfunction in older adults and familiar with appropriate procedures for evaluation and referral.

Currently, the most widely used screening test for cognitive impairment is the Mini-Mental State Examination (MMSE) developed by Folstein, Folstein, and McHugh (1975). This brief (5-10 min.) test was originally developed to

1

evaluate cognitive dimensions such as orientation, registration, attention and calculation, recall, language (object naming, repetition, comprehension, reading, and writing), and visual-spatial abilities. The test yields a 0- to 30-point global score, with lower scores being associated with diminished performance. The authors suggested that patients who score below 24 points are cognitively impaired. The original validation studies demonstrated the utility of the MMSE for differential diagnosis of clinically diagnosed subgroups of dementia, depression, and cognitively impaired depressives (Folstein et al., 1975).

Subsequent work with medical patients who were being screened by a physician for dementia or delirium suggested that the MMSE has adequate overall sensitivity and specificity for detection of cognitive impairment in neurologic patients, but that it shows diminished sensitivity in older and less well-educated persons (Anthony, LeResche, Niaz, Korff, & Folstein, 1982). Other recent work compared the MMSE with other cognitive screening techniques in a sample of patients with documented brain pathology and found that the MMSE yielded a high rate of false negative decisions (Schwamm, Van Dyke, Kiernan, Merrin, & Mueller, 1987). In addition, a recent study which examined the ability of the MMSE to predict performance on a comprehensive neuropsychological test battery found that use of the MMSE as a cognitive screening instrument resulted in numerous

false-negative classifications (Faustman, Moses, & Csernansky, 1990). Furthermore, some subjects with poor MMSE scores demonstrated little cognitive impairment on the neuropsychological tests (Faustman et al., 1990).

Use of the MMSE has now been expanded to samples different from those employed by Folstein et al. (1975) to validate the scale. Although it was originally developed in a hospital setting, it has been widely applied in epidemiological studies in the United States as a screening instrument for cognitive impairment. In 1986, the Medical Research Council convened a Working Group to recommend that a minimum data set be collected in future research on Alzheimer's disease, and the MMSE was adopted as a measure of cognitive functioning (Brayne & Calloway, 1990). In addition, the National Institute of Mental Health Diagnostic Interview Schedule (NIMH-DIS), a highly structured instrument that can be administered by lay interviewers to make psychiatric diagnoses, also incorporates the MMSE as a screen for organic mental disorders in clinical populations (Escobar, Burnham, Karno, Forsythe, Landsverk, & Golding, 1986). Recently, the MMSE has also been adopted by an increasing number of health care professionals in rehabilitation settings and long-term care facilities to assess the cognitive abilities of older persons who have suffered strokes or who exhibit changes in their mental status over

time (Ebrahim, Nouri, & Barer, 1985; Garcia, Tweedy, & Blass, 1984; McDougall, 1990).

In spite of its wide usage, critics of the MMSE have argued that the validation procedures have relied primarily on correlations with informal descriptive diagnoses and have not employed the most rigorous diagnostic means available, such as comprehensive neuropsychological testing. As a screening test, the MMSE should predict performance relative to more thorough testing procedures such as the Luria-Nebraska or Halstead-Reitan Neuropsychological Batteries. These instruments have demonstrated extensive reliability and validity as tests of cortical function in both psychiatric and neurologic populations (Cullum, Thompson, & Heaton, 1989; Robbins, 1989). Recent research suggests that the MMSE may fail where a true screening test would be most needed: in evaluating patients without manifest organic disease in whom the identification of more subtle cognitive impairment might be crucial to diagnosis, case formulation, and treatment planning (Nelson, Fogel, & Faust, 1986).

Although the MMSE is currently the most popular bedside cognitive screening instrument, two other tests, the Cognitive Capacity Screening Examination (CCSE) and the Neurobehavioral Cognitive Status Examination (NCSE), may prove to be more useful for the assessment of cognitive functioning in older adult and geriatric patient groups. The CCSE is a 30-item questionnaire that was specifically

developed to identify organic brain syndromes (Jacobs, Bernhard, Delgado, & Strain, 1977). It differs from the shorter mental status instruments in that five items are included which measure abstracting ability. In a comparative study of the CCSE, the MMSE, and the Short Portable Mental Status Questionnaire (SPMSQ) with medical patients, Foreman (1987) reported the CCSE to be the most valid and reliable measure of cognitive status.

The NCSE, a relatively new test, represents the beginning of a trend within neuropsychology to assess independent areas of cognitive functioning using a screen and metric approach (Kiernan, Mueller, Langston, & Van Dyke, 1987). In a study comparing sensitivities of the NCSE, CCSE, and MMSE in detecting cognitive deficits in patients with documented brain lesions, the NCSE was found to have a significantly lower false-negative rate (Schwamm et al., 1987).

As with the MMSE, however, validation procedures for the CCSE and NCSE have not employed a rigorous criterion measure as the standard by which to judge cognitive functioning. Thus, although research has been favorable regarding the advantages of using the CCSE and NCSE rather than the MMSE with older adults, neither of these instruments can be considered well validated at the present time.

Given the current popularity of the MMSE for assessing the cognitive functioning of older adults and given the advantages that the CCSE and NCSE may offer in terms of

increased sensitivity and specificity, validation of the MMSE, CCSE, and NCSE against a comprehensive neuropsychological test battery, as opposed to clinical judgment or laboratory techniques, would serve to substantiate whether and to what extent these screening instruments are able to assess the presence and severity of cognitive impairment in elderly patients. The purpose of the current chapter, therefore, is to define and explore the meaning of cognitive impairment in the elderly, to review the empirical literature related to the validation of the MMSE, the CCSE, and the NCSE, to explore the use of the Halstead-Reitan Neuropsychological Test Battery (HRNTB) with older adults, and to establish the validity of the HRNTB with a geriatric population suffering from a variety of organic brain syndromes. The goal of the present study was to examine the relationship between scores obtained on the MMSE, the CCSE, and the NCSE and subsequent performance on the HRNTB in a sample of older adults with suspected cognitive dysfunction.

<div align="center">Cognitive Impairment in the Elderly</div>

Cognitive impairment may be broadly defined as a diminished capacity to know the world (Folstein, Anthony, Parhad, Duffy, & Gruenberg, 1985). The syndromes of dementia, delirium, aphasia, amnesia, and mental retardation are all characterized by cognitive impairment. Dementia is the most common syndrome of cognitive decline seen in the

elderly. Current estimates are that two to three million Americans are demented, including 40% to 60% of patients in nursing homes (Winograd & Jarvik, 1986). "Dementia" is a general term used to describe a chronic and substantial decline in two or more areas of cognitive function. It is distinguished from mental retardation, in which cognitive impairment is lifelong; from aphasia and amnesia, in which language and recent memory are specifically and disproportionately affected; and from delirium, in which cognitive impairment occurs in the context of a reduced level of consciousness (Folstein et al., 1985).

The Diagnostic and Statistical Manual of Mental Disorders (DSM III-R, 1987) puts forth the following diagnostic criteria for dementia:

A. Demonstrable evidence of impairment in short- and long-term memory. Impairment in short-term memory (inability to learn new information) may be indicated by inability to remember three objects after five minutes. Long-term memory impairment (inability to remember information that was known in the past) may be indicated by inability to remember past personal information (e.g., what happened yesterday, birthplace, occupation) or facts of common knowledge (e.g., past presidents, well-known dates).

B. At least one of the following:
    (1) impairment in abstract thinking, as indicated by inability to find similarities and differences between related words, difficulty in defining words and concepts, and other similar tasks;
    (2) impaired judgment, as indicated by inability to make reasonable plans to deal with interpersonal, family, and job-related problems and issues;
    (3) other disturbances of higher cortical function, such as aphasia (disorder of language), apraxia

                (inability to carry out motor activities despite intact comprehension and motor function), agnosia (failure to recognize or identify objects despite intact sensory function), and "constructional difficulty" (e.g., inability to copy three-dimensional figures, assemble blocks, or arrange sticks in specific designs);

      (4) personality change, i.e., alteration or accentuation of premorbid traits.

C. The disturbance in A and B significantly interferes with work or usual social activities or relationships with others.

D. Not occurring exclusively during the course of delirium.

E. Either (1) or (2):

      (1) there is evidence from the history, physical examination, or laboratory tests of a specific organic factor (or factors) judged to be etiologically related to the disturbance;

      (2) in the absence of such evidence, an etiologic organic factor can be presumed if the disturbance cannot be accounted for by any nonorganic mental disorder (e.g., major depression accounting for cognitive impairment (p. 107).

The nature of the onset and progression of cognitive deficits differs greatly among the major dementing disorders. Most of the dementias have an insidious onset and develop slowly and gradually. Alzheimer's disease (AD), Pick's disease, Parkinson's dementia, and progressive supranuclear palsy (PSP) are illustrative of this type of dementia. Multi-infarct dementia (MID) exhibits a stepwise deterioration in intellectual functioning that, early in the course of the disease, leaves some cognitive functions relatively intact. The initial symptoms of MID develop acutely; but, because multiple large or small cerebral infarcts are the cause of the cognitive decline, the ultimate clinical picture can take many years to develop (Albert, 1991).

Personality disturbance or psychiatric syndromes such as depression may also accompany dementing disorders. Whether such changes in personality or mood precede or follow the onset of cognitive decline is critical for an accurate diagnosis. Each dementia has a unique history and, at times, a unique pattern of spared and impaired cognitive functions that can help the clinician identify it. The most common dementia seen in the elderly is AD, followed by MID with some patients displaying histopathologic evidence of both disorders (Adams, Craig, & Parsons, 1986). Overall, AD and MID are thought to account for 80% of the dementias of old age (Adams et al., 1986). The cognitive profile associated with these dementias, as well as a syndrome known as pseudodementia, will be discussed in the following paragraphs.

Alzheimer's Disease

The Diagnostic and Statistical Manual of Mental Disorders (DSM III-R, 1987) states that the essential feature of primary degenerative dementia of the Alzheimer type is

the presence of dementia of insidious onset and a generally progressive, deteriorating course for which all other specific causes have been excluded by the history, physical examination, and laboratory tests. The dementia involves a multifaceted loss of intellectual abilities, such as memory, judgment, abstract thought, and other higher cortical functions, and changes in personality and behavior (pp. 119-120).

The DSM III-R (1987) further states that, in the majority of cases of persons with AD, "the brain is atrophied, with

widened cortical sulci and enlarged cerebral ventri-
cles . . . (Postmortem) Microscopic examination usually
reveals three histopathologic changes: senile plaques,
neurofibrillary tangles, and granulovacuolar degeneration of
neurons" (p. 120).

The first and most noticeable symptom generally
observed in patients with AD is a severe anterograde memory
deficit (Cummings & Benson, 1986). Early in the course of
the disease, secondary memory is largely affected; but, as
the disease progresses, deficits in primary memory also
develop (Albert, 1991). In addition to memory impairment,
recent data suggest that the other cognitive deficit most
commonly seen in the early stages of AD is difficulty with
sequencing, monitoring, and shifting behavior (Grady, Haxby,
Horwitz, & Sundaram, 1989; Morris & Fulling, 1983). Such
deficits have typically been attributed to frontal lobe
dysfunction (Damasio, 1985; Stuss & Benson, 1986). However,
problems with complex attentional mechanisms secondary to
parietal lobe abnormalities may also be responsible for such
impairments (Grady et al., 1989).

In the most typical presentation of AD, language defi-
cits and spatial deficits develop after the onset of memory
problems (Bayles & Kaszniak, 1987). In neuropsychological
testing, patients early in the course of AD often score
within the average range on IQ tests but have substantial
difficulty with memory, shifting set, and conceptualization

and slight difficulty with naming (Albert, 1991). There is
often a significant difference between IQ and the memory
quotient obtained on the Wechsler Memory Scale-Revised. A
person's memory quotient should be approximately equal to
the IQ. Impairments in set shifting and abstract thought
are often revealed by performance on the Trail Making Test.
This task requires an individual to first connect a series
of numbers in order and then connect alternating numbers and
letters in order (e.g., 1-A, 2-B, etc.). Mildly impaired AD
patients are generally slow on both tasks and tend to make
errors on the second (Albert, 1991).

Multi-Infarct Dementia

Cerebrovascular disease most commonly presents
clinically as the "stroke syndrome" (Mohr, Fisher, & Adams,
1980). Although not all forms of vascular disease involve
stroke (cardiac arrest, prolonged hypotension), the disor-
ders that produce dementia generally result from multiple
strokes over time (Albert, 1991). These have been labeled
multi-infarct dementia (Hachinski, Lassen, & Marshall, 1974)
to emphasize the fact that cognitive deficits result from
actual infarcts and not from diffuse narrowing of blood
vessels.

According to the DSM III-R (1987), the essential fea-
ture of multi-infarct dementia is

> a dementia due to significant cerebrovascular dis-
> ease . . . The onset is typically abrupt. The course

is stepwise and fluctuating, with rapid changes, rather
than uniformly progressive. The pattern of deficits is
"patchy," depending on which regions of the brain have
been destroyed. Certain cognitive functions may be
affected early, whereas others remain relatively unim-
paired. The dementia typically involves disturbances
in memory, abstract thinking, judgment, impulse con-
trol, and personality (pp. 121-122).

According to Albert (1991), multi-infarct dementia may

be characterized by at least two clinical pictures. When

large-vessel disease produces multiple cerebral emboli,

large discrete cerebral infarcts typically occur. Depending

on the anatomic distribution of the lesions, resulting focal

cognitive deficits may include aphasia, apraxia, agnosia,

and amnesia. Repeated strokes eventually cause the develop-

ment of global cognitive deficits. On the other hand,

medium- or small-vessel disease, secondary to atherosclero-

sis, produces more incomplete, diffuse infarction of brain

tissue. The latter is also known as lacunar disease and is

difficult to differentiate from progressive primary

dementias such as AD. Thus, a careful medical history and

knowledge about the onset of cognitive difficulties and

their course is critical in making an accurate diagnosis.

Neuropsychological testing of patients with lacunar

disease may reveal cognitive deficits reflective of aphasia.

A mildly impaired patient may write "squar" for "square" on

the Reitan-Indiana Aphasia Screen (RIAS) or might write a

sentence like "I came your by automobile" on the MMSE.

Repetition and naming may be impaired, as well as drawings

that require alternation. Memory impairment, however, may be variable, and orientation is likely to be intact (Albert, 1991).

Therefore, MID may present with global deterioration of cognitive functioning or circumscribed cognitive deficits, depending on whether the damage to brain tissue is focal or diffuse. For this reason, a consistent cognitive profile has been difficult to identify. At present, the most useful information in the diagnosis of MID tends to be provided by a careful medical history, neuro-imaging techniques (such as magnetic resonance imaging--MRI), and documented functional impairment on neuropsychological tests.

Pseudodementia

Persons suffering from a major depressive episode may complain of memory problems and difficulty concentrating, and may evidence an overall reduction in intellectual abilities. Cognitive deficits arising secondary to depression characterize a syndrome commonly referred to as "pseudodementia." The differential diagnosis between pro-gressive degenerative dementia such as AD and depression is therefore a difficult clinical task.

Although similar cognitive deficits may be seen in both dementia and pseudodementia, recent research suggests several distinguishing features related to the test perfor-mance of the two populations that can assist in making an

accurate diagnosis. In a comparison of normal elderly subjects with pseudodementia, elderly patients with a major depressive disorder or other major functional psychiatric disorder, and AD patients on a choice reaction-time task, normals performed better than depressed patients with cognitive deficits who, in turn, outperformed AD patients (Rabins, 1983). More importantly, patients with pseudo-dementia and normals, but not AD patients, displayed relatively improved reaction times when given positive feedback about their performance and when instructed to relax before proceeding with the task (Pirozzolo, Christen-sen, & Ogle, 1981).

On memory tasks, depressed elderly patients have been shown to respond more conservatively than AD patients, resulting in fewer errors (Larner, 1977). Compared to AD patients, elderly depressed patients have been shown to make better use of cognitive organizational strategies to facili-tate memory, perform better on minimally demanding cognitive tasks, and have somewhat better access to over-learned memories (Weingartner, Kaye, Smallberg, Cohen, Ebert, Gillin, & Gold, 1982). On the other hand, both AD patients and depressed patients have difficulty on cognitive tasks requiring sustained attention and motivation as well as recalling sequences of recent events (Weingartner et al., 1982). Cognitive functions which remain relatively intact in depression, however, include recognition of high-imagery

words (Silbermann, Weingartner, Laraia, Byrnes, & Post, 1983), recall of related words that have previously been sorted (Weingartner, Gold, Ballenger, Smallberg, Summers, Rubinow, Post, & Goodwin, 1981), paired associate learning (Breslow, Kocsis, & Belkin, 1980), and naming and arithmetic ability (Caine, 1986).

In addition to the differences mentioned above, it is essential to clarify the order in which the cognitive deficits occurred relative to the mood disturbance (LaRue, 1982). For example, if memory difficulties and concentration problems began after the onset of depressive symptoms, this clinical picture is more suggestive of pseudodementia. However, if cognitive problems preceded the onset of depression, it is more likely that a true dementia exists. Also, the test performance of depressed persons is likely to be more variable among tests of a similar nature (e.g., verbal memory tasks) than that of demented persons, who would typically perform poorly across the board within the same functional domain (Albert, 1991). Finally, the possibility of depression superimposed on a true dementia always exists. According to the DSM III-R (1987), "a therapeutic trial with an antidepressant drug . . . may clarify the diagnosis" (p. 106). If cognitive difficulties are arising secondary to depression, improvement in cognitive functioning is usually seen as the mood disorder improves. In the case of depression superimposed on a pre-existing dementing

condition, a drastic improvement in cognitive functioning would not be expected (Adams et al., 1986).

Validation of the Mini-Mental State Examination

In order for a screening instrument to be useful, it must demonstrate acceptable reliability and validity. In the broadest sense, reliability refers to the consistency of scores obtained by the same person when reexamined with the same test, whereas, validity concerns <u>what</u> the test measures and <u>how well</u> it does so (Anastasi, 1982). Although a test may yield consistent scores and thus be reliable, reliability does not ensure validity. For example, if a person obtains the same MMSE score on different days, this can be taken as evidence of test-retest reliability, but obtaining a consistent score on the MMSE does not give information about what the instrument really measures. The trait measured by a given test can be defined only through an examination of the objective sources of information and empirical operations utilized in establishing its validity (Anastasi, 1950). Furthermore, the choice of validation procedure depends on the use to be made of the test scores. The same test, when employed for different purposes, should be validated in different ways (Anastasi, 1982).

In addition to reliability and validity, instruments must have sensitivity, specificity, and predictive value, all of which are necessary for accurate measurement. A

screening instrument is sensitive if it correctly classifies
a characteristic, it is specific if it correctly identifies
the absence of a characteristic, and it is predictive if a
positive characteristic identified is truly present
(McDougall, 1990). There have been varying reports con-
cerning the reliability, validity, sensitivity, specificity,
and predictive utility of the MMSE. In the following para-
graphs, the literature pertaining to the psychometric
properties of the MMSE will be reviewed and critiqued.

Reliability data for the MMSE appear in five studies.
Three of these are found in the original paper by Folstein
et al. (1975) and two in a paper by Dick, Guiloff, Stewart,
Blackstock, Bielawska, and Paul (1984). When the MMSE was
administered twice, 24 hours apart by the same tester on
both occasions, the correlation by a Pearson coefficient was
0.89. Scores were not significantly different using a Wil-
coxon $\underline{T}$. To assess examiner effect on 24-hour test-retest
reliability the MMSE was given twice, 24 hours apart by two
examiners. The Pearson $\underline{r}$ remained high at 0.83, and the
Wilcoxon $\underline{T}$ was not significant. In addition, when elderly
depressed and demented patients chosen for their clinical
stability were given the MMSE twice, an average of 28 days
apart, there was no significant difference in their scores
by the Wilcoxon $\underline{T}$ and the product moment correlation between
scores was 0.98 (Folstein et al., 1975). Similarly, Dick et
al. (1984) found that, in a neurological population using

the Wilcoxon matched-pairs signed rank test, there was no difference between two MMSE scores of 15 patients tested by different examiners at an interval of 24 hours. This was also true for 30 patients tested by the same examiner at a 24-hour interval, and for 14 clinically stable patients tested by the same examiner at a mean interval of 31 days.

For a comparison with the test-retest reliability obtained by Folstein et al. (1975), Pearson correlation coefficients were computed and were close to those quoted by Folstein (0.92 for patients retested by the same examiner and 0.95 for patients retested by a different examiner). Taken together, the results of these studies suggest that both test-retest and inter-rater reliabilities are probably satisfactory for the MMSE.

Although it appears that the MMSE is a reliable instrument, the issue of validity is more critical. Folstein et al. (1975) attempted to establish the validity of the MMSE in five separate studies. In the first study, the MMSE performance of 132 psychiatric inpatients, 69 of whom were diagnosed as depressed or demented (mean age 66.4 years), was compared to the performance of 63 "normal" control subjects (mean age 73.9 years). The mean MMSE score was 9.6 for demented patients, 19.0 for cognitively impaired depressives, 25.1 for depressed patients, and 27.6 for control subjects. Folstein et al. (1975) concluded that MMSE scores "agreed with the clinical opinion of the presence of

cognitive difficulty and as the cognitive difficulty is usually less in depression than in dementia, the scores dispersed in a fashion agreeing with the severity of the difficulty" (p. 192).

In an effort to control for age effects, an age-matched group was drawn from the sample of the first study (8 subjects per group with an age range of 69-86 years) to evaluate whether MMSE scores discriminated between the clinically diagnosed subgroups. Results indicated that MMSE means were significantly different for the three groups (i.e., demented, cognitively impaired depressives, and depressives). In a third analysis, patients with dementia (N = 14), depression (N = 12), and cognitively impaired depressives (N = 7) from the first study were retested after "appropriate treatment." Folstein et al. (1975) predicted that, if the MMSE is a valid test of cognitive state, patients with dementia would be expected to show little change in their scores, whereas those with depression and pseudodementia should improve their scores after treatment. Results supported this hypothesis and found that the mean MMSE scores of the depressed and pseudodementia groups improved significantly, but the mean scores of the demented patients did not.

In a fourth study, Folstein et al. (1975) examined the MMSE score of 137 consecutive private psychiatric hospital admissions. Results suggested that the mean MMSE scores for

demented patients ($\underline{N}$ = 9) differed significantly from the mean scores of the other diagnostic groups. The mean MMSE scores for the clinically diagnosed groups were as follows: dementia = 12.2, depression = 25.9, mania = 26.6, schizophrenia = 24.6, personality disorder = 26.8, and neuroses = 27.6 (Folstein et al., 1975).

In a fifth and final study, concurrent validity was determined by correlating MMSE scores with the Verbal and Performance scores of the Wechsler Adult Intelligence Scale (WAIS) in a group of patients from the original sample ($\underline{N}$ = 26) who had been administered the MMSE and WAIS during the same week. Diagnostic subgroups included dementia ($\underline{N}$ = 8), pseudodementia ($\underline{N}$ = 8), depression ($\underline{N}$ = 8), schizophrenia ($\underline{N}$ = 2), and neurosis ($\underline{N}$ = 1), with ages ranging from 22 to 78 years. For the MMSE vs. Verbal IQ, Pearson $\underline{r}$ was 0.78, and for the MMSE vs. Performance IQ, Pearson $\underline{r}$ was 0.66.

Several conclusions can be drawn from the original validation studies carried out by Folstein et al. (1975). First of all, results suggest that patients with clinically diagnosed dementia will have MMSE scores of less than 24 at least 75% of the time, and non-psychotic psychiatric inpatients without diagnosed organic mental disorders usually score 20 or higher (Nelson, Fogel, & Faust, 1986). Patients suffering from depression with cognitive impairment often scored less than 24 but usually scored higher than 10. In addition, patients diagnosed with depression and

pseudodementia usually improved their MMSE scores following successful treatment of their depression.

Although it would be useful to have data on the average MMSE performance for different diagnostic groups, this information alone cannot be employed to establish the validity of the MMSE has a measure of cognitive functioning. The major flaw of the original validation studies by Folstein et al. (1975) is their use of clinical diagnosis as the sole criterion to establish the presence or absence of cognitive impairment. This is a common procedure in the development of certain personality tests where psychiatric diagnosis, or so-called "clinical judgment," is used both as a basis for the selection of items and as evidence of test validity. In the realm of judging cognitive functioning, however, psychiatric diagnosis cannot be regarded as a criterion measure but, rather, as an indicator or predictor whose own validity would have to be determined. A rigorous, standardized diagnostic procedure should serve as the standard for comparison, as opposed to clinical judgment.

The only study which approaches this was the fifth one, in which MMSE scores were correlated with WAIS Verbal and Performance IQ. This is not without problems, however, because the utility of the WAIS in neuropsychological evaluation is a matter of debate. Some view the WAIS as an index of academically related intellectual skills but not as a valid measure of overall neuropsychological functioning

(Franzen, 1989). Patterns of performance on the WAIS, such as the use of hold-don't hold scales, have been suggested to be useful in the identification of neuropsychological deficits (Fuld, 1983), and Russell (1987) has suggested strategies for interpreting the WAIS in a neuropsychological setting. However, many of these suggestions are in need of rigorous experimental validation. It is interesting to note that the MMSE correlates more highly with Verbal IQ rather than Performance IQ, yet organicity is typically detected more with the Performance subtests (Reitan & Wolfson, 1975). Also, certain profound neuropsychological deficits can coexist without any appreciable loss in measured intelligence (Albert, 1991). Thus, although the WAIS is a better criterion measure than mere informal, descriptive diagnosis, it is not generally considered a rigorous, comprehensive measure of all aspects of cognitive functioning.

Research conducted since the original work by Folstein et al. (1975) has investigated the validity of the MMSE with various populations, including medical patients with known or suspected dementia, neurological patients, psychiatric patients, institutionalized elderly, and community-dwelling elderly. Other variables explored in relation to performance on the MMSE are age, ethnicity, socioeconomic status, and education. Also, recent studies have investigated the utility of performance on the MMSE as a predictor of daily functional abilities in dementia patients. With

one notable exception, these investigations continue to rely upon inadequate criteria such as informal descriptive diagnosis to serve as the criterion measure of cognitive functioning. The results of these validation studies will be reviewed below.

In a study by Anthony, LeResche, Niaz, Korff, and Folstein (1982), hospital patients on a general medical ward ($\underline{N}$ = 97) were administered the MMSE, and scores were correlated with a psychiatrist's standardized clinical diagnosis of delirium or dementia. Results suggest that the MMSE was 87% sensitive and 82% specific in detecting dementia and delirium in this population. The false positive ratio was 39%, and the false negative ratio was 5%. All false positives had less than 9 years of education, and many were 60 years of age or older. The authors concluded that, due to the relatively high false positive ratio and the concentration of older and poorly educated individuals among the false positives, the MMSE should not be used as the sole criterion for diagnosing dementia or delirium (Anthony et al., 1982).

Comparable rates of sensitivity and specificity were also obtained by Kafonek, Ettinger, Roca, Kittner, Taylor, and German (1989) in a study of institutionalized elderly ($\underline{N}$ = 70). These investigators compared the diagnostic accuracy of the MMSE against a standardized psychiatric interview and found that the MMSE was 81% sensitive and 83% specific in

screening for dementia alone. If screening for either dementia or delirium, the MMSE was found to be 86% specific and 79% sensitive. The false positive rate was 7.5, and the false negative rate was 34. Thus, in this population, the likelihood of true dementia in a patient scoring in the impaired range (i.e., positive predictive value) was 93%. However, the likelihood that dementia was absent in patients scoring in the normal range (i.e., negative predictive value) was only 65%. This represents a substantial false negative rate. For any test that is truly to function as a screen, false-negative errors are more serious than false-positive errors (Nelson, Fogel, & Faust, 1986). Results such as these indicate that the MMSE may be insensitive to milder cognitive deficits or moderate focal deficits which are insufficiently evaluated by the item content of the MMSE.

One content area of the MMSE upon which physicians rely heavily in their evaluation of cognitive impairment is the orientation section. It has become a common practice for physicians to ask questions of orientation alone as a quick means of evaluating cognitive function. In a univariate and multivariate analysis of the MMSE in diagnosing dementia (Klein, Roca, McArthur, Vogelsang, Klein, Kirby, & Folstein, 1985), the sensitivity of the orientation items was found to be low (15.3% to 56.9%), although specificity was high (91.7% to 100%). High sensitivity and specificity were

achieved when orientation and non-orientation items were combined (89.6% and 78.1%, respectively). From these results, the authors concluded that orientation items alone are unacceptably insensitive in detecting dementia. They further note that the diagnostic sensitivity of the orientation items is especially poor in patients with milder degrees of impairment (Klein et al., 1985).

Degree of impairment in dementia patients is indeed a critical variable in the validation research on the MMSE. Because Alzheimer's disease is a progressive dementia, one way investigators have studied severity of impairment in relation to MMSE performance has been to compare MMSE scores relative to duration of illness. In a study of 141 AD patients, performance on the MMSE was negatively correlated (r = -.50, p < .001) with duration of illness (Teng, Chui, Schneider, & Metzger, 1987). In addition, subjects whose age at onset was younger than 65 years performed more poorly than those whose age at onset was 65 or older, but the two groups showed comparable slopes of performance decline for duration of illness (Teng et al., 1987). Again, determination of cognitive functioning was assessed subjectively by clinical judgment.

In a similar study, the performance of 92 patients with probable AD was evaluated in a longitudinal design which compared three standardized mental status examinations, including the MMSE. The Information-Memory-Concentration

Test (IMC), the Dementia Rating Scale (DRS), and the MMSE
were administered and then readministered approximately one
year and two years after the initial assessment. Results
indicated that the MMSE was sensitive to the progression of
cognitive dysfunction in AD, although the DRS was found to
be more sensitive in severely demented patients (Salmon,
Thal, Butters, & Heindel, 1990). These results are not
surprising, given that the DRS is a more comprehensive
instrument than the MMSE and, as such, takes more time to
administer and evaluate.

Although the relationship between duration of illness
and subsequent performance decline on the MMSE suggests that
the MMSE is sensitive to deterioration of cognitive func-
tioning over time, it is not direct evidence of the validity
of the instrument because no objective, independent measure
of cognitive processes was carried out. In a more compre-
hensive investigation, Foreman (1987) attempted to address
this problem by evaluating the content, criterion-related,
and construct validity of the MMSE, along with two other
popular mental status tests, the Short Portable Mental
Status Questionnaire (SPMSQ) and the Cognitive Capacity
Screening Examination (CCSE). Subjects were 66 elderly (65
years and older) hospitalized medical-surgical patients
clinically diagnosed with dementia or delirium, or judged to
have normal cognitive functioning. Content validity of the
three instruments was determined by reviewing and

summarizing the psychometric and clinical literature about the three tests. To assess criterion-related validity, scores obtained on the MMSE, SPMSQ, and CCSE were correlated with the clinical diagnosis of global cognitive impairment using Spearman correlation coefficients. Convergent and discriminant validation of the three instruments was used to assess construct validity, based on results obtained on the Dementia Rating Scale (DRS) and the Visual Analogue Scale for Depression (VAS).

Results of the content validation analysis indicated that, of the three instruments, the MMSE was less comprehensive in regard to areas of cognitive functioning assessed than the CCSE, but more comprehensive than the SPMSQ. For criterion-related validity, the Spearman correlation coefficient was 0.78 for the MMSE (p < .001), which was higher than the SPMSQ, but lower than the CCSE. As for construct validity, correlation coefficients among the MMSE, SPMSQ, CCSE, and DRS were all high and statistically significant (p < .001), thereby providing evidence of convergent validity. Conversely, correlations of the MMSE, SPMSQ, CCSE, and DRS with the VAS were low, providing evidence of discriminant validity.

Two important considerations should be kept in mind when evaluating the results of the Foreman (1987) study. First of all, the criterion employed to determine criterion-related validity was clinical diagnosis, not a rigorous,

independent, and well-validated measure of cognitive functioning. Second, for an instrument to demonstrate convergent validity, it must correlate highly with other variables with which it should theoretically correlate (Anastasi, 1982). Although a high correlation between the MMSE and DRS is desirable, the DRS is not a sophisticated measure of cognitive functioning and is often unable to detect mild or focal cognitive deficits, which, as mentioned earlier in this review, is where a true screening test would be most needed. According to Folstein et al. (1975), the MMSE concentrates only on the cognitive aspects of mental functions, and "within the cognitive realm it is thorough" (p. 189). The authors further state that the MMSE "separates patients with cognitive disturbance from those without such disturbance" (p. 195). If this is true, convergent validation procedures should employ rigorous, comprehensive tests of cognitive functioning to serve as the basis of comparison.

Although not intended as a validation study per se, research by Farber, Schmitt, and Logue (1988) evaluated the correlation between scores on the MMSE and WAIS-R full scale IQ in a sample of patients in the early stages of AD. Results indicated that MMSE scores correlated 0.83 with full scale IQ. The authors concluded that the MMSE may be "a reasonable alternative measure of overall intellectual functioning" (p. 509). Given that the validity of the MMSE as a

screen for cognitive functioning has not been well established, it seems like quite a leap to infer that it may serve as a valid measure of overall intellectual abilities as well. The problems associated with using the WAIS or WAIS-R as a measure of neuropsychological functioning have been discussed earlier in this review and will not be repeated here. However, it is important to note that in this study Verbal IQ was not distinguished from Performance IQ, which could at least give some information about how well the MMSE correlates with measures of verbal as opposed to visual-spatial skills.

The correlation of MMSE scores with WAIS scores has also been employed as the criterion in an investigation of the usefulness of the MMSE in a neurological population. As part of a larger investigation, Dick et al. (1984) analyzed the WAIS performance of a subgroup of neurological patients ($\underline{N}$ = 37) judged to have cognitive impairment by a neurologist. Results yielded Spearman correlation coefficients of 0.55 for MMSE scores with Verbal IQ and 0.56 for MMSE scores with Performance IQ. These correlations are substantially lower than those previously reported (Folstein et al., 1975). In the larger investigation, Dick et al. (1984) assessed the MMSE performance of 126 consecutive neurological/neurosurgical admissions (mean age 49.9 years). Diagnosis was established clinically with the aid of computerized axial tomography (CT scans), angiography, and

biochemical tests. Correlation with lesion site and with neurologists' clinical diagnosis of cognitive impairment served as the criteria.

Results indicated that 76% of patients with cognitive impairment ($\underline{N}$ = 50) but only 4.3% of normals ($\underline{N}$ = 93) scored higher than 24 points on the MMSE. However, some of the patients classified as cognitively impaired on clinical grounds had total MMSE scores as high as 27, and 12 out of 50 had MMSE scores of 24 points or more.

In a comparison of MMSE scores in focal versus generalized brain disease, the MMSE total scores in those with right hemisphere disease did not differ from the control group (all scored 24 points or higher), but they were greater than those with left hemisphere and bilateral hemisphere disease. The MMSE total scores in the left hemisphere group did not differ significantly from the bilateral hemisphere group; however, MMSE scores of those with left hemisphere disease and bilateral hemisphere disease were lower than those in the control group. Dick et al. (1984) concluded from these results that the MMSE did not seem useful in differentiating focal from "diffuse" brain disease and may not be "an entirely reliable indicator of cognitive function" (p. 498) in neurological patients. The authors further note that the MMSE was relatively insensitive to damage of the right cerebral hemisphere (Dick et al., 1984).

In an earlier study, DePaulo and Folstein (1978) also attempted to validate the MMSE as a measure related to cerebral disorder. Subjects in this investigation were neurology inpatients ($\underline{N}$ = 126, mean age 50.2 years) clinically diagnosed with cerebral lesions or as having exclusively peripheral disorders with an absence of cerebral disturbance. All patients with exclusively peripheral disorders scored at least 24 points on the MMSE. In contrast, 50% of the patients with cerebral lesions scored less than 24 points on the MMSE. Thus, half of the patients with cerebral abnormality had cognitive defects that were detected by the MMSE.

From these results, the authors concluded that the MMSE is "a valid measure of cognitive defect related to cerebral disorder" (DePaulo & Folstein, 1978, p. 226). If half of the subjects with cerebral abnormality scored below the cutoff for cognitive impairment on the MMSE, half of them scored above the cutoff, and this is not particularly supportive evidence of the claim that the MMSE is sensitive to cognitive impairment in a neurological population. Of those who scored above the cutoff, it may be that the MMSE was unable to detect cognitive impairment in these subjects. To interpret the unimpaired MMSE performance of half of the cerebral abnormality group as evidence of a lack of cognitive impairment without an independent measure to verify the cognitive functioning of subjects is circular reasoning, to

say the least. Clearly, the results do not support the contention that the MMSE demonstrates high sensitivity and specificity in detecting cognitive dysfunction in a neurological population.

Documenting structural brain lesions with CT scans represents the latest trend in the research on the MMSE in neurological populations. Tsai and Tsuang (1979) compared the results of the MMSE with findings from CT scans of the brain to determine whether the MMSE could predict organicity and also discriminate among types of organic brain conditions. Subjects ($\underline{N}$ = 63) were referred for CT scans from the neurology and psychiatry departments of the same medical center due to suspected organic brain syndrome. In order for the CT printout to be judged positive, it had to show evidence of cerebral atrophy and/or focal lesions. Results indicated that 18 patients with diffuse atrophy scored 18.0 ($\underline{SD}$ = 8.6) on the MMSE; 10 patients with focal lesions alone scored 25.3 ($\underline{SD}$ = 5.4); 32 patients with normal CT scans scored 26.4 ($\underline{SD}$ = 5.6). The mean total MMSE score for patients with negative CT scans was significantly higher ($\underline{p}$ < .01) than for patients with positive CT scans. The mean total MMSE score for patients with cerebral atrophy (with or without focal lesions) was significantly lower ($\underline{p}$ < .05) than mean scores for patients with focal lesions only. However, there appeared to be no significant difference in

MMSE performance between patients with negative CT scans and those with only focal lesions (Tsai & Tsuang, 1979).

In a study with similar methodology, Schwamm et al. (1987) compared the MMSE performance of 30 neurosurgical patients with documented brain lesions with their performance on two other screening instruments, the Cognitive Capacity Screening Examination (CCSE) and the Neurobehavioral Cognitive Status Examination (NCSE), to determine which instrument was more sensitive in the detection of cognitive dysfunction. One advantage of this study was that conspicuously demented or delirious patients were excluded, allowing for a better assessment of the utility of the three instruments in detecting cognitive deficits in those patients without clinically obvious impairments. Results indicated that the NCSE identified cognitive impairment in 28 patients, the MMSE in 16, and the CCSE in 13 patients. The CCSE had a false-negative rate of 53%, the MMSE of 43%, and the NCSE of 7%. The authors reported that 6 of the 30 patients had a MMSE score of 27 or higher. These high scores demonstrate that nearly perfect scores on the MMSE cannot be taken as evidence that no significant central nervous system lesions exist. Based on these results, the authors concluded that the MMSE is unacceptably insensitive to the presence of cognitive deficits, and may have limited utility in the assessment of patients with cognitive disorders (Schwamm et al., 1987).

The glaring problem with studies that employ documented brain lesions (via CT scans) as the independent variable, or standard, by which cognitive impairment is determined, is that not every brain lesion or atrophic condition produces cognitive deficits. Cognitive impairment can be determined only behaviorally by observable performance decrements on objective tests specifically designed to assess cognitive abilities. Dementia, in particular, is a diagnosis based on behavior and cannot be determined by CT scans, MRI techniques, electroencephalography (EEG), or other laboratory instruments, although specific causes of dementia may be identified by these means (NINCDS-ADRDA Work Group, 1985). It will be recalled that results of a previously reviewed study indicated that the MMSE scores of patients with negative CT scans (i.e., "normals") were not found to be significantly different from the MMSE scores of patients with focal lesions (Tsai & Tsuang, 1979). Results such as these are inconsistent with the notion that documented structural damage equals cognitive impairment. If this were the case, patients with focal lesions would have performed worse on the MMSE than normals. Thus, documented brain lesions do not represent an "objective" measure of cognitive impairment, and cannot be used to establish the criterion-related validity of cognitive screening instruments.

In the current review, only one study was located which employed modern neuropsychological methods in the validation

procedure. Faustman, Moses, and Csernansky (1990) examined the ability of the MMSE to predict level of performance on the Luria-Nebraska Neuropsychological Battery (LNNB) in a diagnostically mixed sample of 90 psychiatric inpatients. Findings indicated that use of the MMSE as a cognitive screening instrument resulted in numerous false negative classifications when performance on the LNNB was used as the criterion measure. In fact, the MMSE failed to detect cognitive impairment in nearly 80% of the cases where the LNNB showed clear evidence of cognitive deficits (i.e., 5 or more scales above critical level). Furthermore, some patients with poor MMSE scores demonstrated little cognitive impairment on the LNNB.

The latter finding raises the question of how persons could score poorly on the MMSE, yet not have any real cognitive impairment (i.e., false-positives). Cognitive scales are known to be affected by the major sociodemographic variables, and the MMSE is no exception. The influences of age, socioeconomic status (SES), ethnicity, and education have been examined in several studies. Results suggest that lower scores on the MMSE are associated with increasing age, lower SES, and lower educational achievement (Brayne & Calloway, 1990; Cavanaugh & Wettstein, 1983; Escobar et al., 1986; Uhlmann & Larson, 1991). Although Anthony et al. (1982) reported that the specificity of the MMSE was lower for black as compared to white patients (.78 vs. .94), most

of this difference appeared to be an artifact of educational
status. Older persons and low SES individuals also tend to
be less educated. Uhlmann and Larson (1991) found that,
when comparing age, race, and education, only education was
independently associated with MMSE scores at a statistically
significant level.

There has been debate about whether education should be
corrected for by the use of different cut-off points on the
MMSE. Uhlmann and Larson (1991) found that the most accu-
rate lower limits of normal for MMSE scores and their
respective sensitivities and specificities were 21 for
middle school graduates (.82/.94), 23 for high school
graduates (.79/.97), and 24 for college/graduate school
graduates (.83/1.00). These education-specific norms accu-
rately classified over 90% of subjects in all three
educational strata. This would seem to overcome the
influence of education on MMSE scores. However, adjusting
for variables such as education reduces the ability to
examine them as independent risk factors themselves.

According to Berkman (1986), the question of central
importance is whether education and other variables such as
SES are of etiologic significance in the development of
dementia or whether the association reflects a stable
characteristic of the individual in his or her mental per-
formance which leads to differential misclassification and
detection bias. If we are sure that educational level

influences MMSE scores exclusively via the latter pathway,
various adjustment procedures might well be justified.
However, if there is a possibility that some part of the
association between educational level and MMSE scores is the
result of the influence of this factor on a disease process
ultimately resulting in mental deterioration, it would be a
mistake to "adjust" for such a factor.

Berkman (1986) notes that education level reflects many
factors--social, environmental, behavioral, psychological,
and biological. People with different levels of education
are differentially exposed to occupational and environmental
hazards. They experience different social stressors and
have different behaviors with regard to alcohol and ciga-
rette consumption, eating patterns, and physical activity.
In addition, recent neuropsychological research indicates a
relationship between environmental stimulation and brain
function (Greenough & Green, 1981). Although it is unlikely
that lack of stimulation could cause a disease, it does seem
plausible to assume that such environmental conditions could
retard or advance the rate at which dementias progress.
Again, such environmental stimulation may be highly corre-
lated with educational level. Adjusting for education will
not allow investigators to explore the potential etiologic
importance of this variable. At the present time, many
investigators recommend examining factors such as education
and SES as potential risk factors rather than obscuring

their influence by adjustment procedures (Berkman, 1986; Brayne & Calloway, 1990; Heeren & Rooymans, 1991; Uhlmann, Teri, Rees, Mozlowski, & Larson, 1989).

For the purposes of the present study, the important consideration is whether utilizing education-specific norms serves to increase the sensitivity and specificity of the MMSE when performance on a comprehensive neuropsychological evaluation is employed as the criterion standard. Given that false-negative decisions have posed a greater threat to the validity of the MMSE, adjusting for education may only serve to decrease its ability to accurately detect the presence of cognitive deficits.

Studies reviewed thus far have examined the concurrent validity of the MMSE--that is, the majority of research has been concerned with the ability of the MMSE to accurately detect the presence of cognitive impairment. However, MMSE scores are often used to make predictions about an individual's daily functional abilities. This would necessitate the determination of a different type of validity: predictive validity. To investigate whether the MMSE can provide an adequate estimate of daily function, Reed, Jagust, and Seab (1989) evaluated the relationship between MMSE scores and functional abilities as measured by activity of daily living (ADL) scales in a sample of elderly demented patients. Results indicated that MMSE scores explained only about one-third of the variance in both the more basic

"physical" ADLs (e.g., grooming, eating) and the more com-
plex "instrumental" ADLs (e.g., managing money, using the
telephone). Thus, although MMSE scores were broadly predic-
tive of ADLs, they explained less than half of the variance
in ADLs at best, and were a worse predictor of ADLs for
mildly demented patients than for severely demented pa-
tients. These results have been replicated by Aske (1990).

For clinicians, the purpose of measuring cognitive
impairment is often to gain a sense of the patient's func-
tional abilities in order to judge the appropriate level of
care needed. Given that the validity of the MMSE as a mea-
sure of cognitive functioning has generally been considered
adequate, it is not surprising that health care profes-
sionals have used MMSE scores to predict functional
abilities and, thus, make what they considered to be better
decisions regarding treatment, rehabilitation efforts, and
discharge planning. As studies suggest, however, MMSE
scores are not good predictors of functional impairment,
especially in the early stages of dementia.

It is the assumption that the MMSE demonstrates ade-
quate validity that has prompted clinicians and researchers
to routinely use the MMSE as a measure of cognitive func-
tioning and also to use the MMSE to validate other screening
instruments (Burch & Andrews, 1987; Herst, Voss, & Waldman,
1990). As can be seen from the present review, however,
high sensitivity and specificity are achieved only when MMSE

scores are correlated with the clinical diagnosis of
delirium or dementia. Even when clinical diagnosis defines
the criterion measure, research suggests that the MMSE evi-
dences a higher false positive rate with less well educated,
lower SES persons and older persons. When the criterion
measure is documented brain pathology, the MMSE appears to
be insensitive to milder cognitive deficits or moderate
focal deficits. In addition, the MMSE appears to be re-
latively insensitive to damage of the right cerebral
hemisphere and is not useful in differentiating focal from
diffuse brain disease. When a comprehensive neuropsycho-
logical battery was used as the standard for comparison in a
diagnostically mixed sample of psychiatric inpatients, the
MMSE failed to detect cognitive impairment in 80% of the
cases. Clearly, the validity of the MMSE has not been well
established at the present time.

<div align="center">Alternative Cognitive Screening Instruments</div>

Cognitive Capacity Screening Examination

Although the MMSE is currently the most widely used
bedside cognitive screening instrument, several other brief
tests have been used increasingly in clinical and research
settings to assess the presence and severity of cognitive
impairment. In two review articles of popular screening
instruments for assessing cognition in older adults
(McDougall, 1990; Yazdanfar, 1990), it was noted that the

Cognitive Capacity Screening Examination (CCSE), developed
by Jacobs, Bernhard, Delgado, and Strain (1977), offers the
brevity of the MMSE but also differs from other screening
instruments in that it includes five items which tap "ab-
stracting ability," a higher cortical function, of the
patient. The CCSE was developed specifically for use in
identifying organic brain syndromes and has been extensively
tested with geriatric patients (Jacobs et al., 1977;
McCartney & Palmateer, 1985; Omer, Foldes, Toby, & Menczel,
1983). Because this instrument requires only verbal re-
sponses from the patient, it can be given to patients with
motor, visual, or auditory impairments (in the last
instance, with written questions or written answers). Like
the MMSE, the highest possible score on the CCSE is 30, and
a score of less than 20 strongly suggests cognitive
dysfunction.

Several studies support the cutoff of 19 or less as
being suggestive of cortical impairment. In the original
validation studies, Jacobs et al. (1977) reported a signifi-
cant correlation between clinical diagnosis of organic
mental disorder and a CCSE score higher than 20 points in
medical patients evaluated by psychiatrists for organic
disorder. In addition, Kaufman, Weinberger, Strain, and
Jacobs (1979) administered the CCSE to 59 medical-surgical
patients admitted to a neurology service. Using the cutoff
score criterion of less than 20 for cognitive deficits, the

CCSE correctly identified, either true positive or true negative, 71% (42) when compared to results of clinical examinations.

Omer et al. (1983) gave the CCSE to 65 hospitalized medical-surgical patients (average age 76.1 years), and to 60 non-hospitalized persons as a control group (average age 71.6 years). The mean score for the control group was 22.3 with only 11 persons scoring below the cut-off of 20, whereas the mean score for the hospitalized group was 14.5 with 48 patients scoring below 20. Of note was the fact that the subsamples of hospitalized stroke patients (N = 18) achieved a mean score of 13.8, the organic mental syndrome group (N = 15) a mean score of 8.5, and the hip fracture group (N = 28) a mean score of 13.5 on the CCSE. The authors concluded that their study supports the importance of utilizing a measure of cognitive functioning on medical wards.

Haddad and Coffman (1987) cross-validated the CCSE in a sample of 87 psychiatric-geriatric patients (mean age 74.6 years). The CCSE total scores were compared across patients independently classified into functional and organic groups. The functional group (N = 46) achieved a mean score of 16.87, whereas the organic group (N = 41) achieved a mean score of 7.15. The differences were highly statistically significant, but both groups scored well below the cutoff of 20 points.

McCartney and Palmateer (1985) used the CCSE to identify and document cognitive deficits in elderly patients at the time of admission to a university hospital, with the goal of improving quality of life since some cases of cognitive dysfunction are reversible. Of 182 elderly patients, 36% (65) had CCSE scores of less than 20, and 64% (117) had scores of 20 or higher. Mental status findings were compared with chart review for notes regarding altered mental status in the elderly patients. Of the 65 patients identified with the CCSE as possibly having cognitive impairment, physician notes indicated deficits in only 14 cases (23.1%). One patient was identified by his physician as having a recent memory problem, making the total 15 patients. The mean CCSE score for these 15 patients was 9.9, indicating severe cognitive deficit. The remaining 50 patients not commented on by their physicians had a mean CCSE score of 15.5. The authors concluded that the need for assessment of the presence or absence of any degree of cognitive deficit was supported.

Webster, Scott, Nunn, McNeer, and Varnell (1984) compared the utility of the CCSE and a geometric copying task in a sample of 43 patients with documented brain impairment (i.e., determined by EEGs and CT scans) and 19 normal subjects. Findings revealed an overall accuracy rate of 61% for the CCSE. Results also suggested that the CCSE was better able to predict left cerebral hemisphere deficits,

whereas the design copying task predicted right hemisphere deficits better.

In a comparative study described in detail earlier in this review, Foreman (1987) evaluated the reliability and validity of the CCSE, the MMSE, and the Short Portable Mental Status Questionnaire (SPMSQ) developed by Pfeiffer (1975) with medical-surgical patients 65 years of age or older. Results indicated that the CCSE had the highest level of internal consistency reliability measured at 0.969. The CCSE was also reported to be the most comprehensive of the three mental status questionnaires studied. Criterion-related validity was reported to be 0.87 (p < .001), with excellent sensitivity and specificity (i.e., the CCSE accurately classified all patients). As with the SPMSQ and the MMSE, the CCSE's convergent validity was significant (p < .001), with evidence of discriminant ability. In a recent study, however, the CCSE was found to yield a higher false-negative rate (53%) than the MMSE (43%) in a sample of neurosurgical patients with documented brain lesions (Schwamm et al., 1987).

In summary, results of research generally tend to support the ability of the CCSE to detect cognitive impairment in geriatric patients when clinical diagnosis has been employed as the criterion standard. No studies examining the effects of socioeconomic status and education were located in the current review. Thus, at present,

sociodemographic effects on CCSE performance are unknown.
Compared to the MMSE, research suggests that the CCSE is
more comprehensive. However, when documented brain pa-
thology was used to establish the presence or absence of
cognitive dysfunction, the CCSE evidenced a higher rate of
false-negative decisions than the MMSE. Although the inclu-
sion of items on the CCSE that measure abstract thinking
abilities would seemingly increase its sensitivity to mild
or focal cognitive deficits, additional validation studies
utilizing a more adequate criterion measure are needed to
investigate this notion.

Neurobehavioral Cognitive Status Examination

In addition to comparing the validity of the MMSE and
the CCSE, the previously cited investigation by Schwamm and
his colleagues (1987) also examined the sensitivity of a
relatively new test called the Neurobehavioral Cognitive
Status Examination (NCSE). Results yielded a far superior
false negative rate (7%) for the NCSE when documented brain
lesions were employed as the criterion measure.

In a recent study comparing the NCSE to the MMSE in a
geriatric inpatient population, the NCSE was found to be
more sensitive than the MMSE in detecting cognitive impair-
ment, but its specificity and positive predictive values
were lower (Fields, Fulop, Sachs, Strain, & Fillit, 1992).
The latter investigation employed a psychiatrist's

determination of the presence of cognitive impairment as the criterion standard. Lower specificity and positive predictive value for the NCSE were achieved because, in many cases, the NCSE identified the presence of cognitive impairment, but the psychiatrist's assessment of cognitive dysfunction was negative. A higher proportion of false positives, therefore, tended to decrease the NCSE's specificity and positive predictive value. Although the authors attributed this finding to the NCSE's tendency to "overdiagnose" cognitive impairment, it may well be that it was the psychiatrist's judgment that was a poor predictor of cognitive functioning. As has been reiterated many times in the present review, clinical diagnosis is an inadequate standard by which to establish the presence or absence of cognitive impairment.

The NCSE is based on recent trends within neuropsychology that emphasize the importance of assessing independent areas of cognitive functioning rather than quantifying intactness of cognitive function with a single, global score (Kiernan et al., 1987). The NCSE assesses level of consciousness, orientation, and attention, as well as constructions, memory, calculations, and reasoning. Points given for correct responses are summed within each cognitive ability area to provide independent scores as opposed to a single overall score. Scores below a

predetermined criterion are interpreted as reflecting impairment within that particular area of functioning.

Standardization of the NCSE was based on a sample of 60 volunteers (age 20 to 66 years) and a second group of 59 geriatric volunteers (age 70 to 92 years), both without history of medical or psychiatric conditions that might have affected performance. These normative data are expressed as ranges of normal performance on the NCSE test summary sheet and are adjusted for the slightly poorer performance in normal elderly persons on memory, constructions, and similarities (Kiernan et al., 1987). No test-retest reliability data have been published for the NCSE, and, so far, the NCSE has reportedly been used only to test patients with known neurologic dysfunction and a normal cohort group.

Conceptually, the NCSE offers some clear advantages over current screening instruments. First, the test is based on an abilities model of brain function that emphasizes independent assessment of five major areas of cognitive function. Thus, it avoids a simplistic conceptualization of "organicity" as being either present or absent. The NCSE does not combine the results of performance in different cognitive areas into one total score. Therefore, successful performances in several areas do not obscure deficits in others. By specifying a patient's cognitive strengths and weaknesses, the NCSE provides information

usually attainable only through lengthy and extensive neuro-psychological test batteries.

Second, the NCSE increases the likelihood of detecting mild deficits by using a graded series of test items within each cognitive domain. Patients who fail the challenging screen items are evaluated for potential dysfunction with a series of increasingly difficult test items. Thus, when subjects are identified as cognitively impaired in a particular area, the degree of impairment is quantified.

Third, the NCSE independently assesses more areas of cognitive function than any of the other brief cognitive screening instruments (Schmitt, Ranseen, & DeKosky, 1989) and, as a result, should be able to detect isolated cognitive deficits with greater frequency. The inclusion of tasks designed to tap judgment should also increase the NCSE's sensitivity to mild cognitive dysfunction. Judgment is a higher-level, "executive" cortical function which is frequently impaired in the early stages of several progressive dementias (Albert, 1991).

Although research has been favorable regarding the use of the NCSE and the CCSE with neurosurgical populations and with older adult and geriatric groups, validation of both of these instruments, as with the MMSE, has not employed rigorous criterion measures as the standard by which to judge cognitive functioning. As established earlier in this review, clinical judgment; laboratory techniques including

the EEG, CT scan, MRI; and traditional intellectual measures
such as the WAIS-R are inadequate measures by which to
establish the criterion-related validity of cognitive
screening instruments. Although certain subtests of the
WAIS-R and some individual neuropsychological tests such as
the Boston Naming Test or the Wechsler Memory Scale-Revised
(WMS-R) are sensitive to the effects of cerebral damage,
they are significantly less sensitive to brain dysfunction
than a comprehensive examination like the Halstead-Reitan
Neuropsychological Test Battery (Reitan & Wolfson, 1986).

It is clear that the NCSE represents a major departure
from previous instruments for testing cognitive function,
and it may prove to provide the "missing link" between
brief, yet relatively insensitive mental status instruments
and lengthier, yet highly sensitive instruments like the
Halstead-Reitan battery. Likewise, the CCSE includes items
which measure abstracting ability and thus may prove to have
greater utility in the identification of more subtle cogni-
tive impairment than the MMSE. As Nelson et al. (1986)
concluded at the end of their exhaustive review of bedside
cognitive screening instruments, studies of criterion-based
validity employing comprehensive neuropsychological testing
are badly needed to clarify the utility of these
instruments.

Use of the Halstead-Reitan Battery with Older Adults

There is general agreement that comprehensive neuropsychological testing represents the most rigorous diagnostic means available to determine cognitive functioning (Nelson et al., 1986). The Halstead-Reitan Neuropsychological Test Battery (HRNTB) is one of the most widely used instruments for the assessment of brain dysfunction. This set of clinical tests was originally developed and validated by Ward Halstead and was later extended and modified by Ralph Reitan. An initial validation of the test procedures comprising the Halstead battery was reported by Reitan (1955b) in a study presenting a cross-validation of Halstead's work. In this investigation of 50 pairs of subjects (brain-damaged and normals), Reitan found significant differences in performance between the two groups on all instruments. Comparable results were reported by Vega and Parsons (1967) in their cross-validation study. Studies using populations from different geographical areas (i.e., midwest, east, southwest, and Norway) have produced similar findings (Chapman & Wolff, 1959; Klove, 1974).

There is evidence that the Halstead-Reitan battery can predict to a high degree right- versus left-hemispheric involvement (Reitan, 1955a, 1966), focal diffuse or bilateral focal damage (Reitan, 1959), lobular localization (Reitan, 1966); static versus rapidly growing lesions (Fitzhugh, Fitzhugh, & Reitan, 1961); and the disease

process, including cerebrovascular disease, neoplasm,
trauma, or degenerative disease (Reitan, 1966). Support for
the statistical and clinical validity and utility of the
HRNTB has also been reported by other researchers in a
variety of geographical and neuropsychological settings
(Goldstein, Deysach, & Kleinknecht, 1973; Klonoff, Fibiger,
& Hutton, 1970; Matthews, Shaw, & Klove, 1966; Schreiber,
Goldman, Kleinman, Goldfader, & Snow, 1976).

Although there is general agreement that the HRNTB is a
highly reliable and valid measure of brain dysfunction, a
critical issue in the neuropsychological evaluation of older
individuals involves discriminating between a pathologic
deterioration in performance and benign decrements thought
to be associated with normal aging (Flicker, Ferris, Crook,
Bartus, & Reisberg, 1986). The concepts of fluid and
crystallized cognitive abilities are useful in examining
cognitive changes associated with the normal aging process.
Crystallized abilities refer to overlearned, stored
knowledge functions such as those required by the Vocabulary
and Information subtests of the WAIS-R. Such measures have
been shown to be relatively resistant to age effects and
acquired brain damage, and may even show a slight age-
related increase until late in life (Hochanadel & Kaplan,
1984; Kallman & May, 1989). Fluid abilities, on the other
hand, refer to adaptive, new learning, and problem-solving
capacities. Such abilities have consistently proven to be

highly sensitive to cerebral dysfunction as well as the effects of normal aging (Cullum et al., 1989).

The Category Test, Tactual Performance Test (TPT), and Trail Making Test-Part B represent tasks from the HRNTB that require more fluid types of abilities, although crystallized abilities are also utilized to a lesser extent. These measures have consistently shown the strongest (negative) relationships with age (Fromm-Auch & Yeudall, 1983; Heaton, Grant, & Matthews, 1986; Reitan & Wolfson, 1985). Interestingly, Heaton et al. (1986) found that age accounted for less than 10% of the variance on measures of simple motor speed (Finger Tapping), grip strength (Hand Dynamometer), and sensory abilities (Sensory-Perceptual Exam). In addition, the use of standard cutoff scores for the Aphasia Screening Test and Sensory-Perceptual Exam remain appropriate in normal elderly individuals, based on a study of a large sample of healthy older persons age 65 to 75 years (Ernst, 1988).

The critical question to be addressed in this review is whether the validity of the HRNTB differs in older versus younger patient groups. To determine the sensitivity of the HRNTB to brain damage in these groups, Cullum et al. (1989) compared the neuropsychological test performance of younger versus older groups of patients with documented cerebral lesions. There were 196 younger patients with a mean age of 31.5 (SD = 9.4) years and 130 older patients with a mean age

of 60.8 ($\underline{SD}$ = 7.7) years. The two groups were comparable with respect to education (means of 12.6 versus 12.9) and sex distribution (62% versus 71% males). Results revealed that the Halstead Average Impairment Rating (AIR) correctly classified 80% and 78% of the patients in the respective groups. This finding suggests virtually identical sensitivity rates of the overall HRNTB to structural cerebral abnormalities in the younger and older groups. In addition, Cullum et al. (1989) found that elderly patients were able to complete the HRNTB with only a modest increase in test time (20 to 30 minutes). On the basis of these findings, the authors (1989) concluded that "the use of the HRNTB in evaluating most older patients is both feasible and psychometrically justifiable" (p. 606).

Although it is clear from these results that the HRNTB demonstrates respectable validity and utility as a measure of cerebral dysfunction in older adults, it is important to mention that the known effects of normal aging on neuropsychological test performance are always taken into consideration when interpreting individual test results. Careful examination of the pattern of test performance and level of performance across measures can substantially reduce the likelihood of misclassifying a patient as brain impaired. In addition, because the HRNTB is a comprehensive measure of brain functioning, there is a reduced likelihood that various age-related performance decrements on certain

tasks will result in the <u>overall</u> classification of a normal elderly person in the brain-impaired range.

The General Neuropsychological Deficit Scale (G-NDS) derived from the HRNTB evaluates the subject's performance on 42 different variables. The G-NDS yields scores for four subcategories (i.e., level of performance, pathognomonic signs, patterns and relationships among test results, and right-left differences) as well as a total score. If a normal elderly person scored in the severely impaired range on the Category Test, TPT (total time, memory, and localization), and Trails B, using the standard norms, the total G-NDS score would still fall well within the normal range of performance if the other performance variables were unimpaired. Thus, even though older individuals tend to perform more poorly than their younger counterparts on tasks from the HRNTB that require more fluid cognitive abilities, these age effects generally do not invalidate the HRNTB as a sensitive indicator of global cognitive functioning in the elderly (Cullum et al., 1989).

Need for the Study

As the literature reviewed thus far indicates, a clear need exists for a reliable and valid screening instrument for detecting cognitive impairment. Failure to recognize cognitive impairment inevitably leads to inappropriate interventions or no interventions at all (Folstein & Rovner,

1986). Elderly patients, in particular, may be viewed as "uncooperative," "lazy," or "manipulative" by professional staff and caregivers if these behaviors are not understood in the context of cognitive dysfunction. Failure to detect cognitive limitations can also lead to serious consequences for the patient and family upon discharge. A cognitively impaired person may be a danger to himself or herself and/or others if left unsupervised, allowed to drive, or left to assume major financial management responsibilities. Likewise, a person without any real cognitive deficits may have unnecessary restrictions placed on him or her based on "impaired" scores on mental status testing. An appreciation for the extent of cognitive disability and a clear understanding of the functional correlates of such impairment are crucial for accurate diagnosis, case formulation and management, rehabilitation, and discharge planning.

As previously cited literature suggested, the Mini-Mental State Examination (MMSE) was developed in response to the need for a standardized cognitive mental status examination. The MMSE's brevity and ease of administration have contributed to its popularity, and it is currently the most widely used screening test for cognitive impairment. The original validation studies by Folstein et al. (1975) demonstrated the utility of the MMSE for differential diagnosis of clinically diagnosed subgroups of dementia, depression, and cognitively impaired depressives. The research reviewed

thus far, however, suggests that the MMSE may show diminished sensitivity with older and less well-educated persons (Anthony et al., 1982) and may yield an unacceptably high rate of false negative decisions (Faustman et al., 1990; Schwamm et al., 1987). Indeed, recent research suggests that the MMSE may fail where a true screening test would be most needed--in evaluating patients without obvious cognitive deterioration.

As the research reviewed thus far indicates, the CCSE and NCSE may serve as viable alternatives to the MMSE. Both instruments are relatively brief and are more comprehensive in scope than other screening examinations. The NCSE, in particular, represents a new approach to rapid cognitive assessment. As mentioned earlier in this review, the NCSE independently assesses multiple domains of cognitive functioning and thereby provides the clinician with a differentiated profile of the patient's cognitive status. Although research to date has been favorable regarding the utility of both the CCSE and NCSE, validation procedures, as with the MMSE, have relied primarily on correlations with informal descriptive diagnoses or with other inadequate criterion measures such as WAIS-R scores, CT scans, MRI techniques, and EEGs, rather than employing the most rigorous diagnostic means available.

As screening instruments, the CCSE and NCSE as well as the MMSE should predict performance relative to a more

thorough testing procedure such as the HRNTB. As has been established in this review, the HRNTB demonstrates extensive reliability and validity as a test of cognitive functioning; and, in spite of its lengthy administration time and susceptibility to certain age effects, this instrument represents one of the most rigorous criterion measures available (Cullum et al., 1989; Nelson et al., 1986). A fundamental requirement of any screening examination is that it have high sensitivity (i.e., a low rate of false-negative results). To determine whether the CCSE and NCSE are more sensitive instruments for the detection of cognitive impairment than the MMSE, comparison of these instruments with the results of comprehensive neuropsychological testing would serve as a useful addition to the validation literature. Given the crucial importance that decisions based on the results of mental status testing have for patients and their caregivers, the identification of a truly sensitive, yet brief, cognitive screening test would be invaluable. Likewise, the identification of a truly insensitive instrument may help to prevent misuse and misinterpretation of test results that ultimately lead to erroneous treatment decisions and inadequate patient care.

Research Question and Hypotheses

The present study established the criterion-based validity of the Mini-Mental State Examination (MMSE), the

Cognitive Capacity Screening Examination (CCSE), and the
Neurobehavioral Cognitive Status Examination (NCSE). Per-
formance on the Halstead-Reitan Neuropsychological Test
Battery (HRNTB) served as the standard for comparison. The
sensitivity, specificity, and predictive value of each in-
strument, as well as how well each screening test correlated
with the HRNTB, were investigated. In addition, an explora-
tory analysis investigated what combination of subtests from
each of the three screening instruments most accurately
predicted performance on the HRNTB. The following
hypotheses were postulated.

1) The NCSE will correlate with the HRNTB to a
   significantly greater degree than either the MMSE
   or the CCSE.

2) The NCSE will demonstrate significantly greater
   sensitivity, specificity, and predictive value than
   either the MMSE or the CCSE.

3) The CCSE will demonstrate significantly greater
   sensitivity, and positive and negative predictive
   value than the MMSE.

4) The CCSE and MMSE will yield similar specificity
   rates.

CHAPTER 2

METHOD

Subjects

Subjects for the present study were recruited from the geriatric inpatient and outpatient population of the Dallas Veterans Administration Medical Center (DVAMC) Nursing Home Care Unit (NHCU) referred for psychological evaluation due to suspected cognitive dysfunction. Patients admitted to the NHCU constitute an eclectic group of primarily elderly medical patients in whom progressive dementing conditions such as multi-infarct dementia and other neurological diagnoses (e.g., cerebrovascular accident, anoxia) are common. Outpatients of the NHCU are community-residing elderly veterans who are referred for evaluation, usually by family members who have noticed symptoms of mental deterioration (e.g., forgetfulness, confusion, mood and personality changes). Additional subjects were recruited from the Texas College of Osteopathic Medicine's (TCOM) Gerontology Assessment and Planning Program (GAP). The TCOM Psychiatry Clinic provides neuropsychological testing services to both community-residing elderly referred through the GAP and to inpatients.

There were 52 subjects in the present investigation, 35 males and 17 females. Ages ranged from 55 years to 86 years, with a mean age of 67.25 years. Fifty percent of the subjects were from 55 to 66 years of age. Forty-two percent of the subjects were from 68 to 79 years of age. Only eight percent of the sample were 80 years of age or older. The majority of subjects (86.5%) were Caucasian, while 11.5% were Afro-American and 1.9% were Hispanic. Fifty percent of the subjects reported being currently married, while 23.1% reported being divorced, 19.2% were widowed, 5.8% were single, and 1.9% reported being currently separated from a spouse.

Years of education within this sample ranged from 6 to 20, with 9.6% having completed the eighth grade or less, 15.4% having attended some high school, 30.8% having graduated from high school, 17.3% having earned some college credit, 13.5% having graduated from college, and 13.4% having earned a graduate degree. The average number of years of education was 13.

Of the 52 subjects evaluated, the vast majority (96.2%) were cognitively impaired based on their performance on the Halstead-Reitan Neuropsychological Test Battery (HRNTB). Of that percentage, the majority (59.5%) were moderately impaired, with 19% falling within the mildly impaired range and 17.1% falling within the severely impaired range.

Eighteen of the subjects (34.6%) had the presumptive diagnosis of Alzheimer's disease, and 17 (32.7%) had either suffered a single cerebrovascular accident (CVA) or multiple CVAs, in which case the diagnosis was multi-infarct dementia. Seven of the subjects (13.5%) were referred for evaluation due to closed head injuries. Three of the subjects (5.8%) were diagnosed with atypical dementias including Parkinson's disease, Korsakoff's syndrome, and frontal lobe dementia. Six of the subjects (11.5%) had suffered a variety of medical problems prompting the referral for testing. Of these, two were diagnosed as having seizure disorders, and one had suffered anoxia following a myocardial infarct. Another subject was recovering from brain surgery that removed a tumor from the right cerebral hemisphere. One subject had received maintenance electro-convulsive shock treatment (ECT) for several years as treatment for major depression and was referred due to complaints of memory loss. Another subject had rheumatoid arthritis with systemic involvement believed to be etiologically related to her neuropsychological deficits.

In the present sample, 65.4% of the subjects were outpatients, and 34.6% were inpatients. The majority of outpatients (74%) were tested through the TCOM clinic, and the majority (67%) of inpatients were tested at the DVAMC NHCU.

The Geriatric Depression Scale (GDS) was used in the current study for exclusionary purposes. In the present sample, 36.5% of the subjects scored within the normal range, with the remaining 63.4% scoring in the range suggestive of mild depression. No subject scored within the range indicative of moderate to severe depression. Also, no subject in the present sample met the DSM III-R criteria for major depression. Therefore, it was not necessary to exclude any of the subjects due to severe depressive symptomatology.

Because the intent of the present study was to compare the validity of three cognitive screening instruments in older adult and geriatric groups, younger patients (54 years and younger) were excluded from participation. Also, patients who were severely demented, as determined by the initial clinical interview conducted with all patients at the NHCU and TCOM, would have been too severely impaired to participate in testing and thus were excluded. Patients with severe visual and hearing decrements or severe receptive and expressive speech difficulties that would have interfered with their ability to accurately perceive and/or respond to test instructions or stimuli were also excluded. Information about potential sensory deficits was available from a routine physical examination conducted by a physician at both the NHCU and TCOM. In addition, subjects were not taking medications (e.g., sedatives, high dosage of

antipsychotic drugs) that would have significantly inter-
fered with their ability to remain alert and attentive
during testing procedures. Subjects also spoke English as
their first language.

Instruments

Geriatric Depression Scale (GDS)

The GDS is a brief paper-and-pencil inventory designed
to screen for depression in an elderly population. The GDS
has an acceptably high sensitivity and specificity for
detecting major depression in both inpatients and out-
patients (Koenig, Meador, Cohen, & Blazer, 1988). A
detailed explanation of the construction of the GDS and its
reliability and validity data is available in Brink,
Yesavage, Lum, Heersema, Adey, and Rose (1982). The GDS
yields a 0- to 30-point global score with scores of 0 to 10
points being considered normal, 11 to 20 points being sug-
gestive of mild depression, and 21 to 30 points being
indicative of moderate to severe depression. According to
La Rue, Yang, and Osato (1992), research findings over the
past decade have indicated that the majority of older de-
pressed patients do not show severe or generalized cognitive
problems when they become depressed. It is only among
severely depressed individuals, most of whom are inpatients,
that cognitive impairment is relatively common (McAllister,
1983; Rabins, 1983). In the present study, a score of 21

points or above on the GDS defined the range to be utilized for exclusionary purposes.

Mini-Mental State Examination (MMSE)

The MMSE is a brief (5-10 min.) screening test of cognitive functioning. It evaluates the following cognitive dimensions: 1) orientation, 2) registration, 3) attention and calculation, 4) recall, 5) language (object naming, repetition, comprehension, reading and writing), and 6) constructional abilities. The test yields a 0- to 30-point global score with lower scores being associated with diminished performance. Folstein et al. (1975) suggest that patients who score below 24 points are cognitively impaired. MMSE scores in the range of 18 to 23 points and 0 to 17 points are indicative of mild and moderate/severe degrees of impairment, respectively.

Cognitive Capacity Screening Exam (CCSE)

The CCSE (Jacobs et al., 1977) contains 30 items and requires 5 to 15 minutes to administer. The CCSE was developed as a sensitive instrument for detecting diffuse organic mental syndromes. It measures domains that other screening instruments do not measure (abstraction and language) and has been tested with geriatric patients (McCartney & Palmateer, 1985). Content areas assessed by the CCSE are 1) orientation, 2) digit span, 3) concentration, 4) serial sevens, 5) repetition, 6) verbal concept formation, and 7)

short-term verbal recall. Scoring is done on the basis of number of correct responses with 30 representing a perfect score. The authors suggest that scores of 20 and above indicate normal cognitive functioning, 10 to 19 points represents mild cognitive impairment, and 0 to 9 points is suggestive of moderate to severe cognitive impairment.

Neurobehavioral Cognitive Status Examination (NCSE)

The NCSE was designed to provide a brief, yet comprehensive, assessment of cognitive functioning (Kiernan et al., 1987). Test content for the NCSE includes standard domains such as orientation, attentional ability, and level of consciousness in addition to independent tests to evaluate functioning within five major ability areas. Language assessment comprises four major areas: comprehension, naming, fluency, and repetition. Abstract verbal reasoning is also assessed with similarities and judgment tasks. Constructional skills are evaluated with an analogue of the WAIS-R Block Design subtest. Calculation ability is assessed by arithmetic problems, and memory is assessed in both verbal and visual spheres for immediate and delayed recall skills. The NCSE uses a screen and metric approach. Items are arranged so that, if a subject passes the screening question, performance in that area is assumed to be intact, and the examiner moves on to evaluate other cognitive areas. If the screen is failed, the metric is

administered. The metric consists of a series of items of graded difficulty. Thus, the screen items permit a brief examination in areas of normal cognitive functioning, while the metric items provide a quantitative evaluation whenever the question of disability is raised by failure on the screen. In addition, subjects who fail the screen may still demonstrate abilities in the normal range through their performance on the metric.

The NCSE takes approximately 20 minutes to administer to cognitively impaired patients. It consists of 11 independent tests, each of which is scored separately. The developers of the NCSE suggest that patients who have scores that are lower than those in the average range on any test are impaired in that specific skill. Thus, the NCSE yields a cognitive profile which visually depicts which cognitive domains are intact and which fall within the impaired performance range.

In order to compute sensitivity and specificity, the NCSE was considered positive for cognitive impairment if the subject scored below the age-adjusted norm for any one of the 10 cognitive domains. To determine the total score, points earned within each of the 10 subscales were summed. A score of 82 represents perfect performance on the NCSE. Level of impairment on the NCSE is further classified as mild, moderate, or severe within each of the 10 cognitive domains, but a global level of impairment rating is not

provided. For exploratory purposes, a level of impairment rating for overall performance on the NCSE was derived by computing an average impairment rating utilizing the ratings of the individual impaired scales. A scale in the mildly impaired range received a score of 1, a scale in the moderately impaired range received a score of 2, and a scale in the severely impaired range received a score of 3. The severity of impairment "scores" were added and then divided by the total number of impaired cognitive scales. Scores below 2 represent mild cognitive impairment, 2 to 2.99 represent moderate cognitive impairment, and a score of 3 represents severe cognitive impairment.

Halstead-Reitan Neuropsychological Test Battery (HRNTB)

The HRNTB consists of a number of individual tests which, when administered together, comprise a battery. The HRNTB has demonstrated extensive reliability and validity as a test of cortical function in psychiatric, neurologic, and geriatric populations (Cullum et al., 1989; Robbins, 1989). Critics have argued, however, that the HRNTB does not thoroughly assess memory and that it provides a poor measure of parietal lobe functioning (Franzen, 1989). Despite these shortcomings, the HRNTB was chosen as the criterion measure in the present study because it represents a widely used neuropsychological battery with proven utility in the documentation of brain impairment and because it is the standard

neuropsychological test battery regularly used at both data collection sites. In addition, the primary investigator was trained in the administration and interpretation of the HRNTB, but has limited familiarity with other neuropsychological assessment approaches. Thus, although an instrument such as the Luria-Nebraska Neuropsychological Battery could have served equally well as the criterion standard, its use was not feasible in the present investigation.

The HRNTB assesses the following areas associated with brain function: 1) complex problem-solving, 2) attention and concentration, 3) learning and delayed recall, 4) concept formation, 5) perseveration, 6) perceptual motor speed, 7) visuo-spatial abilities, 8) sequencing efficiency, 9) sensory-perceptual functions, 10) expressive and receptive language, and 11) motor proficiency. Administration of the HRNTB for Adults (age 15 and older) usually takes between four and six hours, with individual tests ranging from 10 minutes to one hour or more. The HRNTB involves interactive testing between the examiner and subject. The battery used in the present study was comprised of the following tests.

Sensory-Perceptual Exam (SPE). The SPE measures tactile, auditory, and visual perceptual abilities. The procedures require the subject to perceive unilaterally presented stimuli on each side of the body. Next, stimuli are represented in a bilateral, simultaneous manner to determine the subject's ability to perceive both stimuli.

Tactile, visual, and auditory sensitivity are measured, as well as left-right comparisons.

Reitan-Indiana Aphasia Screening Test (RIAST). The RIAST provides a measure of several aspects of language ability and usage, including the ability to name common objects, spell, read, identify numbers and letters, write, do arithmetic calculations, articulate, identify body parts, perform pretended movements, understand the meaning of spoken language, follow directions, and differentiate left from right. It also tests visual constructional abilities and provides samples of the subject's attempts at reproducing the spatial configuration of several forms.

Category Test. The Category Test is a measure of abstracting ability. Stimulus figures, which vary in size, location, shape, number, color, and intensity, and are grouped by abstract principles, are projected on a screen. The task of the subject is to figure out the principle relating stimulus subtests and signal the answer by pressing the appropriate key on a board. The test is an excellent discriminator between brain-damaged and neurologically intact groups. The test measures the ability to sustain attention, remember past performance, evaluate past performance and learn from feedback, concentrate, analyze visually presented material, understand spatial relationships, and demonstrate cognitive flexibility in handling a complex, changing problem.

Speech Sounds Perception Test. The Speech Sounds Test consists of 60 spoken nonsense syllables, the beginning and ending consonant sounds of which vary while their "ee" vowel sound remains constant. The subject must underline the spoken syllable, selecting from four alternatives printed on the test form. The test requires attention, auditory perception of verbal material, and the ability to match phonemes with their written equivalents (graphemes).

Seashore Rhythm Test. The Rhythm Test is a subtest of the Seashore Test of Musical Talent. The subject must discriminate between 30 pairs of rhythmic beats which are sometimes the same and sometimes different. The task measures alertness to nonverbal, auditory stimuli; sustained attention to the task; and the ability to perceive and compare different rhythmic sequences. Because of the attentional component, the Rhythm Test is not believed to be an effective discriminator between functional psychiatric disorders and organicity.

Tactual Performance Test (TPT). The TPT uses a modification of the Sequin-Goddard form board. The subject is blindfolded and is not permitted to see the stimulus material. The task is to fit blocks into the proper spaces on the board using first the dominant hand, then the non-dominant hand, then both hands. Times are recorded for each trial and also for the total time required for all three trials. The stimulus materials are put away, and the

blindfold is removed. The subject is then asked to draw a diagram of the board representing the blocks in their proper spaces. The drawing is scored based upon the number of blocks correctly reproduced (Memory score) and the number of blocks correctly placed (Localization score). The test requires the ability to recognize forms tactually, perform coordinated movements of the hands and arms, perceive one's own movement in space without visual cues, and plan and solve a nonverbal, kinesthetic problem. Performance of the right and left sides of the body is evaluated on the TPT. The Memory and Localization components require both spatial and incidental memory. The Localization score, in particular, is very sensitive to organicity.

Finger Oscillation Test. Finger tapping is a measure of fine motor speed which uses a mounted tapper equipped with a counter. The subject is administered several consecutive trials with each hand until attaining a criterion of five trials within a five-point range of each other. The score for the dominant hand and the nondominant hand is the average of the five trials. Fine motor speed and coordination and left-right differences are assessed.

Grip Strength Test. A plunger-type dynamometer with a grip adjustable for hand size is used to measure strength. Trials alternating between the dominant and non-dominant hand are administered, with the score being the average between two trials for each hand. The test measures grip

strength as well as providing a comparison of the right and left sides of the body.

Trail Making Test. This test consists of two parts, Trails A and Trails B. Trails A consists of 25 circles distributed randomly over a sheet of paper and numbered from 1 to 25. The subject is required to connect the circles with a pencil line in ascending numerical order. Part B also consists of 25 circles which are numbered from 1 to 13 and lettered from A to L. The task is to connect the circles in sequence, alternating between numbers and letters. The scores obtained are the time taken to complete each task and the number of errors. The test measures visual-motor tracking skills, counting ability, spatial skills, sequential and planning ability, cognitive flexibility in shifting between letter and number sets, and the ability to handle verbal material. The Trail Making Test, Trails B in particular, is a sensitive indicator of brain dysfunction.

Although interpretation of individual test performance on the HRNTB is a complex process, for the purpose of the present study, one global performance score, the General Neuropsychological Deficit Scale (G-NDS), was computed and served as the criterion standard. The G-NDS differentiates strikingly between control (normals) and brain-damaged subjects and serves as an excellent general indicator of neuropsychological impairment without being unduly influenced by diffuse, left, or right cerebral damage (Reitan

& Wolfson, 1988). The G-NDS score represents the subject's performance on 42 variables derived from the HRNTB. These variables are divided into four groups to reflect deficits in accordance with each of the methods of neuropsychological inference: 1) Level of Performance (variables 1-19), 2) Pathognomonic Signs (variables 20-31), 3) Patterns and Relationships among Test Results (variables 32 and 33), and 4) Right-Left Differences (variables 34-42). Of the 42 variables, computation of four (variables 1, 2, 32 and 33) requires Full Scale, Verbal and Performance IQ scores from the Wechsler Adult Intelligence Scale-Revised (WAIS-R). Because it was not feasible to administer the WAIS-R in the present investigation, it was assumed that subjects had at least average intelligence and that there was not a significant difference between Verbal IQ and Performance IQ. These assumptions served not to penalize subjects for performance decrements that might have been picked up on the WAIS-R.

A score is obtained for each of the G-NDS categories as well as a total score. A score of 0 represents perfect performance, with scores ranging from 0 to 116. Persons with scores of 25 or less are classified by the G-NDS as normal in terms of the adequacy of their neuropsychological functioning. A G-NDS score of 26 to 40 represents mild impairment, 41 to 67 represents moderate impairment; and a G-NDS score of 68 or more points represents severe

impairment. A copy of the rules for calculating the G-NDS score appears in Appendix A.

For exploratory analyses, two other performance indices were computed based on performance on the HRNTB. The first is called the Impairment Index (II) and is a measure of the consistency of performance decrements on the HRNTB. The II compares performance on seven subtests (Category; TPT--Total Time, Memory, and Localization; Seashore Rhythm; Speech Sounds; and Finger Oscillation). It is calculated by counting the number of impaired performance scores and dividing by the total number of subtest scores. Thus, scores range from 0 to 1.0, with 1.0 representing impaired performance on all tests which comprise the II. A cutoff score of .40 or higher is used to identify persons with brain damage.

The second score is called the Average Impairment Rating (AIR) and is also a measure of global neuropsychological functioning developed by Russell, Neuringer, and Goldstein (1970). The AIR rates performance on 11 tests from the Halstead battery (Category; TPT--Total Time, Memory, and Localization; Speech Sounds; Seashore Rhythm; Finger Oscillation; Trails B; Reitan-Indiana Aphasia Screen; Spatial Relations; and the Sensory Perceptual Exam) and one subtest from the WAIS-R (Digit Symbol) as better than average (scored 0), normal (scored 1), mildly impaired (scored 2), moderately impaired (scored 3), moderately severe (4), and severe (scored 5), based on revised norms for rating

equivalents of the raw scores. These performance ratings are then added and divided by 12 to derive the AIR score. Scores of 0.00 to 1.35 are considered normal, 1.36-2.00 suggests mild impairment, 2.01-2.85 suggests moderate impairment, 2.86-3.50 suggests moderately severe impairment, and 3.51-5.00 suggests severe impairment. In the present study, only the 11 tests from the Halstead battery were used to compute the AIR because subjects were not administered the WAIS-R. A copy of the rating equivalents for 11 tests from the Halstead-Reitan Battery appears in Appendix B.

Procedure

Each subject was asked to complete a brief questionnaire designed to gather basic demographic data that were utilized later to describe the research sample. Each subject at the DVAMC was individually tested by the primary investigator, and each subject at TCOM was tested by Dr. Andrew Houtz, who routinely conducts neuropsychological evaluations. Testing was divided into two sessions in order to reduce fatigue effects. During the first session, each subject completed the demographic questionnaire, the GDS, the MMSE, the CCSE, and the NCSE. A previous investigation comparing the MMSE, the CCSE, and the NCSE found no significant order effects for mean scores of these three instruments (Schwamm et al., 1987). Examination scores, therefore, were unaffected by the order of test

administration of these instruments. In the present study, the MMSE was administered first, followed by the CCSE and the NCSE. To promote patient cooperation, minimize learning effects, and avoid needless repetition, items common to all three screening instruments were combined. The first section of each instrument consists of shared orientation questions. The CCSE and MMSE overlap on the serial sevens subtraction task, so this task was administered only in the MMSE section, where it serves as a distraction between registration and recall on the memory task.

After a 15-minute break, three tests of the HRNTB were also administered in the following order during the first testing session: 1) Sensory-Perceptual Exam, 2) Reitan-Indiana Aphasia Screening Test, and 3) Category Test. During the second testing session, the remaining tests of the HRNTB were administered in the following order: 4) Speech Sounds Perception Test, 5) Seashore Rhythm Test, 6) Tactual Performance Test (TPT), 7) Finger Oscillation Test, 8) Grip Strength Test, and 9) Trail Making Test. Another 15-minute break was given after the TPT. The order of administration of the various tests which comprise the HRNTB was chosen in order to reduce fatigue effects. After consultation with Dr. Michael Eppinger, a neuropsychologist at the DVAMC who frequently assesses geriatric patients, it was determined that this ordering of the tests intersperses the more difficult and lengthy tasks with easier and less

time-consuming ones. This served to substantially increase the likelihood that subjects remained attentive and motivated to do their best.

A summary of the instruments administered and the order in which the tests were given is provided below.

First Session:
    1) Geriatric Depression Scale
    2) Mini-Mental State Examination
    3) Cognitive Capacity Screening Examination
    4) Neurobehavioral Cognitive Status Examination
    5) Halstead-Reitan Battery:
        a) Sensory-Perceptual Exam
        b) Reitan-Indiana Aphasia Screening Test
        c) Category Test

Second Session:
    6) Halstead-Reitan Battery continued:
        d) Speech Sounds Perception Test
        e) Seashore Rhythm Test
        f) Tactile Performance Test
        g) Finger Oscillation Test
        h) Grip Strength Test
        i) Trail Making Test

Statistical Analysis

In order to investigate the relationship between scores obtained on the three cognitive screening instruments and the criterion measure, Pearson correlation coefficients were computed between the MMSE, CCSE, NCSE, and the G-NDS from the Halstead-Reitan Battery. To test for significant differences between the correlation coefficients, t-tests for a dependent sample were computed. To explore the intercorrelations among the three cognitive screening instruments, additional correlation coefficients were computed between the MMSE, CCSE, and NCSE. Correlation coefficients were

also computed between the various cognitive dimensions of the MMSE, CCSE, and NCSE and the G-NDS.

Other exploratory analyses were conducted to examine the correlation between the G-NDS and the Impairment Index (II) from the Halstead-Reitan Battery and between the G-NDS and the Average Impairment Rating (AIR). The MMSE, CCSE, and NCSE were further tested for sensitivity, specificity, and predictive value by computing the true positive, false negative, false positive, and true negative rates and utilizing formulas adapted from Larson's (1986) work on evaluating the validity of screening tests. Differences in sensitivity, specificity, and predictive value for the three screening instruments were examined with a series of tests for significance of difference between two proportions. In addition, a stepwise multiple regression equation was performed, utilizing subtests from the MMSE, CCSE, and NCSE as predictor variables, in order to explore which combination of subtests "explained" the largest proportion of variance in the dependent measure. The latter analysis was done in an attempt to identify which subcomponents of each cognitive screening test may be the most clinically useful to identify cognitive impairment.

A series of one-tailed univariate analyses of variance (ANOVAs) was performed in order to investigate the effects of age, education, and sex on performance on the HRNTB, as well as the three cognitive screening instruments. A series

of one-tailed ANOVAs was also used to explore the effect of inpatient versus outpatient status on performance on the Halstead, MMSE, CCSE, and NCSE. Two additional ANOVAs were computed to explore differences in performance related to subject population (i.e., TCOM versus VA) and differences in performance related to diagnostic group. A final exploratory analysis presented level of impairment ratings from the MMSE, CCSE, NCSE, and the G-NDS in the form of contingency tables in order to investigate whether the three screening tests could accurately predict level of cognitive impairment as established by the G-NDS.

CHAPTER 3

RESULTS

The first hypothesis predicted that the NCSE total
score would correlate with the G-NDS from the HRNTB to a
significantly greater degree than either the MMSE total
score or the CCSE total score. To investigate this hypothe-
sis, Pearson correlation coefficients were computed between
the MMSE, CCSE, and NCSE and the G-NDS. Figures 1, 2, and 3
(Appendix C) display the relationship between the MMSE,
CCSE, NCSE, and the criterion measure, the G-NDS.

The correlation between the NCSE and the G-NDS was
-0.67. This correlation is statistically significant ($p <$
.01, two-tailed), with a moderate portion (45%) of the vari-
ance explained. The correlation between the MMSE and G-NDS
was -0.60, which is also significant ($p <$ .01, two-tailed),
with a moderate portion (36%) of the variance explained.
The correlation obtained between the CCSE and the G-NDS was
-0.62. This correlation is also significant ($p <$ .01, two-
tailed) and accounts for 38% of the variance. To test for
significant differences among the correlation coefficients,
t-tests were performed. Although the NCSE yielded a
slightly higher correlation with the G-NDS than did the MMSE

or CCSE, the differences were statistically nonsignificant. Therefore, results do not support the first hypothesis.

Additional correlation coefficients were computed to explore the intercorrelations between the three cognitive screening tests, as well as between the G-NDS and the Impairment Index (II) and between the G-NDS and the Average Impairment Rating (AIR). The correlation between the MMSE and the CCSE was 0.89 with 79% shared variance. The correlation between the NCSE and MMSE was 0.75 with 56% shared variance. The correlation between the NCSE and the CCSE was 0.77 with 59% shared variance. The G-NDS and AIR were highly correlated (r = 0.93), and the G-NDS and the II were moderately correlated (r = 0.71). All of these correlations were statistically significant. A correlation matrix listing coefficients computed between the MMSE, CCSE, NCSE, and G-NDS, as well as coefficients calculated between these measures and the II and AIR is presented in Table 1 (Appendix C).

Correlation coefficients were also calculated between each cognitive dimension assessed by the three cognitive screening tests and the G-NDS. On the MMSE, 6 out of 10 cognitive dimensions were significantly correlated with the G-NDS. However, correlation coefficients could not be computed between three of the domains (Registration, Naming, and Repetition) because all subjects correctly answered the items comprising these content areas. A negligible and

nonsignificant correlation (r = -0.03) was obtained between the MMSE 3-Stage Command task and the G-NDS. Correlation coefficients calculated between the MMSE cognitive dimensions and the G-NDS are presented in Table 2 (Appendix C).

On the CCSE, the Serial-Sevens task, followed by the Orientation items and the Abstract Reasoning items achieved the highest correlations with the G-NDS (r = -0.50, -0.463, and -0.461, respectively). All cognitive domains on the CCSE were significantly correlated with the G-NDS. Table 3 (Appendix C) presents the correlation coefficients calculated between the CCSE content areas and the G-NDS.

On the NCSE, only the Repetition subscale was not significantly correlated with the G-NDS. The highest correlations were between the Constructions subscale and the G-NDS (r = -0.69) and between the Memory subscale and the G-NDS (r = -0.47). Table 4 (Appendix C) presents the correlation coefficients calculated between the NCSE cognitive domains and the G-NDS.

In order to investigate what combination of cognitive dimensions from the three screening tests "explained" the largest proportion of variance in the dependent measure, an exploratory stepwise multiple regression analysis was performed. The potential independent or "predictor" variables included 7 cognitive dimensions measured by the MMSE (Orientation, Attention and Calculation, Recall, 3-Stage Command, Reading, Writing, and Copy Design), the 6 cognitive

dimensions measured by the CCSE (Orientation, Attention, Calculations, Abstract Reasoning, Memory, and Serial Sevens), and the 10 cognitive domains assessed by the NCSE (Orientation, Attention, Comprehension, Repetition, Naming, Constructions, Memory, Calculations, Similarities, and Judgment). Three MMSE cognitive domains (Registration, Naming, and Repetition) were deleted from the analysis because correlation coefficients could not be computed due to the lack of variability among scores. The dependent variable was G-NDS total score.

In the first step, the independent variable entered into the equation was the one with the highest Pearson correlation with the G-NDS, the Constructions subscale of the NCSE ($r = 0.687$, $F (1,50) = 44.68$, $p < .0000$). The Adjusted r-square for the NCSE Constructions subscale was 0.46, indicating that 46% of the variance was explained by this variable. In the second step, the NCSE Memory subscale was entered, with a resulting significant increase in the Adjusted r-square from 0.46 to 0.54 ($F (2,49) = 31.62$, $p < .0000$), indicating that 54% of the variance in the G-NDS was explained by these two cognitive domains. In the third and final step, the Attention and Calculation dimension of the MMSE was entered, with a resulting significant increase in the adjusted r-square from 0.54 to 0.58 ($F (3,48) = 24.61$, $p < .0000$), indicating that, with the addition of the Attention and Calculation dimension of the MMSE, the proportion

of "explained" variance increased to 58%. The addition of the remaining cognitive dimensions from the MMSE, CCSE, and NCSE resulted in no significant increase in r-square.

The second hypothesis predicted that the NCSE would demonstrate significantly greater sensitivity, specificity, and positive and negative predictive value than either the MMSE or the CCSE. To investigate this hypothesis, the frequency of False Positives (FP), True Positives (TP), True Negatives (TN), and False Negatives (FN) for each cognitive screening test was computed by comparing all subjects' scores on the screening test to their performance on the criterion measure, the G-NDS from the Halstead battery. Table 5 (Appendix C) lists the age, gender, years of education, medical diagnosis prompting referral, and total score on the MMSE, CCSE, NCSE, G-NDS, II, and AIR for each subject. Figures 4, 5, and 6 (Appendix C) show the frequency of FP, TP, TN, and FN cases for the MMSE, the CCSE, and the NCSE. The formulas for calculating sensitivity, specificity, and predictive values for a positive or negative test are shown in Appendix D.

Of the 52 subjects evaluated in the present study, the G-NDS identified 50 as cognitively impaired and 2 as normal. G-NDS total scores ranged from 18 to 90 points. The mean G-NDS total score was 52.83 ($\underline{SD}$ = 16.48). MMSE total scores ranged from 15 to 30 points ($\underline{M}$ = 25.67, $\underline{SD}$ = 3.72), with 42 subjects scoring in the normal range (24-30 points) and 10

subjects scoring in the cognitively impaired range (23 points or less). Of the 50 cases that were identified by the G-NDS as cognitively impaired, the MMSE detected only 10. The MMSE did accurately identify the two normal cases as unimpaired, however, and did not identify any cases as impaired that were normal according to the G-NDS (i.e., False Positives). Therefore, results indicate that the MMSE achieved 20% sensitivity, 100% specificity, 100% positive predictive value, and 5% negative predictive value. The False Positive Rate was 0, and the False Negative Rate was 80%.

CCSE total scores ranged from 6 to 30 points ($\underline{M}$ = 23.15, $\underline{SD}$ = 5.33), with 41 subjects scoring in the normal range (20-30 points), and 11 subjects scoring in the cognitively impaired range (19 points or less). Of the 50 cases that were identified by the G-NDS as cognitively impaired, the CCSE accurately identified only 11. Like the MMSE, the CCSE did accurately classify the two normal cases as unimpaired, and did not make any False Positive errors. Thus, results indicate that the CCSE demonstrated 22% sensitivity, 100% specificity, 100% positive predictive value, and 5% negative predictive value. The False Positive Rate was 0, and the False Negative Rate was 78%.

The total number of impaired scales on the NCSE ranged from 0 to 8 ($\underline{M}$ = 2.25, $\underline{SD}$ = 2.08), with 11 subjects scoring in the normal range (0 impaired scales) and 41 subjects

scoring in the cognitively impaired range (1 or more im-
paired cognitive scales). The NCSE accurately identified 43
of 52 subjects as being either normal ($\underline{N}$ = 2) or cognitively
impaired ($\underline{N}$ = 41). Nine subjects were classified as normal
on the NCSE while demonstrating cognitive impairment on the
criterion measure (False Negatives). Therefore, results
indicate that the NCSE evidenced 82% sensitivity, 100% spe-
cificity, 100% positive predictive value, and 18% negative
predictive value. The False Positive Rate for the NCSE was
0, and the False Negative Rate was 18%.

The NCSE, MMSE, and CCSE yielded identical rates of
specificity and positive predictive value (all 100%). The
MMSE and CCSE also demonstrated identical negative predic-
tive values (5%). To test for significant differences in
sensitivity between the NCSE and the MMSE and between the
NCSE and the CCSE and to test for significant differences in
negative predictive value between the NCSE and the value
obtained by the other two screening tests, a series of tests
for significance of the difference between two proportions
was performed. Results indicated a significant difference
between the sensitivity of the NCSE (82%) and the sensi-
tivity of the MMSE (20%), with $\underline{z}$ = 6.89 ($\underline{p}$ < .01, two-
tailed). Results also indicated a significant difference
between the sensitivity of the NCSE and the CCSE (22%), with
$\underline{z}$ = 6.00 ($\underline{p}$ < .01, two-tailed). The difference between the
negative predictive value of the NCSE (18%) and the negative

predictive value of the MMSE and CCSE (both 5%) was non-significant. Therefore, the results partially support the second hypothesis.

The third hypothesis predicted that the CCSE would demonstrate significantly greater sensitivity and positive and negative predictive value than the MMSE. As the prior analysis showed, the MMSE and CCSE yielded virtually identical positive (100%) and negative predictive values (5%). A test for the significance of the difference between two proportions was performed to examine differences in sensitivity between the MMSE and CCSE (20% versus 22%, respectively). Results indicated that the differences were nonsignificant. Thus, the third hypothesis was not supported.

The fourth hypothesis predicted that the CCSE and MMSE would yield similar specificity rates. Both screening instruments demonstrated 100% specificity. Therefore, results support the fourth hypothesis.

In order to investigate whether adjusting the cutoff scores on the MMSE and CCSE would substantially improve their sensitivity, a series of exploratory analyses was performed whereby the cutoff score was raised for each instrument. On the MMSE, the cutoff score was systematically raised to 25, 26, and 27 points. When the cutoff was defined as 26 points or less to be suggestive of cognitive impairment, the sensitivity of the MMSE was improved from

20% to 54%, with a resulting lower False Negative Rate of
46%. On the CCSE, the cutoff score was systematically
raised to 20, 21, 22, 23, 24, and 25 points. When the cut-
off was defined as 24 points or less to be suggestive of
cognitive dysfunction, the sensitivity of the CCSE improved
from 22% to 56%, with a resulting lower False Negative Rate
of 44%. Table 6 (Appendix C) presents a comparison of the
sensitivity and specificity for the MMSE and CCSE using
standard and adjusted cutoff scores.

In order to investigate whether the MMSE, CCSE, and
NCSE could predict level of cognitive impairment, as defined
by the G-NDS, contingency tables were created that compared
the ratings from the G-NDS to each cognitive screening test.
As can be seen from Table 7 (Appendix C), the MMSE clas-
sified 100% of the mildly impaired cases as normal. The
MMSE classified 84% of the moderately impaired cases as
normal, and the remaining 16% as mildly impaired. In addi-
tion, the MMSE classified 44% of the severely impaired cases
as normal, 22% as mildly impaired, and 33% as moderately to
severely impaired. Out of a total of 52 cases, the MMSE
rating of level of impairment was discrepant from the cri-
terion standard in 47 cases (90%).

Like the MMSE, the CCSE classified 100% of the mildly
impaired cases as normal. The CCSE classified 81% of the
moderately impaired cases as normal and the remaining 19% as
mildly impaired. Of the severely impaired cases, 44% were

classified as normal by the CCSE, 44% were classified as mildly impaired, and the remaining 11% were classified as moderate to severely impaired. Out of 52 cases, the CCSE rating of level of cognitive impairment was discrepant from the criterion standard in 49 cases (94%). Table 8 (Appendix C) presents the comparisons between CCSE level of impairment ratings and the G-NDS.

When level of impairment ratings between the G-NDS and the NCSE were compared, the NCSE was found to classify 50% of the mildly impaired cases as normal, 20% as moderately impaired, and the remaining 30%--in agreement with the G-NDS--as mildly impaired. Of the moderately impaired cases, 48% were also classified as moderately impaired by the NCSE, with 13% being classified as normal, 26% as mildly impaired, and 13% as severely impaired. Of the cases rated as severely impaired by the G-NDS, the NCSE also classified 33% as severely impaired, with 22% being classified as mildly impaired and the remaining 44% as moderately impaired. Out of 52 cases, the NCSE rating of level of impairment was discrepant from the criterion standard in 29 cases (56%). Table 9 (Appendix C) presents the comparisons between the NCSE level of impairment ratings and the G-NDS.

Additional exploratory analyses investigated each screening test's ability to predict cognitive impairment compared to the Impairment Index (II) from the Halstead battery. Of 48 subjects rated as impaired by the II (II >

0.4), the MMSE classified 38 (79%) as normal, the CCSE classified 37 (77%) as normal, and the NCSE classified 7 (15%) as normal. Thus, the NCSE identified 41 (85%) of the 48 subjects classified as brain impaired by the II, whereas the MMSE identified only 10 (21%) and the CCSE identified only 11 (23%).

Each screening test's ability to predict cognitive impairment based on Average Impairment Rating (AIR) scores was also explored. Of 49 cases identified as cognitively impaired by the AIR (total score $\geq$ 1.36), the MMSE classified 39 (80%) as normal, the CCSE classified 38 (78%) as normal, and the NCSE classified 8 (16%) as normal. Thus, the NCSE identified 41 (84%) of the 49 subjects classified as impaired by the AIR. The MMSE, however, identified only 10 (20%) and the CCSE identified only 11 (22%) of the 49 impaired subjects. Tables 10, 11, and 12 (Appendix C) present the comparisons between the AIR and each of the cognitive screening tests' rating of level of impairment.

Although the G-NDS was used as the criterion standard in the present study, exploratory comparisons investigated to what extent the G-NDS, the II, and the AIR agreed in their identification of cognitive impairment. When a comparison of the G-NDS and II was made, it was found that the G-NDS and II disagreed on only two cases. The G-NDS identified two subjects as mildly impaired, whereas their scores on the II fell within the normal range. When the G-NDS was

compared to the AIR, it was found that the G-NDS and AIR disagreed on only one case. The G-NDS identified one subject as mildly impaired, whereas that subject's score on the AIR was normal. Likewise, when the AIR was compared to the II, they disagreed on only one case. The AIR identified one subject as mildly impaired, whereas that subject's score on the II was within the normal range.

A series of exploratory one-tailed, univariate analyses of variance (ANOVAs) was performed to investigate the effects of age, education, and sex on performance on the MMSE, CCSE, and NCSE, as well as on the criterion measure, the G-NDS. In the first analysis, two age groups were created: young-old, ages 55-65 years (N = 24) and old-old, ages 66-86 years (N = 28). Differences between the young-old and old-old groups on the G-NDS, MMSE, CCSE, and NCSE mean total scores were nonsignificant.

In the second analysis, two educational groups were created: a low group with 6-12 years of education (N = 29) and a high group with 13-20 years of education (N = 23). Results were significant for education, with the G-NDS total score (F (1,50) = 7.104, p < .01), indicating that the low educational group had a higher mean G-NDS total score (M = 57.90) than did the high educational group (M = 46.30). Results were also significant for education with the MMSE total score (F (1,50) = 4.339, p < .04), indicating that the low educational group scored lower (M = 24.79) on the MMSE

than did the high educational group ($\underline{M}$ = 26.87). For the CCSE, results were significant ($\underline{F}$ (1,50) = 6.855, $\underline{p}$ < .01), indicating that the low educational group also scored lower on the CCSE ($\underline{M}$ = 21.45) than did the high educational group ($\underline{M}$ = 25.22). Results were also significant for education with the NCSE total score ($\underline{F}$ (1,50) = 4.183, $\underline{p}$ < .04), suggesting that the low educational group also scored lower ($\underline{M}$ = 62.24) than did the high educational group ($\underline{M}$ = 68.09). Therefore, results indicate that the low educational group performed significantly worse on all three cognitive screening tests, as well as on the criterion measure, than did the high educational group.

In the third analysis, no significant differences were found between males and females on the G-NDS, MMSE, CCSE, or NCSE mean total scores.

Additional one-tailed ANOVAs were performed to investigate the effect of inpatient ($\underline{N}$ = 18) versus outpatient ($\underline{N}$ = 34) status on performance on the Halstead battery, as well as on each cognitive screening test. In the first ANOVA, the dependent variable (DV) was G-NDS total score. Results were significant for inpatient versus outpatient status with the G-NDS ($\underline{F}$ (1,50) = 7.654, $\underline{p}$ < .008), indicating that the inpatient group scored significantly worse ($\underline{M}$ = 60.94) on the G-NDS did than the outpatient group ($\underline{M}$ = 48.44). Subsequent ANOVAs revealed nonsignificant differences between

inpatient and outpatient groups on MMSE, CCSE, and NCSE mean total scores.

In order to investigate differences between the TCOM subject population (N = 31) and the DVAMC subject population (N = 21), a series of one-tailed ANOVAs was performed, treating age, education, performance on the Halstead battery (G-NDS mean total scores), and performance on each cognitive screening test (MMSE, CCSE, and NCSE mean total scores) as dependent variables. Results were significant for age by subpopulation status (F (1,50) = 8.688, p < .005), indicating that the TCOM subpopulation group was significantly older (M = 69.77 years) than the DVAMC subpopulation group (M = 63.52 years). Results were also significant for performance on the Halstead battery by subpopulation status (F (1,50) = 4.352, p < .04), indicating that the TCOM subject group performed better (G-NDS mean total score = 48.97) on the Halstead battery than did the DVAMC subject group (G-NDS mean total score = 58.38). Differences between the groups on education, MMSE, CCSE, and NCSE mean total scores were all nonsignificant.

A final series of exploratory analyses investigated the effects of diagnostic status on performance on the criterion standard, the G-NDS, as well as on MMSE, CCSE, and NCSE performance. Due to small and unequal numbers in some of the diagnostic groups, only the two major diagnostic categories, Alzheimer's disease (N = 18) and multi-infarct

dementia and stroke patients ($\underline{N}$ = 17), were included. A series of one-tailed ANOVAs was performed treating mean total scores on the G-NDS, MMSE, CCSE, and NCSE as dependent variables. No significant differences between the two diagnostic groups were found.

CHAPTER 4

DISCUSSION

As cognitive screening instruments, the MMSE, CCSE, and
NCSE should predict performance relative to a more thorough
testing procedure such as a comprehensive neuropsychological
evaluation. The goal of the present study was to examine
the relationship between scores obtained on the MMSE, CCSE,
and NCSE and subsequent performance on the Halstead-Reitan
Neuropsychological Test Battery (HRNTB) in a sample of older
adults with suspected cognitive impairment. Scores on the
General Neuropsychological Deficit Scale (G-NDS), a global
performance measure computed from the HRNTB, served as the
criterion standard by which to judge the presence or absence
of cognitive deficits. The sensitivity, specificity, and
predictive value of each screening test, as well as how well
each test correlated with the G-NDS, were investigated.
Four major hypotheses were postulated.

The first hypothesis predicted that the NCSE total
score would correlate with the G-NDS to a significantly
greater degree than either the MMSE or CCSE total scores.
The results of the present study do not support this hypoth-
esis. Correlations between the MMSE and the G-NDS, the CCSE
and the G-NDS, and the NCSE and the G-NDS were all

95

moderately high (ranging from -0.60 to -0.67), statistically significant, and similar to one another.

These correlational values are substantially higher than those reported by Faustman et al. (1990) in their investigation of the ability of the MMSE to predict performance on the Luria-Nebraska Neuropsychological Battery (LNNB). The correlation between the MMSE and the LNNB Global measure was -0.27. Although this correlation was statistically significant, it accounts for only a small proportion (7%) of variance in LNNB scores. In the present study, the proportion of variance explained by the relationship between the MMSE and G-NDS was 36%. A slightly higher proportion of variance (45%) was explained by the NCSE.

Several explanations for these results may be offered. First of all, although the present study and the investigation by Faustman et al. (1990) both utilized a standardized neuropsychological battery as the criterion measure, the LNNB and HRNTB are composed of different neuropsychological tests. Although there is substantial overlap in the cortical functions assessed by the two batteries and both batteries represent very sensitive measures of brain impairment, they are not identical. Thus, correlation values reported in the present study may have been higher, in part, because of the instrumentation used.

Second, the subject population of the present study was comprised of geriatric patients, the vast majority of whom

were diagnosed with Alzheimer's disease or multi-infarct dementia and scored in the cognitively impaired range on the HRNTB. The sample in the Faustman et al. (1990) study was comprised of a younger, diagnostically mixed sample of psychiatric inpatients, the majority of whom did not evidence cognitive impairment on the LNNB. Therefore, the lower correlations in the Faustman et al. (1990) study may reflect a more heterogeneous sample, whereas higher correlations in the present study may reflect the homogeneity of the research sample. In addition, scores on the cognitive screening tests in the current study demonstrated restricted variability, which could also have inflated correlational values.

Not only were the correlations between each of the three cognitive screening tests and the G-NDS moderately high, but no significant differences among the correlations were found. The similarity among the correlations between the MMSE, CCSE, NCSE, and G-NDS suggests that the screening tests measure many of the same cognitive functions. All three tests assess memory, attention, calculations, and orientation. There was an overlap in the Serial Sevens task between the MMSE and CCSE, and the MMSE and NCSE both assess language skills and constructional abilities. In addition, the NCSE and CCSE both assess abstract reasoning, although the NCSE also includes a specific measure of judgment. Therefore, shared items and the inclusion of similar

cognitive domains are likely to have produced similar correlations with the G-NDS. The high intercorrelations between the MMSE, CCSE, and NCSE further support this notion.

In spite of achieving significant correlations with the G-NDS, the MMSE, CCSE, and NCSE failed to explain a substantial proportion of variance in the dependent measure. As screening tests, however, the MMSE, CCSE, and NCSE do not assess certain abilities that a comprehensive neuropsychological battery like the Halstead does. For example, motor functioning, sensory-perceptual functioning, complex non-verbal problem-solving abilities, and measures of cognitive flexibility are neglected in brief cognitive evaluations. Screening tests like the MMSE, CCSE, and NCSE typically emphasize crystallized verbal skills more than fluid cognitive skills. The ability to live independently and successfully engage in the more complex instrumental activities of daily living requires intact fluid abilities. Failure to comprehensively assess these skills, therefore, may seriously hinder rehabilitation efforts and lead to inappropriate expectations regarding the elderly person's functional capacity.

Results of the exploratory analysis examining the correlations between the various cognitive dimensions assessed by each screening test and the G-NDS revealed that the highest correlations were achieved between the

constructional components of the NCSE and MMSE and the criterion measure. In addition, results of an exploratory stepwise multiple regression analysis found that the best predictors of G-NDS scores were the Constructions subscale of the NCSE, the NCSE Memory subscale, and the MMSE Attention and Calculations task. Combined, these variables accounted for 58% of the variance in the criterion measure. It is believed that performance within the three cognitive domains may have resulted in better prediction of performance on the HRNTB because these tasks are more complex and rely less upon overlearned verbal abilities.

The second hypothesis predicted that the NCSE would demonstrate significantly greater sensitivity, specificity, and positive and negative predictive value than either the MMSE or the CCSE. Results of the present study partially support this hypothesis in that the NCSE demonstrated significantly greater sensitivity (82%) than either the MMSE (20%) or the CCSE (22%).

The findings of low sensitivity for the MMSE and CCSE in the present study contrast sharply with the majority of prior investigations, which have reported sensitivities ranging from 61% to 87% (Anthony et al., 1982; Fields et al., 1992; Kafonek et al., 1984; Kaufman et al., 1979; Webster et al., 1984). The results of the present study indicate that the False Negative rates for the MMSE and CCSE are substantially higher (80% and 78%, respectively) than

previously reported False Negative rates of 43% and 53%, respectively (Schwamm et al., 1987). Findings of superior sensitivity (i.e., a low rate of False Negative decisions) for the NCSE are in agreement with a recent comparative study of the MMSE, CCSE, and NCSE that utilized documented brain pathology as the criterion standard (Schwamm et al., 1987).

The present findings of extremely low sensitivity for the MMSE and CCSE and high sensitivity for the NCSE may be explained by considering several factors. First of all, the present investigation was unique in that it compared performance on all three cognitive screening tests to performance on a widely used, comprehensive neuropsychological battery. Prior investigations have utilized inadequate measures of cognitive functioning such as clinical judgment or laboratory techniques as the criterion standard. When a rigorous criterion standard like performance on the Halstead battery was employed, the inability of brief screening tests like the MMSE and CCSE to accurately identify the presence of cognitive deficits was highlighted.

Second, although the present investigation found the NCSE to be highly sensitive to cognitive impairment, its superiority is due, in part, to the multidimensional scoring system. An example will help to clarify this point. On the MMSE, a person can miss all of the verbal recall items, fail the Copy Design task, and miss two other items, and still

receive a total score in the normal range. On the NCSE, however, performance is considered impaired if deficits are observed on any one of the 10 cognitive domains assessed. Thus, deficits in one ability area are not obscured by computation of a global score that relies on various cutoff points to determine the presence of cognitive impairment. Impairment in any one domain constitutes a "positive" test for the NCSE.

Resulting differences in sensitivity between the NCSE, MMSE, and CCSE are due not only to different scoring systems, however. The NCSE is more comprehensive than either the MMSE or CCSE and includes a graded series of items of increasing difficulty within each ability area. The items within each domain tend to be more demanding, and the degree of impairment can thus be quantified. In contrast, the cognitive dimensions assessed by the MMSE and CCSE include fewer items, and the tasks tend to be less demanding. For these reasons, even adjusting the cutoff scores on the MMSE and CCSE did not substantially improve their sensitivity.

In spite of demonstrating superior sensitivity, the NCSE did not evidence greater specificity or positive or negative predictive value as was postulated in the second hypothesis. All three cognitive screening tests achieved 100% specificity, 100% positive predictive value, and similar negative predictive values (5% for the MMSE and CCSE; 18% for NCSE). These results can be explained by examining

what specificity and predictive value measure in relation to a cognitive screening instrument.

Specificity of a test refers to the proportion of subjects without cognitive deficits who score as "normals" on the screening test (i.e., True Negatives). It is calculated by comparing the number of True Negatives to the number of False Positives, that is, normals who are mistakenly identified as cognitively impaired by the screening test. In the present study, only two subjects were identified as normals according to their performance on the Halstead, and both of these subjects were identified as normals by the screening tests. None of the screening tests identified a subject as impaired who was "normal" (i.e., False Positive). Therefore, specificity for the MMSE, CCSE, and NCSE was perfect because all normals were identified as such, and no False Positive errors were made.

The False Positive rate is also crucial in understanding why all three screening tests evidenced perfect positive predictive value. The predictive value of a positive test is the proportion of those who score as impaired on the screening test who are indeed cognitively impaired. It is calculated by comparing the number of True Positives to False Positives. Because the False Positive rate was 0 for all three screening tests, 100% positive predictive value was achieved, regardless of the number of False Negative errors. Likewise, the negative predictive value compares

the number of True Negatives to False Negatives. Similar values were achieved by all three screening tests because only two normals were included in the present sample. Thus, although the NCSE made far fewer False Negative errors, the small number of normal cases tended to artificially lower this value for the NCSE.

The distinction between the relative importance of specificity and sensitivity is a critical one. A fundamental requirement of any screening test is that it demonstrate high sensitivity (i.e., low rate of False Negative results). For cognitive screening tests in particular, False Negative errors are more serious than False Positive errors. A cognitive screen has limited utility if impairment tends to go undetected in the vast majority of cases. In the present study, the MMSE and CCSE misclassified 80% and 78% of the impaired cases, respectively, as normal. Results such as these suggest that these two instruments demonstrate poor criterion-based and incremental validity.

The findings of equally low sensitivities for the MMSE and CCSE and identical predictive values do not support the third hypothesis, which predicted that the CCSE would demonstrate significantly greater sensitivity and positive and negative predictive value than the MMSE. With the inclusion of items on the CCSE that tap the abstract reasoning ability of subjects, it was thought that more subjects with subtle cognitive disturbances would be identified. In actuality,

all of the abstract reasoning items were passed by 65% of the subjects, with only 35% missing one or more items. Again, reliance upon a unidimensional scoring system allowed numerous persons who may have failed some or all of the items within this domain to receive an overall score in the normal range.

Although it was predicted that the CCSE would be more sensitive to cognitive deficits than the MMSE, the fourth hypothesis predicted that their specificities would not differ. Results of the present study support this hypothesis. This prediction was based on clinical experience, which has suggested that, due to the simplicity of these tests, few individuals are likely to be identified as impaired when indeed they are not.

The primary issue addressed in the present study was whether the MMSE, CCSE, and NCSE could accurately identify the presence of cognitive impairment. Additional exploratory analyses, however, investigated whether the three screening tests could predict level of impairment, as defined by the G-NDS. Two additional criterion standards, the Impairment Index and the Average Impairment Rating, were further employed for this purpose. Results suggested that the MMSE and CCSE were unable to detect any cases of mild impairment and misclassified 80% of the moderately impaired subjects as normal. Although a relatively greater number of

severely impaired individuals were recognized as such by the
MMSE and CCSE, almost half were misclassified as normal.

The NCSE fared somewhat better in that it detected 30%
of the mildly impaired cases and misclassified only a small
number of moderately impaired persons as normal. No se-
verely impaired subjects were misclassified as normal by the
NCSE. The general trend was for all three screening tests
to underestimate the degree of cognitive impairment for the
majority of cases, and these results were consistent across
different criterion standards. It is important to note,
however, that the NCSE does not claim to be able to predict
level of impairment for overall performance. It rates se-
verity of cognitive impairment within each cognitive domain.
A global rating was computed in the present study only for
comparative purposes.

Comparison of the results of the present study with
prior studies that have addressed the effects of the major
sociodemographic variables on cognitive test performance
yielded mixed results. Although prior investigations have
found poorer performance associated with increasing age and
lower educational achievement (Brayne & Calloway, 1990;
Cavanaugh & Wettstein, 1983; Escobar et al., 1986; Uhlmann &
Larson, 1991), results of the present study suggest that
only education was associated with poorer performance on all
three cognitive screening tests, as well as performance on
the Halstead battery.

The impact of education has been a concern of those who want to reduce the number of False Positive errors for cognitive screening tests. Results of the present investigation, however, strongly suggest that the rate of False Negative errors poses a much greater threat to the validity of these instruments. Given the high rate of False Negative errors for the MMSE and CCSE, adjusting cutoff scores for different educational levels would serve only to decrease their already unacceptably low sensitivity. Before the influence of education and other sociodemographic variables that may mediate performance on cognitive screening tests can be meaningfully addressed, attention needs to be redirected to the basic criterion-related validity of these instruments.

In summary, the results of this investigation found that, although the MMSE, CCSE, and NCSE were all significantly correlated with the G-NDS, only the NCSE demonstrated an appropriate balance between high sensitivity and specificity. When a rigorous neuropsychological evaluation was employed as the criterion standard, the NCSE accurately detected the presence of cognitive impairment in 82% of the cases. The MMSE and CCSE, however, failed to detect the presence of cognitive deficits in approximately 80% of the cases. The insensitivity of the MMSE and CCSE appears to be related to the limited number of cognitive abilities assessed, the restricted number of items within the domains,

and the simplicity and undemanding nature of the tasks.
Furthermore, reliance upon a global score often served to
obscure deficits within one or several ability areas.

On the other hand, the heightened sensitivity of the
NCSE appears to be a result of several unique characteris-
tics of this instrument. The NCSE does not combine the
results of performance in different cognitive areas into one
total score. Thus, failure within one cognitive domain is
not masked by intact functioning in others. In addition,
the NCSE assesses more areas of cognitive function than
either the MMSE or CCSE and does so more comprehensively. A
graded series of increasingly difficult items within each
cognitive area increases the likelihood of detecting mild or
isolated deficits and allows the degree of impairment to be
quantified. Results further support the notion that cogni-
tively complex tasks are better predictors of overall
neuropsychological functioning than relatively simple verbal
tasks.

The current findings clearly indicate that the MMSE and
CCSE may have limited utility in the identification of
cognitive impairment in older adults. The present study,
however, had certain limitations. With regard to internal
validity, the present study did not allow for the random
selection of subjects. Thus, selective sampling undoubtedly
occurred. Subjects were selected, however, on the basis of
consecutive referrals for evaluation during a specified time

period and, as such, were fairly representative of the type
of clinical population often seen in medical inpatient and
outpatient settings. Furthermore, the way the data were
collected represents the way that cognitive screening tests
are utilized in clinical practice. Given the lengthy and
demanding nature of the Halstead battery, fatigue may have
influenced performance on the criterion measure; but it
should be noted that testing was divided into two sessions
and that the order of administration of the various tests
comprising the Halstead battery interspersed more difficult
tasks with less demanding ones. Both of these precautions
served to minimize the effects of fatigue and maximize the
likelihood that subjects performed to the best of their
ability.

External validity was reduced in the present study by
the modest sample size and lack of random sampling
procedures. In addition, the present sample was fairly
homogeneous in relation to certain personal variables such
as race, sex, and education. Consequently, the present
sample may not be representative of the population of cogni-
tively impaired elderly in general.

Suggestions for future research on the utility of cog-
nitive screening instruments would emphasize the crucial
importance of utilizing a rigorous criterion standard,
rather than a clinician's judgment or laboratory techniques,
to establish the presence and severity of cognitive

impairment. Standardized neuropsychological batteries like the Halstead-Reitan or Luria-Nebraska have been the subject of extensive research that has demonstrated their accuracy in the detection of cognitive deficits, and they appear to be well suited for this purpose. In addition, the reduction of False Negatives should be a high priority. Results of the present study suggest that merely adjusting the cutoff scores on the MMSE and CCSE did not substantially improve their sensitivity. Future research, however, could investigate whether the inclusion of a wider range of cognitively more complex items and the implementation of a multidimensional scoring system could improve the utility of these instruments. More rigorous assessment of right hemisphere function, including visual-spatial skills and constructional abilities, could also serve to substantially increase their sensitivity. Finally, given that the NCSE has demonstrated impressive sensitivity in the detection of cognitive impairment in a sample of elderly adults, future studies need to explore the validity of this instrument with more diverse populations suffering from a variety of organic mental syndromes.

APPENDIX A

RULES FOR COMPUTING THE GENERAL NEUROPSYCHOLOGICAL

DEFICIT SCALE SCORE

Level of Performance

Variable	0	1	2	3
1. Verbal IQ	90+	82–89	73–81	≤72
2. Performance IQ	90+	82–89	73–81	≤72
3. Impairment Index	0–.2	.3–.4	.5–.7	.8–1.0
4. Category Test	0–25	26–45	46–64	65+
5. TPT — Total Time	0'–9.0'	9.1'–15.0'	15.1'–25.0'	25.1'+
6. TPT — Memory	8–10	7	4–6	0–3
7. TPT — Localization	7–10	6	3–5	0–2
8. Seashore Rhythm Test (# correct)	28–30	25–27	20–24	0–19
9. Speech-sounds Perception Test (errors)	0–6	7–10	11–15	16+
10. Finger Tapping — Dominant Hand	55+	50–54	41–49	0–40
11. Finger Tapping — Non-dominant Hand	49+	45–48	37–44	0–36
12. Trail Making Test — Part A	0"–26"	27"–39"	40"–51"	52"+
13. Trail Making Test — Part B	0"–65"	66"–85"	86"–120"	121"+
14. Tactile Form Recognition — Total Time	0"–16"	17"–23"	24"–33"	34"+
15. Bilateral Tactile Stimulation — Total errors	0	1	2–3	4+
16. Bilateral Auditory Stimulation — Total errors	0	1	2	3+
17. Bilateral Visual Stimulation — Total errors	0	1	2–3	4+
18. Tactile Finger Recognition — Both hands (errors)	0–2	3–4	5–8	9+
19. Finger-tip Number Writing — Both hands (errors)	0–3	4–6	7–11	12+

Pathognomonic Signs

Variable	Score
20. Dysnomia	3
21. Auditory verbal dysgnosia	3
22. Visual number dysgnosia	3
23. Visual letter dysgnosia	3
24. Body dysgnosia	3
25. Dyscalculia	2
26. Dysgraphia	2
27. Dyslexia	2
28. Constructional dyspraxia	2
29. Central dysarthria	1
30. Spelling dyspraxia	1
31. Right-Left confusion	1

Patterns

Variable	0	1	2	3
32. Verbal IQ/Performance IQ Difference	0–5	6–10	11–19	20 +
33. If Impairment Index 0.0–.4, score is 0. If Impairment Index .5–1.0, then derive score from Full Scale IQ.	If FS IQ: <90	90–95	96–100	101 +

Right-Left Differences

Variable	0	1	2	3
For Variables 34, 35, and 36: Divide non-dominant hand by dominant hand and subtract from 1.0				
34. Finger tapping	.08–.12	.13–.16 .07–.05	.17–.21 .04–(–.03)	.22 or more (–.04 or less)
35. TPT	.38–.26	.25–.15 .39–.42	.14–.05 .43–.50	.04 or less .51 or more
36. Grip strength	.08–.12	.13–.17 .07–.06	.18–.20 .05–.00	.21 or more (–.01 or less)
37. Tactile Form Recognition	0–1″	2″–3″	4″–5″	6″ +
The score is the difference in seconds between the two hands				
For Variables 38, 39, and 40: The score is the difference in errors between the right and left side				
38. Bilateral Tactile Stimulation	0	1	2	3 +
39. Bilateral Auditory Stimulation	0	1	2	3 +
40. Bilateral Visual Stimulation	0	1	2	3 +
For Variables 41 and 42: (1) Determine the percentage of errors by the hand with the greater number of errors; (2) Enter the table in the row with the total errors made; (3) Determine the score				

41. Tactile Finger Recognition

42. Finger-tip Number Writing

Table for Variables 41 and 42:

Total Errors	0	1	2	3
21 or more	50–54	55–57	58–60	61 +
18–20	50–54	55–57	58–62	63 +
15–17	50–54	55–58	59–63	64 +
12–14 %	50–55	56–58	59–63	64 +
9–11	50–56	57–59	60–63	64 +
6–8	50–56	57–63	64–70	71 +
3–5	50–59	60–67	68–79	80 +
0–2	50	1 error	100	—

APPENDIX B

RATING EQUIVALENTS OF RAW SCORES USED TO COMPUTE THE

AVERAGE IMPAIRMENT RATING

Test	0 Normal	1	2 Mild	3 Moderate	4 Moderately Severe	5 Severe
Average impairment rating	0.00-1.00	1.01-1.35	1.36-2.00	2.01-2.85	2.86-3.50	3.51-5.00
Category errors	≤ 25	26-52	53-75	76-105	106-131	132+
TPT total time	≤ 9.0	9.1-15.6	15.7-21	21.1-29.9	30 min. (14-30 blocks)	30 min. (13-0 blocks)
TPT memory (no. correct)	10-9	8-6	5-4	3-2	1	0
TPT location (no. correct)	10-7	6-5	4-3	2-1	0 and TPT memory > 0	0 and TPT memory = 0
Speech errors	0-3	4-7	8-14	15-25	26-30	31+
Rhythm errors	0-2	3-5	6-9	10-13	14-18	19+
Tapping (no.)						
Dom M	≥ 55	54-50	49-43	42-32	31-20	19-0
Dom F	≥ 51	50-46	45-39	38-28	27-16	15-0
Nondom M	≥ 49	48-44	43-37	36-26	25-14	13-0
Nondom F	≥ 45	44-40	39-33	32-22	21-10	9-0
Trails B (time)	≤ 57	58-87	88-123	124-186	187-275	276+
Aphasia (errors)	0	1-6	7-15	16-25	26-40	41+
Spatial relations (errors)	1	2-3	4-5	6-7	8-9	10-12
Perceptual (errors)	0-4	5-12	13-30	31-50	51-80	81+

APPENDIX C

TABLES AND ILLUSTRATIONS

Table 1

Intercorrelation Matrix of Measures of Cognitive Impairment

	AIR	MMSE	CCSE	NCSE	G-NDS	II
AIR	1.0000	-.5933**	-.6151**	-.6756**	.9264**	.7057**
MMSE	-.5933**	1.0000	.8872**	.7521**	-.5996**	-.4115**
CCSE	-.6151**	.8872**	1.0000	.7753**	-.6160**	-.4545**
NCSE	-.6756**	.7521**	.7753**	1.0000	-.6655**	-.5020**
G-NDS	.9264**	-.5996**	-.6160**	-.6655**	1.0000	.7070**
II	.7057**	-.4115**	-.4545**	-.5020**	.7070**	1.0000

Note. AIR = Average Impairment Rating; MMSE = Mini-Mental State Examination; CCSE = Cognitive Capacity Screening Examination; NCSE = Neurobehavioral Cognitive Status Examination; G-NDS = General Neuropsychological Deficit Scale; II = Impairment Index.

*p < .05, two-tailed.
**p < .01, two tailed.

Table 2

Correlation Coefficients for Cognitive Dimensions Assessed
by the MMSE with the G-NDS Total Score

	G-NDS
MMSEOR	-.4329**
MMSERG	X
MMSEAC	-.4960**
MMSERC	-.2757*
MMSENA	X
MMSECO	-.0277
MMSERP	X
MMSERD	-.2966*
MMSEWR	-.2934*
MMSECD	-.5147**

Note. MMSE = Mini-Mental State Examination; OR = Orienta-
tion; RG = Registration; AC = Attention/Calculation; RC =
Recall; NA = Naming; CO = 3-Stage Command; RP = Repetition;
RD = Reading; WR = Writing; CD = Copy Design; X = coeffi-
cient cannot be computed.

 *p < .05, two-tailed.
**p < .01, two-tailed.

Table 3

Correlation Coefficients for Cognitive Dimensions Assessed by the CCSE with the G-NDS Total Score

	G-NDS
CCSEOR	-.4634**
CCSEAT	-.3536*
CCSECA	-.3503*
CCSEAR	-.4617**
CCSEME	-.2830*
CCSESS	-.5015**

Note. CCSE = Cognitive Capacity Screening Examination; OR = Orientation; AT = Attention; CA = Calculation; AR = Abstract Reasoning; ME = Memory; SS = Serial Sevens.

 *p < .05, two-tailed.
**p < .01, two-tailed.

Table 4

Correlation Coefficients for Cognitive Domains Assessed by
the NCSE with the G-NDS Total Score

	G-NDS
NCSE 1	-.3293*
NCSE 2	-.4272**
NCSE 3	-.3194*
NCSE 4	-.0818
NCSE 5	-.4654**
NCSE 6	-.6870**
NCSE 7	-.4702**
NCSE 8	-.3781**
NCSE 9	-.4191**
NCSE 10	-.4531**

Note. NCSE = Neurobehavioral Cognitive Status Examination;
1 = Orientation; 2 = Attention; 3 = Comprehension; 4 = Repe-
tition; 5 = Naming; 6 = Constructions; 7 = Memory; 8 = Cal-
culations; 9 = Similarities; 10 = Judgment.

*p < .05, two-tailed.
**p < .01, two-tailed.

Table 5

Profile of Subjects, Medical Diagnoses Prompting Referral, and Test Scores on the MMSE, CCSE, NCSE, G-NDS, II, and AIR

Subj No	Age	Sex	Educ	MMSE Tot Score	CCSE Tot Score	NCSE No Imp Scales	G-NDS Tot Score	II	AIR	Medical Problem or Diagnosis Prompting Referral
1	77	M	16	21	18	3	53	0.9	3.00	AD
2	59	F	9	27	28	0	50	1.0	2.91	AD
3	76	F	12	25	21	2	45	0.9	2.55	AD
4	81	M	18	30	29	2	42	0.9	2.27	AD
5	86	F	18	27	27	4	87	1.0	3.91	MD
6	70	F	12	30	26	1	36	0.7	1.73	MD
7	85	F	18	25	24	2	47	0.9	2.64	AD
8	57	F	12	29	29	0	38	0.3	1.56	Head injury (closed)
9	79	F	14	26	26	1	40	0.9	1.82	AD
10	69	M	14	27	23	1	40	0.7	2.36	Parkinson's Disease
11	70	F	12	30	24	2	47	1.0	2.00	AD
12	70	F	12	24	22	1	58	1.0	3.36	AD
13	74	F	12	25	22	2	43	0.9	2.09	Rheumatoid arthritis-- systemic disease
14	70	M	12	25	20	4	74	0.9	3.00	R-CVA
15	60	M	18	30	30	0	18	0.4	1.09	Head injury (closed)
16	58	F	9	25	25	0	39	0.9	1.64	Cryptogenic epilepsy
17	76	M	14	27	19	0	43	0.9	2.45	R-CVA
18	68	M	16	29	30	0	61	0.9	3.18	MD
19	61	M	16	27	25	2	39	0.7	1.73	AD
20	59	M	16	28	26	1	43	0.9	1.64	R-CVA
21	65	M	8	20	14	4	54	0.7	2.55	Head injury (closed)-- bilateral frontal damage
22	55	M	12	29	22	2	54	0.9	2.45	Frontal-CVA
23	73	M	11	24	15	3	63	1.0	2.91	MD
24	57	M	8	17	10	8	80	1.0	3.73	Anoxia following MI
25	66	M	14	29	25	0	36	0.7	2.00	AD
26	65	M	8	26	26	2	56	0.9	2.64	AD
27	62	M	12	28	22	2	64	1.0	3.45	MD

Subj No	Age	Sex	Educ	MMSE Tot Score	CCSE Tot Score	NCSE No Imp Scales	G-NDS Tot Score	II	AIR	Medical Problem or Diagnosis Prompting Referral
28	61	M	16	29	29	1	27	0.6	1.64	AD
29	62	M	14	30	30	4	47	0.7	2.55	AD
30	70	M	8	26	24	2	65	1.0	3.00	AD
31	68	M	12	26	22	4	63	0.9	2.91	R-Brain tumor (post surgery)
32	75	F	9	20	18	8	64	1.0	3.27	AD
33	66	M	14	28	29	0	30	0.6	1.64	Frontal lobe dementia
34	76	F	12	15	8	7	90	1.0	4.09	AD
35	70	M	11	26	27	1	68	0.9	2.73	R-CVA
36	65	M	12	26	26	2	53	0.9	2.55	Maintenance ECT for major depression
37	68	M	18	24	23	1	43	1.0	1.45	Unspecified seizure disorder
38	59	M	16	28	25	1	56	1.0	3.00	Head injury (closed)
39	57	M	16	27	27	2	45	0.9	2.64	Head injury (closed)
40	62	M	12	28	29	0	20	0.4	1.27	Head injury (closed)
41	58	M	10	24	20	5	62	1.0	3.36	L-CVA
42	77	M	10	20	18	1	81	1.0	3.73	MD
43	59	F	20	30	30	0	26	0.3	1.09	Head injury (closed)
44	80	F	6	26	24	6	73	1.0	3.55	AD
45	63	M	10	29	22	4	58	1.0	2.91	Pons-CVA
46	74	M	18	29	29	0	53	1.0	2.45	Korsakoff's syndrome
47	72	M	14	23	20	4	53	1.0	2.55	AD
48	64	M	12	16	10	5	75	1.0	3.91	MD
49	57	M	12	25	24	2	64	1.0	2.91	MD
50	55	M	13	23	19	1	80	1.0	3.45	R-CVA
51	70	F	15	19	17	5	56	0.9	3.00	AD
52	61	F	12	28	26	2	45	1.0	2.36	MD

Note. MMSE = Mini-Mental State Examination; CCSE = Cognitive Capacity Screening Exam; NCSE = Neuro-behavioral Cognitive Status Exam; G-NDS = General Neuropsychological Deficit Scale; II = Impairment Index; AIR = Average Impairment Rating. AD = Alzheimer's Disease; MD = Multi-Infarct Dementia; CVA = Cerebrovascular Accident; MI = Myocardial Infarct; R = Right Cerebral Hemisphere; L = Left Cerebral Hemisphere.

Table 6

Sensitivity and Specificity for the MMSE and CCSE with Ad-
justment of Cutoff Scores

MMSE Standard Cutoff < 24 Points		MMSE Adjusted Cutoff < 27 Points	
Sensitivity:	20%	Sensitivity:	54%
Specificity:	100%	Specificity:	100%
Positive predictive value:	100%	Positive predictive value:	100%
Negative predictive value:	5%	Negative predictive value:	8%
False positive rate:	0	False positive rate:	0
False negative rate:	80%	False negative rate:	46%
CCSE Standard Cutoff < 20 Points		CCSE Adjusted Cutoff < 25 Points	
Sensitivity:	22%	Sensitivity:	56%
Specificity:	100%	Specificity:	100%
Positive predictive value:	100%	Positive predictive value:	100%
Negative predictive value:	5%	Negative predictive value:	8%
False positive rate:	0	False positive rate:	0
False negative rate:	78%	False negative rate:	44%

Table 7

Comparisons between Level of Impairment Ratings for the MMSE
with the G-NDS

G-NDS	MMSE			Row Total
	Normal	Mild	Moderate/ Severe	
Normal	2			2 3.8
Mild	10			10 19.2
Moderate	26	5		31 59.6
Severe	4	2	3	9 17.3
Column Total	42 80.8	7 13.5	3 5.8	52 100.0

Note. N = 52 cases.

Table 8

Comparisons between Level of Impairment Ratings for the CCSE
with the G-NDS

G-NDS	CCSE Normal	Mild	Moderate/ Severe	Row Total
Normal	2			2 3.8
Mild	10			10 19.2
Moderate	25	6		31 59.6
Severe	4	4	1	9 17.3
Column Total	41 78.8	10 19.2	1 1.9	52 100.0

Note. N = 52 cases.

Table 9

Comparisons between Level of Impairment Ratings for the NCSE
with the G-NDS

	NCSE				
G-NDS	Normal	Mild	Moderate	Severe	Row Total
Normal	2				2 3.8
Mild	5	3	2		10 19.2
Moderate	4	8	15	4	31 59.6
Severe		2	4	3	9 17.3
Column Total	11 21.2	13 25.0	21 40.4	7 13.5	52 100.0

Note. N = 52 cases.

Table 10

Comparisons between Level of Impairment Ratings for the MMSE
with the Average Impairment Rating (AIR)

AIR

MMSE	Normal	Mild	Moderate	Moderately Severe	Severe	Row Total
Normal	3	11	14	12	2	42 80.8
Mild			2	4	1	7 13.5
Moderate/ Severe					3	3 5.8
Column Total	3 5.8	11 21.2	16 30.8	16 30.8	6 11.5	52 100.0

Note. N = 52 cases.

Table 11

Comparisons between Level of Impairment Ratings for the CCSE
with the Average Impairment Rating (AIR)

CCSE	AIR					
	Normal	Mild	Moder-ate	Moder-ately Severe	Severe	Row Total
Normal	3	11	14	11	2	41 78.8
Mild			2	5	3	10 19.2
Moderate/ Severe					1	1 1.9
Column Total	3 5.8	11 21.2	16 30.8	16 30.8	6 11.5	52 100.0

Note. N = 52 cases.

Table 12

Comparisons between Level of Impairment Ratings for the NCSE
with the Average Impairment Rating (AIR)

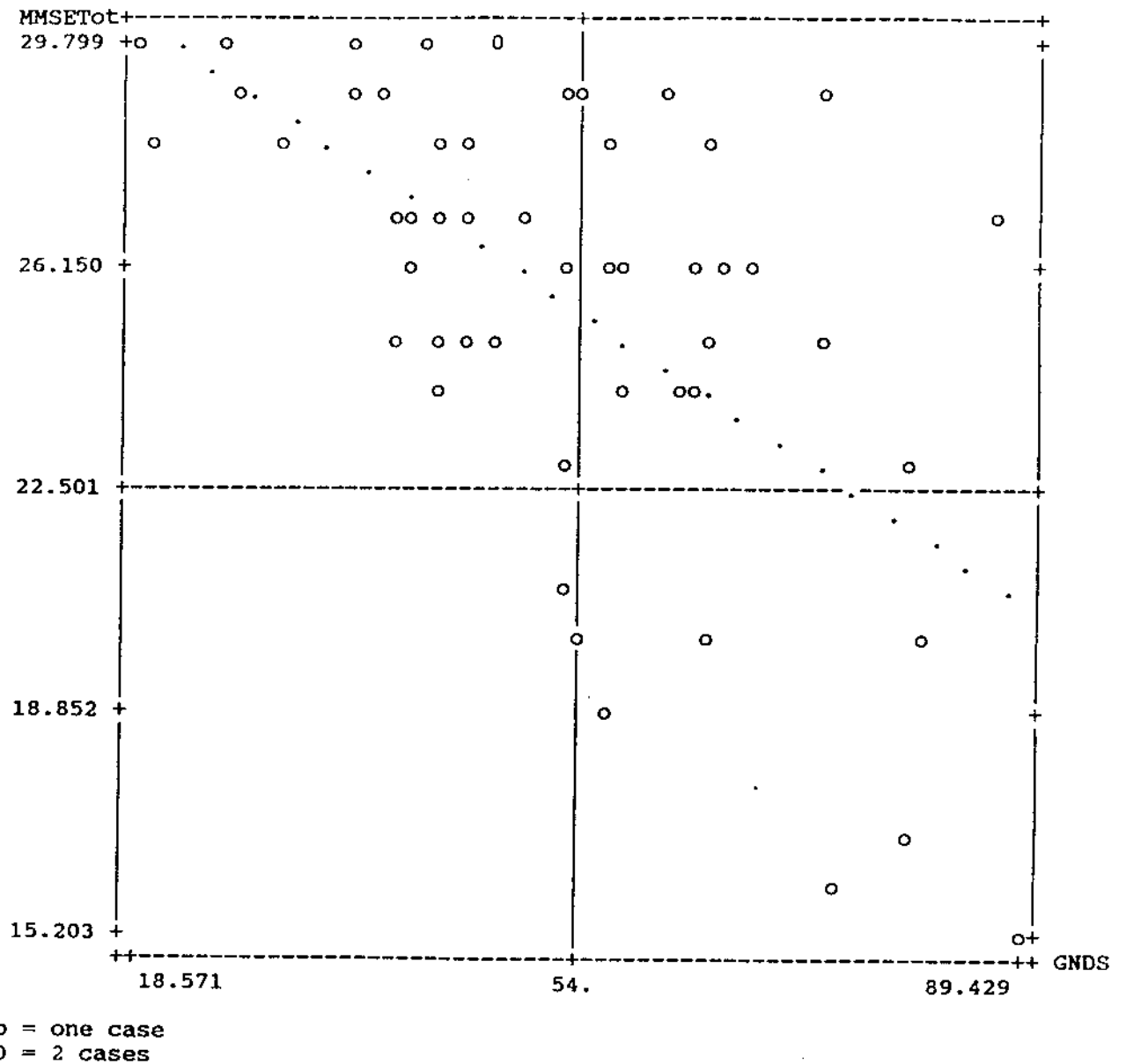| AIR | NCSE | | | | Row |
	Normal	Mild	Moderate	Severe	Total
Normal	3				3 5.8
Mild	4	4	2	1	11 21.2
Moderate	2	5	7	2	16 30.8
Moderately Severe	2	2	8	4	16 30.8
Severe		2	4		6 11.5
Column Total	11 21.2	13 25.0	21 40.4	7 13.5	52 100.0

Note. N = 52 cases.

Figure 1. Scattergram of the relationship between MMSE scores and scores on the General Neuropsychological Deficit Scale (G-NDS); 52 cases.
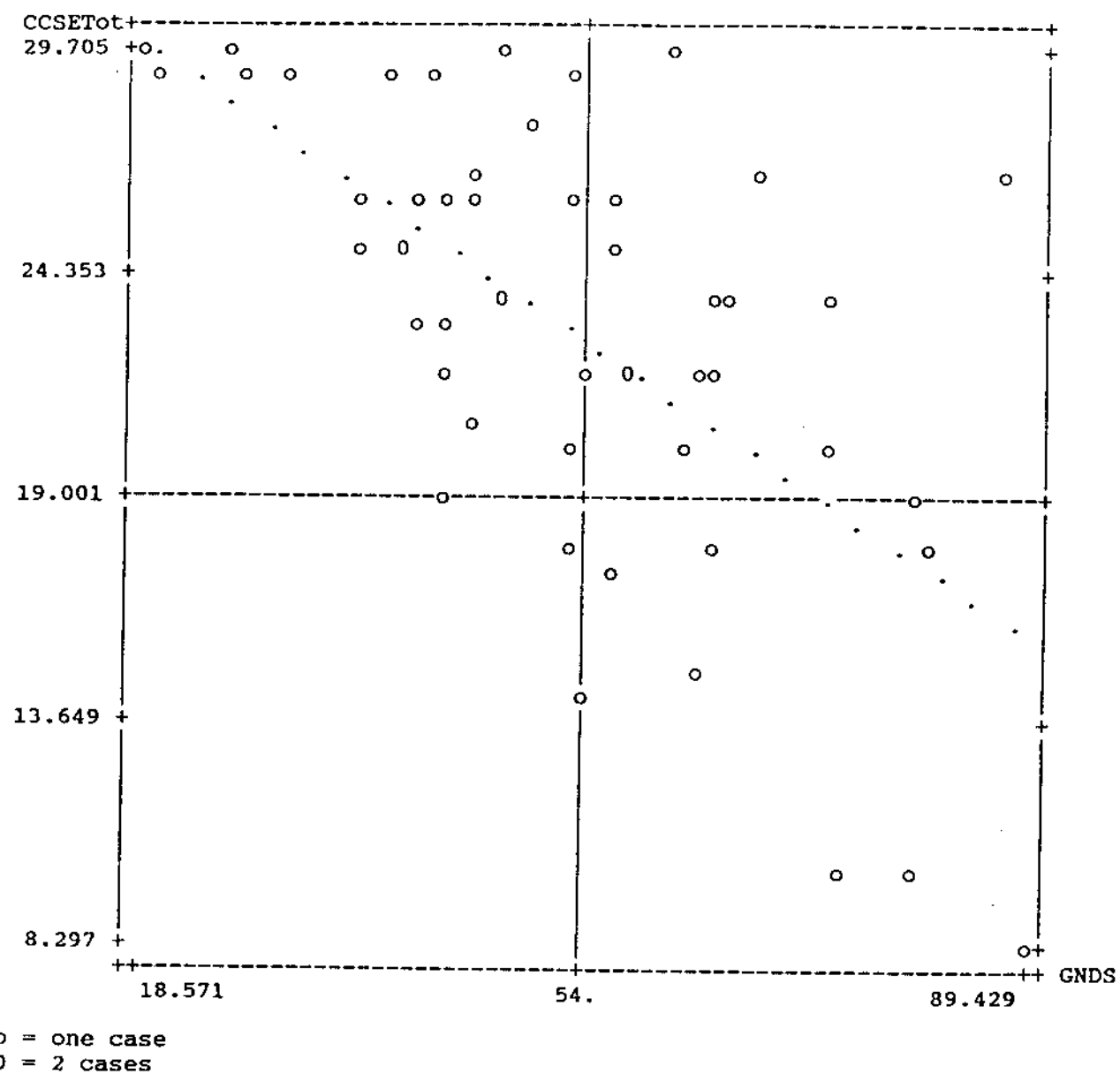
Figure 2. Scattergram of the relationship between CCSE scores and scores on the General Neuropsychological Deficit Scale (G-NDS); 52 cases.
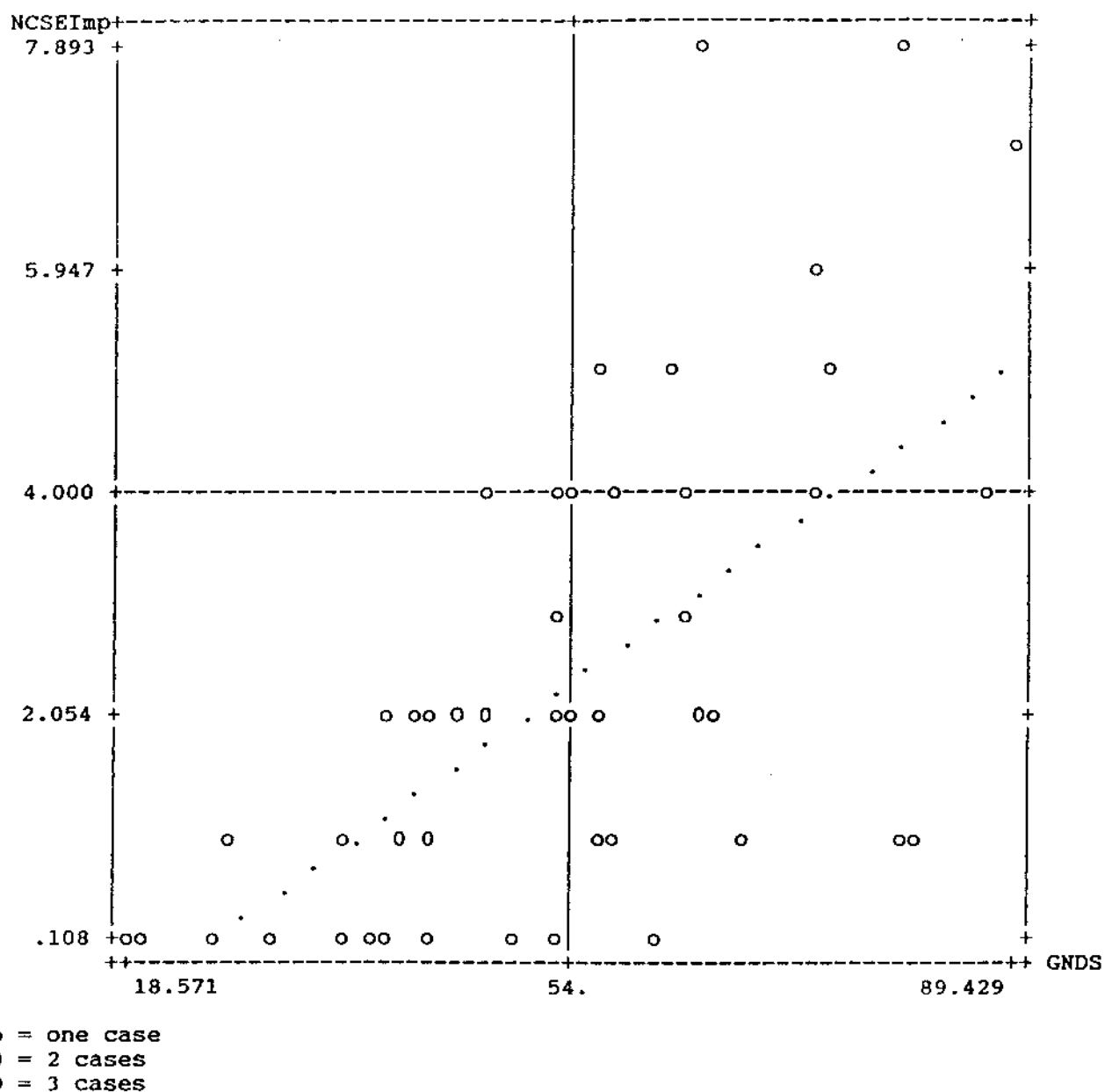
Figure 3. Scattergram of the relationship between NCSE scores and scores on the General Neuropsychological Deficit Scale (G-NDS); 52 cases.

MMSE

	FP	TP
	0	10
	2	40
	TN	FN

Figure 4. Number of false positive (FP), true positive (TP), true negative (TN), and false negative (FN) decisions for the MMSE, based on a comparison with the General Neuropsychological Deficit Scale (G-NDS); total N = 52.

CCSE

	FP	TP
	0	11
	2	39
	TN	FN

Figure 5. Number of false positive (FP), true positive
(TP), true negative (TN), and false negative (FN) decisions
for the CCSE, based on a comparison with the General Neuro-
psychological Deficit Scale (G-NDS); total N = 52.

NCSE

	FP	TP
	0	41
	2	9
	TN	FN

Figure 6. Number of false positive (FP), true positive (TP), true negative (TN), and false negative (FN) decisions for the NCSE, based on a comparison with the General Neuropsychological Deficit Scale (G-NDS); total $\underline{N}$ = 52.

APPENDIX D

DEFINITIONS OF TERMS AND FORMULAS FOR CALCULATING

SENSITIVITY, SPECIFICITY, AND POSITIVE AND

NEGATIVE PREDICTIVE VALUES OF A TEST

TERM	DEFINITION	FORMULA
Sensitivity	Fraction of time a test (in this case, cognitive screen) makes a positive diagnosis when the disorder (cognitive impairment) is present	$\dfrac{TP}{TP + FN}$
Specificity	Fraction of time a test makes a negative diagnosis when the disorder is absent	$\dfrac{TN}{TN + FP}$
Positive Predictive Value	Proportion of those with a positive test who have the disorder	$\dfrac{TP}{TP + FP}$
Negative Predictive Value	Proportion of those with a negative test who do not have the disorder	$\dfrac{TN}{TN + FN}$
False Positive Rate	Proportion of those without the disorder who have a positive test	$\dfrac{FP}{FP + TN}$ or $1 - $ specificity
False Negative Rate	Proportion of those with the disorder who have a negative test	$\dfrac{FN}{FN + TP}$ or $1 - $ sensitivity

TP = True Positives
FP = False Positives
TN = True Negatives
FN = False Negatives

Example--50 cases, ideal results

FP	TP
0	25
25	0
TN	FN

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{25}{25} = 1.0 = 100\%$$

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{25}{25} = 1.0 = 100\%$$

$$\text{Positive Predictive Value} = \frac{TP}{TP+FP} = \frac{25}{25} = 1.0 = 100\%$$

$$\text{Negative Predictive Value} = \frac{TN}{TN+FN} = \frac{25}{25} = 1.0 = 100\%$$

$$\text{False Positive Rate} = \frac{FP}{FP+TN} = \frac{0}{25} = 0$$

$$\text{False Negative Rate} = \frac{FN}{FN+TP} = \frac{0}{25} = 0$$

REFERENCES

Adams, R. L., Craig, P. L., & Parsons, O. A. (1986). Neuropsychology of dementia. <u>Neurologic Clinics</u>, <u>4</u>(2), 387-404.

Albert, M. S. (1991). Neuropsychological testing. In J. Sadavoy, L. W. Lazarus, & L. F. Jarvik (Eds.), <u>Comprehensive review of geriatric psychiatry</u> (pp. 223-244). Washington, DC: American Psychiatric Press.

Anastasi, A. (1950). The concept of validity in the interpretation of test scores. <u>Educational and Psychological Measurement</u>, <u>10</u>, 67-68.

Anastasi, A. (1982). <u>Psychological testing</u> (5th ed.). New York: Macmillan.

Anthony, J. C., LeResche, L., Niaz, U., Korff, M. R., & Folstein, M. F. (1982). Limits of the "Mini-Mental State" as a screening test for dementia and delirium among hospital patients. <u>Psychological Medicine</u>, <u>12</u>, 397-408.

Aske, D. (1990). The correlation between Mini-Mental State Examination scores and Katz ADL status among dementia patients. <u>Rehabilitation Nursing</u>, <u>15</u>(3), 140-142.

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

Berkman, L. (1986). The association between educational attainment and mental status examinations: Of etiologic significance for senile dementia or not? <u>Journal of Chronic Disease</u>, <u>39</u>(3), 171-175.

Brayne, C., & Calloway, P. (1990). The association of education and socioeconomic status with the Mini-Mental State Examination and the clinical diagnosis of dementia in elderly people. <u>Age and Ageing</u>, <u>19</u>(2), 91-96.

Breslow, R., Kocsis, J., & Belkin, B. (1980). Memory deficits in depression: Evidence utilizing the Wechsler Memory Scale. <u>Perceptual and Motor Skills</u>, <u>51</u>, 541-542.

Brink, T. L., Yesavage, J. A., Lum, D., Heersema, P., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. Clinical Gerontologist, 1(1), 37-43.

Burch, E. A., & Andrews, S. R. (1987). Comparison of two cognitive rating scales in medically ill patients. International Journal of Psychiatry in Medicine, 17, 193-200.

Caine, E. (1986). The neuropsychology of depression: The pseudodementia syndrome. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp. 177-184). New York: Oxford University Press.

Cavanaugh, S. V., & Wettstein, R. M. (1983). The relationship between severity of depression, cognitive dysfunction, and age in medical inpatients. American Journal of Psychiatry, 140(4), 495-496.

Chapman, L. F., & Wolff, H. G. (1959). The cerebral hemisphere and the highest integrative functions of man. Archives of Neurology, 1, 357-424.

Cullum, C. M., Thompson, L. L., & Heaton, R. K. (1989). The use of the Halstead-Reitan Test Battery with older adults. Clinics in Geriatric Medicine, 5(3), 595-610.

Cummings, J. L., & Benson, F. (1986). Dementia of the Alzheimer type: An inventory of diagnostic clinical features. Journal of the American Geriatrics Society, 34, 12-19.

Damasio, A. (1985). The frontal lobes. In K. Heilman & E. Valenstein (Eds.), Clinical neuropsychology (pp. 339-375). New York: New York University Press.

DePaulo, J. R., & Folstein, M. F. (1978). Psychiatric disturbances in neurological patients: Detection, recognition, and hospital course. Annals of Neurology, 4, 225-228.

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

Dick, J. P. R., Guiloff, R. J., Stewart, A., Blackstock, J., Bielawska, C., & Paul, E. A. (1984). Mini-Mental State Examination in neurological patients. Journal of Neurology, Neurosurgery, and Psychiatry, 47, 496-499.

Ebrahim, S., Nouri, F., & Barer, D. (1985). Cognitive impairment after stroke. Age and Ageing, 14, 345-350.

Ernst, J. (1988). Language, grip strength, sensory-perceptual, and receptive skills in a normal elderly sample. Clinical Neuropsychology, 2, 30-40.

Escobar, J. I., Burnham, A., Karno, M., Forsythe, A., Landsverk, J., & Golding, J. M. (1986). Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. Journal of Nervous and Mental Disease, 174, 607-614.

Farber, J. F., Schmitt, F. A., & Logue, P. E. (1988). Predicting intellectual level from the Mini-Mental State Examination. Journal of the American Geriatrics Society, 36, 509-510.

Faustman, W. O., Moses, J. A., & Csernansky, J. G. (1990). Limitations of the Mini-Mental State Examination in predicting neuropsychological functioning in a psychiatric sample. Acta Psychiatrica Scandinavica, 81, 126-131.

Fields, S. D., Fulop, G., Sachs, C. J., Strain, J., & Fillit, H. (1992). Usefulness of the Neurobehavioral Cognitive Status Examination in the hospitalized elderly. International Psychogeriatrics, 4(1), 93-102.

Fitzhugh, K. B., Fitzhugh, L. C., & Reitan, R. M. (1961). Psychological deficits in relation to acuteness of brain dysfunction. Journal of Consulting Psychology, 25, 61-66.

Flicker, C., Ferris, S. H., Crook, T. Bartus, R. T., & Reisberg, B. (1986). Cognitive decline in advanced age: Future directions for the psychometric differentiation of normal and pathological age changes in cognitive function. Developmental Neuropsychology, 2(4), 309-322.

Folstein, M. F., Anthony, J. C., Parhad, I., Duffy, B., & Gruenberg, E. M. (1985). The meaning of cognitive impairment in the elderly. Journal of the American Geriatrics Society, 33, 228-235.

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 Psychiatric Research, 12, 189-198.

Folstein, M. F., & Rovner, B. (1986). Implications of changing concepts of dementia for nursing homes. In M. S. Harper & B. D. Lebowitz (Eds.), Mental illness in nursing homes: Agenda for research (pp. 309-320). Washington, DC: U.S. Department of Health and Human Services.

Foreman, M. D. (1987). Reliability and validity of mental status questionnaires in elderly hospitalized patients. Nursing Research, 36(4), 216-220.

Franzen, M. D. (1989). Reliability and validity in neuropsychological assessment. New York: Plenum Press.

Fromm-Auch, D., & Yeudall, L. T. (1983). Normative data for the Halstead-Reitan neuropsychological tests. Journal of Clinical Neuropsychology, 5, 221-238.

Fuld, P. A. (1983). Psychometric differentiation of the dementias: An overview. In B. Reisberg (Ed.), Alzheimer's disease (pp. 201-210). New York: Free Press.

Garcia, C. A., Tweedy, J. R., & Blass, J. P. (1984). Underdiagnosis of cognitive impairment in a rehabilitation setting. Journal of the American Geriatrics Society, 32, 339-342.

Goldstein, S. G., Deysach, R. E., & Kleinknecht, R. A. (1973). Effects of experience and amount of information on identification of cerebral impairment. Journal of Consulting and Clinical Psychology, 41, 30-34.

Grady, G. L., Haxby, J. V., Horwitz, B., & Sundaram, M. (1989). Longitudinal study of the early neuropsychological changes in dementia of the Alzheimer type. Journal of Clinical and Experimental Neuropsychology, 10, 576-596.

Greenough, W. T., & Green, E. J. (1981). Experience and the changing brain. In J. McGaugh & S. Kiesler (Eds.), Aging: Biology and behavior (pp. 159-193). New York: Academic Press.

Hachinski, V. C., Lassen, N. A., & Marshall, J. (1974). Multi-infarct dementia: A cause of mental deterioration in the elderly. Lancet, 2, 207-209.

Haddad, L. B., & Coffman, T. L. (1987). A brief neuropsychological screening exam for psychiatric-geriatric patients. Clinical Gerontologist, 6(3), 3-10.

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp. 101-120). New York: Oxford University Press.

Heeren, T. J., & Rooymans, H. G. M. (1991). Reply--The Mini-Mental Status Examination. Journal of the American Geriatrics Society, 39(7), 733.

Herst, L. D., Voss, C. B., & Waldman, J. (1990). Cortical function assessment in the elderly. Journal of Neuropsychiatry, 2(4), 385-390.

Hochanadel, G., & Kaplan, E. (1984). Neuropsychology of normal aging. In M. L. Albert (Ed.), Clinical neurology of aging (pp. 231-244). New York: Oxford University Press.

Jacobs, J. W., Bernhard, M. R., Delgado, A., & Strain, J. J. (1977). Screening for organic mental syndromes in the medically ill. Annals of Internal Medicine, 86, 40-46.

Kafonek, S., Ettinger, W. H., Roca, R., Kittner, S., Taylor, N., & German, P. S. (1989). Instruments for screening depression and dementia in a long-term care facility. Journal of the American Geriatrics Society, 37(1), 29-34.

Kallman, H., & May, H. J. (1989). Mental status assessment in the elderly. Primary Care, 16(2), 329-347.

Kaufman, D. M., Weinberger, M., Strain, J. J., & Jacobs, J. W. (1979). Detection of cognitive deficits by a brief mental status examination. General Hospital Psychiatry, 1, 247-255.

Kiernan, R. J., Mueller, J., Langston, J. W., & Van Dyke, C. (1987). The Neurobehavioral Cognitive Status Examination: A brief but differentiated approach to cognitive assessment. Annals of Internal Medicine, 107, 481-485.

Klein, L. E., Roca, R. P., McArthur, J., Vogelsang, G., Klein, G. B., Kirby, S., & Folstein, M. (1985). Diagnosing dementia: Univariate and multivariate analyses of the mental status examination. Journal of the American Geriatrics Society, 33(7), 483-488.

Klonoff, H., Fibinger, C. H., & Hutton, G. (1970). Neuropsychological patterns in chronic schizophrenia. The Journal of Nervous and Mental Disease, 150(4), 291-300.

Klove, H. (1974). Validation studies in adult clinical neuropsychology. In R. M. Reitan & L. A. Davison (Eds.), Clinical neuropsychology: Current status and application. Washington, DC: Winston.

Koenig, H. G., Meador, K. G., Cohen, H. J., & Blazer, D. G. (1988). Self-rated depression scales and screening for major depression in older hospitalized patients with medical illness. Journal of the American Geriatrics Society, 36, 699-706.

La Rue, A. (1982). Memory loss and aging. Psychiatric Clinics of North America, 5, 89-103.

La Rue, A., Yang, J., & Osato, S. (1992). Neuropsychological assessment. In J. E. Birren & R. B. Sloane (Eds.), Handbook of mental health and aging (2nd ed.) (p. 653). New York: Academic Press.

Larner, S. (1977). Encoding in senile dementia and elderly depressives: A preliminary study. British Journal of Social and Clinical Psychology, 16, 379-390.

Larson, E. (1986). Evaluating the validity of screening tests. Nursing Research, 35, 186-188.

Matthews, C. G., Shaw, D. J., & Klove, H. (1966). Psychological test performance in neurologic and "pseudoneurologic" subtests. Cortex, 2, 244-253.

McAllister, T. W. (1983). Overview: Pseudodementia. American Journal of Psychiatry, 140, 528-533.

McCartney, J. R., & Palmateer, L. M. (1985). Assessment of cognitive deficit in geriatric patients: A study of physician behavior. Journal of the American Geriatrics Society, 33(7), 467-471.

McDougall, G. J. (1990). A review of screening instruments for assessing cognition and mental status in older adults. Nurse Practitioner, 15(11), 18-28.

Mohr, J. P., Fisher, C. M., & Adams, R. D. (1980). Cerebrovascular diseases. In K. Isselbacher & R. D. Adams (Eds.), Harrison's principles of internal medicine (pp. 1911-1941). New York: McGraw-Hill.

Morris, J. C., & Fulling, K. (1983). Early Alzheimer's disease: Diagnostic considerations. Archives of Neurology, 45, 345-356.

Nelson, A., Fogel, B. S., & Faust, D. (1986). Bedside cognitive screening instruments: A critical assessment. The Journal of Nervous and Mental Disease, 174(2), 73-83.

NINCDS-ADRDA Work Group. (1985). Criteria for the clinical diagnosis of Alzheimer's disease: Excerpts from the NINCDS-ADRDA Work Group report. Journal of the American Geriatrics Society, 33, 2-3.

Omer, H. J., Foldes, J., Toby, M., & Menczel, J. (1983). Screening for cognitive deficits in a sample of hospitalized geriatric patients: A reevaluation of a brief mental status questionnaire. Journal of the American Geriatrics Society, 31, 266-268.

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. Journal of the American Geriatrics Society, 23, 433-441.

Pirozzolo, F. J., Christensen, K. L., & Ogle, K. M. (1981). Simple and choice reaction time in dementia: Clinical implications. Neurobiology of Aging, 2, 113-117.

Rabins, P. V. (1983). Reversible dementia and the misdiagnosis of dementia: A review. Hospital and Community Psychiatry, 34, 830-835.

Reed, B. R., Jagust, W. J., & Seab, J. P. (1989). Mental status as a predictor of daily function in progressive dementia. The Gerontologist, 29(6), 804-807.

Reitan, R. M. (1955a). Certain differential effects of left and right cerebral lesions in human adults. Journal of Comparative and Physiological Psychology, 48, 474-477.

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

Reitan, R. M. (1959). The comparative effects of brain damage on the Halstead Impairment Index and the Wechsler-Bellevue Scale. Journal of Clinical Psychology, 15, 281-285.

Reitan, R. M. (1966). A research program on the psychological effects of brain lesions in human being. In N. R. Ellis (Ed.), International review of research in mental retardation (Vol. 1, pp. 153-218). New York: Academic Press.

Reitan, R. M., & Wolfson, D. (1975). The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation. Tucson, AZ: Neuropsychology Press.

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

Reitan, R. M., & Wolfson, D. (1986). The Halstead-Reitan Neuropsychological Test Battery and aging. Clinical Gerontologist, 5 (1-2), 39-61.

Reitan, R. M., & Wolfson, D. (1988). Traumatic brain injury: Volume 2. Recovery and rehabilitation. Tucson, AZ: Neuropsychology Press.

Robbins, D. E. (1989). The Halstead-Reitan Neuropsychological Battery. In M. D. Franzen (Ed.), Reliability and validity in neuropsychological assessment (pp. 91-118). New York: Plenum Press.

Russell, E. W. (1987). Neuropsychological interpretation of the WAIS. Neuropsychology, 1, 2-6.

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

Salmon, D. P., Thal, L. J., Butters, N., & Heindel, W. C. (1990). Longitudinal evaluation of dementia of the Alzheimer type: A comparison of three standardized mental status examinations. Neurology, 40(8), 1225-1230.

Schmitt, F. A., Ranseen, J. D., & DeKosky, S. T. (1989). Cognitive mental status examinations. Clinics in Geriatric Medicine, 5(3), 545-564.

Schreiber, D. J., Goldman, H., Kleinman, K. M., Goldfader, P. R., & Snow, M. Y. (1976). The relationship between independent neuropsychological and neurological detection and localization of cerebral impairment. Journal of Nervous and Mental Disease, 162, 360-365.

Schwamm, L. H., Van Dyke, C., Kiernan, R. J., Merrin, E. L., & Mueller, J. (1987). The Neurobehavioral Cognitive Status Examination: Comparison with the Cognitive Capacity Screening Examination and the Mini-Mental State Examination in a neurosurgical population. Annals of Internal Medicine, 107(4), 486-491.

Silbermann, E. K., Weingartner, H., Laraia, M., Byrnes, S., & Post, R. M. (1983). Processing of emotional properties of stimuli by depressed and normal subjects. Journal of Nervous and Mental Disease, 171, 10-14.

Stuss, D. T., & Benson, D. F. (1986). The frontal lobes. New York: Raven Press.

Teng, E. L., Chui, H. C., Schneider, L. S., & Metzger, L. E. (1987). Alzheimer's dementia: Performance on the Mini-Mental State Examination. Journal of Consulting and Clinical Psychology, 55, 96-100.

Tsai, L., & Tsuang, M. T. (1979). The Mini-Mental State test and computerized tomography. American Journal of Psychiatry, 136, 436-439.

Uhlmann, R. F., & Larson, E. B. (1991). Effect of education on the Mini-Mental State Examination as a screening test for dementia. Journal of the American Geriatrics Society, 39, 876-880.

Uhlmann, R. F., Teri, L., Rees, T. S., Mozlowski, K. J., & Larson, E. B. (1989). Impact of mild to moderate hearing loss on mental status testing. Journal of the American Geriatrics Society, 37(3), 223-228.

Vega, A., Jr., & Parsons, O. A. (1967). Cross-validation of the Halstead-Reitan tests for brain damage. Journal of Consulting Psychology, 31, 619-625.

Webster, J. S., Scott, R. R., Nunn, B., McNeer, M. F., & Varnell, N. (1984). A brief neuropsychological screening procedure that assesses left and right hemispheric function. Journal of Clinical Psychology, 40, 237-240.

Weingartner, H., Gold, P., Ballenger, J. D., Smallberg, S. A., Summers, R., Rubinow, D. R., Post, R. M., & Goodwin, F. K. (1981). Effects of vasopressin on human memory function. Science, 211, 601-603.

Weingartner, H., Kaye, W., Smallberg, S., Cohen, R., Ebert, M. H., Gillin, J. C., & Gold, P. (1982). Determinants of memory failures in dementia. In S. Corkin, K. L. Davis, J. H. Growdon, E. Usdin, & R. J. Wurtman (Eds.), Alzheimer's disease: A report of progress in research (pp. 171-176). New York: Raven Press.

Winograd, C. H., & Jarvik, L. F. (1986). Physician management of the demented patient. Journal of the American Geriatrics Society, 34, 295-308.

Yazdanfar, D. (1990). Assessing the mental status of the cognitively impaired elderly. _Journal of Gerontological Nursing, 16_(9), 32-36.